Prof.ssa Elisabetta Ulivi

Terza dispensa

APPUNTI DI PROBABILITÀ E DI STATISTICA

INDICE

Elementi di Probabilità e di Statistica

Da M. Martelli, *Dispense di Istituzioni di Matematiche*
Con integrazioni tratte da: P. Marcellini-C. Sbordone, *Elementi di Calcolo*, e da S. C. Ross, *Calcolo delle Probabilità*.

Da L. Barletti, *Calcolo delle Probabilità e Statistica* e R. Ricci, *Appunti di Probabilità e Statistica*

Da M. Abate, *Matematica e statistica. Le basi per le scienze della vita*

Elaborazione statistica dei dati sperimentali
Si rimanda a P. Marcellini-C. Sbordone, *Elementi di Calcolo*, p. 385

Elementi di Probabilità e di Statistica

da M. Martelli, *Dispense di Istituzioni di Matematiche*

Eventi

Se di fronte ad ogni esperimento od osservazione non esistessero alternative sul risultato non avrebbe senso occuparsi di probabilità e di statistica. Invece in generale ogni prova, ogni esperimento, ogni osservazione può condurre a diversi risultati. In generale sappiamo quali essi possono essere ma ignoriamo quale di essi si verificherà realmente. Così se da un'urna contenente i primi 90 numeri naturali estraiamo a caso uno di essi sappiamo benissimo che abbiamo davanti a noi 90 casi possibili, ma ignoriamo quale di essi si verificherà. Analogamente di fronte ad una partita di calcio fra due squadre A e B sappiamo bene che (salvo imprevisti) potrà verificarsi uno solo di questi tre casi: vince A, vince B, pareggio; ma ignoriamo quale sarà dei tre a verificarsi.

Tuttavia mentre nel caso dell'estrazione di un numero da un'urna che ne contiene 90 non sappiamo proprio che risultato aspettarci e pensiamo che ognuno dei 90 numeri abbia le stesse probabilità di essere prescelto, nel caso invece della partita possiamo azzardare delle previsioni sulla base della conoscenza che abbiamo delle due squadre. Cioè l'esperienza può fornir ci delle indicazioni sulla possibilità che un certo risultato sia più probabile di un altro.

D'altra parte se compiamo 4000 estrazioni dalla nostra urna, sempre rimettendo a posto, dopo ogni estrazione, il numero sorteggiato e mescolando bene il tutto e se dopo 4000 estrazioni il numero 3 non è mai uscito siamo portati a pensare che all'estrazione 4001 esso abbia più probabilità degli altri di venire sorteggiato.

La probabilità e la statistica si incaricano di quantificare queste semplici osservazioni che noi abbiamo fatto e di scoprire quella parte di determinismo che si nasconde in ogni situazione e che può fornire delle indicazioni su come essa si evolverà.

Supponiamo allora che di fronte ad un esperimento, ad una prova qualunque si sia in grado di precisarne tutti i risultati possibili. Ciascuno di essi è tradotto in una <u>proposizione</u>. Ad esempio nel caso del lancio di un dado i sei risultati possibili, tradotti ciascuno in una proposizione, sono

esce l'uno, esce il due, esce il tre,
esce il quattro, esce il cinque, esce il sei.

Nel caso invece della nascita di un bambino i due risultati possibili, con riferimento al suo sesso, sono, tradotti ciascuno in una proposizione

è una femmina, è un maschio.

Si osservi che le proposizioni devono essere formulate in modo che sia possibile poi stabilire con certezza se si è verificato o meno l'evento da esse annunciato.

Con tali premesse si capisce perchè chiamiamo <u>evento ogni fatto del quale si possa decidere senza ambiguità sul suo verificarsi o sul suo non verificarsi</u>.

Così non è un evento il seguente: "il missile C partirà da capo Kennedy oggi alle 17" perchè manca nella sua formulazione un corretto riferimento temporale. (ora italiana, ora locale?).

Un evento si dirà certo se esso si verifica senz'altro e impossibile se esso senz'altro non si verifica. Ogni altro evento sarà detto possibile. Due eventi si diranno incompatibili se non possono mai verificarsi contemporaneamente. Un evento si dirà elementare se non può essere considerato come unione di altri eventi.

E' conveniente considerare l'evento certo come l'insieme S formato da tutti gli eventi elementari, identificare ogni altro evento possibile con un sottoinsieme non vuoto di S e l'evento impossibile con l'insieme vuoto.

Esempio 1 - Nel lancio di un dado possiamo identificare S con l'insieme dei primi 6 numeri naturali $S = \{1,2,3,4,5,6\}$. Ciascuno di essi rappresenta un evento elementare. L'evento "esce un numero pari" è rappresentato dal sottoinsieme A di S formato dal 2, dal 4 e dal 6. L'evento "esce un numero primo" è rappresentato dal sottoinsieme B di S che consiste degli elementi 2,3,5. S è l'evento certo perchè esce sicuramente un numero compreso fra 1 e 6.

Esempio 2 - Nel gioco del lotto possiamo identificare S con l'insieme dei primi 90 numeri naturali cioè $S = \{1,2,\dots,90\}$. Ciascuno di essi rappresenta un evento elementare. L'evento "esce un numero minore o eguale di 30" è rappresentato dal sottoinsieme $A = \{1,2,\dots,30\}$. L'evento "esce un numero più grande di 60" è rappresentato dal sottoinsieme $B = \{61,62,\dots,90\}$.

Nei due esempi precedenti abbiamo visto come tradurre un evento possibile in un sottoinsieme di S. E' naturale chiedersi se viceversa possiamo interpretare ogni sottoinsieme di S come un possibile evento. La risposta è affermativa nel caso che S sia un insieme finito oppure numerabile. Se S è più che numerabile si possono dare degli esempi in cui certi suoi sottoinsiemi non sono degli eventi.

Operazioni logiche sugli eventi.

Da ora in avanti identificheremo gli eventi con i sottoinsiemi di S che li rappresentano.

Supponiamo che $A, B \subset S$ siano due eventi.

(a) $A \cup B$ è l'evento che si verifica se si verifica o A o B (non escludendo che si verifichino ambedue)

(b) $A \cap B$ è l'evento che si verifica se si verificano sia A che B. Gli eventi A, B sono quindi incompatibili se e solo se $A \cap B = \emptyset$

(c) A^c è l'evento che si verifica se non si verifica A.

Esempio - I due eventi "esce un numero pari" "esce un numero primo" nel gioco dei dadi portano agli insiemi $A = \{2,4,6\}$ e $B = \{2,3,5\}$. Allora

$A \cup B = \{2,3,4,5,6\}$ è l'evento "non esce l'1".

$A \cap B = \{2\}$ è l'evento "esce il 2"

$A^c = \{1,3,5\}$ è l'evento "esce un numero dispari".

Ad ogni evento possibile è spontaneo cercar di associare un grado di possibilità, che esprima la maggiore o minor fiducia che abbiamo del suo verificarsi. Tale grado di possibilità sarà concretizzato in un numero chiamato probabilità dell'evento considerato. In certi casi questo può essere fatto senza ricorrere all'esperienza. In altri casi invece non possiamo farlo senza l'ausilio dell'esperienze passate. In ambedue i casi le esperienze future saranno utilissime per verificare se il grado di possibilità assegnato ad un certo evento è corretto o meno.

Il concetto di probabilità

Ci sono tre modi per introdurre e stimare la probabilità di un evento:

Approccio elementare, detto anche classico o a priori:
Se un evento A si può verificare in k modi diversi (casi favorevoli) su n casi possibili, essendo questi tutti equivalenti cioè ugualmente possibili, la probabilità p(A) dell'evento è:

p(A) = casi favorevoli / casi possibili = k/n dove $0 \leq k \leq n$ e quindi $0 \leq p \leq 1$

Approccio frequentistico o statistico o a posteriori:
Ripetiamo n volte un esperimento e supponiamo che un certo evento A si verifichi k volte. Diciamo k frequenza dell'evento A. Definiamo frequenza relativa

f = casi favorevoli / casi osservati = k/n

La base della teoria della probabilità consiste nel presupporre, cosa del resto confortata dall'esperienza, che, salvo casi eccezionali, si verifica che, quando $n \to +\infty$, f tende a stabilizzarsi attorno ad un ben determinato valore p che si dice probabilità dell'evento.

Sia l'approccio elementare che quello frequentistico vanno incontro a serie difficoltà. Ad esempio il primo a causa dell'espressione "ugualmente possibili". Per questo si preferisce l'"approccio assiomatico" che si basa sugli insiemi.

<u>Assiomi della probabilità.</u>

Sia $\mathcal{E}$ l'insieme i cui elementi sono tutti quei sottoinsiemi di S che sono eventi. Come abbiamo notato in precedenza se S è finito o numerabile allora ogni suo sottoinsieme è un evento. In tal caso $\mathcal{E}$ coincide con l'insieme di tutte le parti di S. E' poi evidente che $\emptyset$, $S \in \mathcal{E}$.

Una funzione $p : \mathcal{E} \to \mathbb{R}$ è detta <u>funzione di probabilità</u> se soddisfa alle seguenti condizioni:

P_1. $0 \leq p(A) \leq 1$ qualunque sia $A \in \mathcal{E}$

P_2. $p(S) = 1$

P_3. Se A e $B \in \mathcal{E}$ e si ha $A \cap B = \emptyset$ (in S) allora $p(A \cup B) = p(A) + p(B)$.

Se p soddisfa alle tre condizioni imposte allora si può dimostrare facilmente che

(1) $p(\emptyset) = 0$

(2) se $A_1, A_2, \ldots, A_n \in \mathcal{E}$ e si ha $A_i \cap A_j = \emptyset$ (in S) ogni volta che $i \neq j$ allora

$p(A_1 \cup A_2 \cup A_3 \ldots \cup A_n) = p(A_1) + p(A_2) + \ldots + p(A_n)$

In altre parole la P_3 supposta vera per due insiemi si dimostra vera per un numero n di insiemi purchè in S essi siano a 2 a 2 disgiunti.

(3) Se $A, B \in \mathcal{E}$ e $A \subset B$ (in S) allora $p(A) \leq p(B)$

(4) se $A, B \in \mathcal{E}$ allora

$$\boxed{p(A \cup B) = p(A) + p(B) - p(A \cap B)}$$

Per provare quest'ultima eguaglianza si può, ad esempio, procedere così. Poichè $A = (A \setminus B) \cup (A \cap B)$ avremo, in base alla P_3 (essendo i due insiemi $A \setminus B$ e $A \cap B$ disgiunti

$$p(A) = p(A \setminus B) + p(A \cap B)$$

Quindi anche $\quad p(B) = p(B \setminus A) + p(A \cap B)$

D'altra parte $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$ e poichè $A \setminus B$, $B \setminus A$, $A \cap B$ sono a due a due disgiunti avremo, in base al risultato (2)

$$p(A \cup B) = p(A \setminus B) + p(B \setminus A) + p(A \cap B)$$

Sostituendo $p(A \setminus B)$ con $p(A) - p(A \cap B)$ e $p(B \setminus A)$ con $p(B) - p(A \cap B)$ si ottiene infine

$$\boxed{p(A \cup B) = p(A) + p(B) - p(A \cap B)}$$

In particolare si osservi che se $B = A^c$ si ottiene (si ricordi che $A \cup A^c = S$ e $A \cap A^c = \emptyset$)

$$1 = p(S) = p(A \cup A^c) = p(A) + p(A^c)$$

cioè

$$\boxed{p(A^c) = 1 - p(A)}$$

Osservazione:
Dagli assiomi della probabilità si deduce facilmente che la probabilità secondo la concezione classica (p = casi favorevoli / casi possibili) si ottiene come caso particolare della probabilità secondo l'impostazione assiomatica, quando si suppone che gli eventi elementari siano equiprobabili.

Esempio 1 - Scegliamo a caso 2 articoli da un insieme di 12 articoli dei quali 4 sono difettosi. Vogliamo sapere la probabilità che

(a) ambedue gli articoli scelti siano difettosi

(b) ambedue gli articoli scelti siano non difettosi.

Lo spazio degli eventi S è costituito da $\binom{12}{2}$ eventi semplici, cioè tanti quanti sono i sottoinsiemi distinti formati da due elementi in un insieme di n elementi.
Quindi S è costituito da $\frac{12!}{2!10!} = 66$ eventi elementari.
E' ragionevole supporre che ognuno di essi abbia la stessa probabilità. Siccome, d'altra parte, gli articoli difettosi sono 4 allora gli eventi elementari interessati alla scelta (a) saranno $\binom{4}{2} = 6$. Perciò la probabilità di scegliere due articoli di-

fettosi sarà $\frac{6}{66} = \frac{1}{11}$.

Invece la probabilità che nessuno dei due sia difettoso sarà $\frac{28}{66} = \frac{14}{33}$ poichè $\binom{8}{2} = 28$.

A questo punto possiamo anche stabilire la probabilità che almeno 1 sia difettoso. Essa sarà $1 - \frac{14}{33} = \frac{19}{33}$, perchè i due eventi "nessuno dei due difettoso" e almeno uno difettoso" sono complementari.

Esempio 2 - (Problema classico delle nascite). Cerchiamo la probabilità p che n persone ($n \leq 365$) abbiano tutte compleanni diversi.

Lo spazio S degli eventi elementari è costituito da tutti i possibili modi in cui si possono distribuire i compleanni delle n persone. Se si trascurano gli anni bisestili avremo che tali modi sono 365^n. Supponiamo ora che ogni persona abbia eguale probabilità di nascere in un qualunque giorno dell'anno. Allora perchè si abbiano compleanni diversi alla prima persona daremo la possibilità di essere nata in uno qualunque dei 365 giorni dell'anno, alla seconda in uno qualunque dei 364 giorni che rimangono, alla terza in uno qualunque dei 363 giorni che rimangono etc. ...

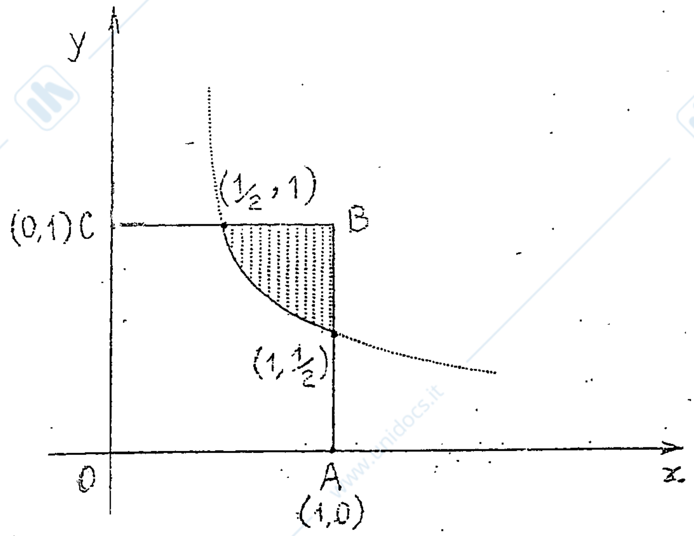
Perciò dei 365^n modi diversi di distribuire i compleanni delle nostre n persone conducono a compleanni fra loro diversi solo $365.364.363...(365-n+1)$ modi.

Quindi la probabilità che le n persone celebrino il loro compleanno in giorni diversi è

$$p = \frac{365.364...(365-n+1)}{365^n}.$$

Si può dimostrare che se $n \geq 23$ allora $p < \frac{1}{2}$. In altre parole fra 23 o più persone è più facile trovare che almeno due hanno lo stesso compleanno di quanto non lo sia il contrario.

Esempio 3 - Supponiamo che un tiratore spari su un bersaglio quadrato come in figura e che si possa essere certi che lo colpirà. Si vuole la probabilità che egli colpisca la regione tratteggiata



S sarà l'insieme dei punti del quadrato OABC. La probabilità che venga colpito un certo sottoinsieme di S sarà eguale all'area del sottoinsieme stesso (supponendo che se ne possa definire l'area).

L'area della regione che interessa in questo problema è

$$\frac{1}{2} - \int_{1/2}^{1} \frac{1}{2x} dx = \frac{1}{2} - \frac{1}{2} \ln x \Big|_{\frac{1}{2}}^{1} = \frac{1}{2}(1-\ln 2)$$

- Probabilità condizionata.

Cominciamo con un esempio semplice.

Si gettano due dadi. Supponiamo che la somma ottenuta sia 6 (evento A). Vogliamo sapere la probabilità che uno dei due dadi sia un 2 (evento B).

Qui l'evento: uno dei due dadi è un 2, è subordinato all'altro evento: la somma dei due dadi è eguale a 6. Si tratta cioè di determinare una probabilità condizionata. Non ci interessa in assoluto che esca un 2, ma ci interessa che esca in un lancio che ha dato per somma 6. E' come se il nuovo insieme di eventi fosse diventato: lancio con somma 6 e noi vogliamo sapere quale è la probabilità che uno dei due numeri sia un 2.

Poichè per avere somma 6 occorre che siano date le seguenti 5 coppie, tra le quali la coppia (3,3) conta due volte

$$(1,5),\ (5,1),\ (3,3),\ (2,4),\ (4,2)$$

e il 2 compare in due di esse avremo che la probabilità richiesta, che viene indicata con $p(B|A)$, è $\frac{2}{6} = \frac{1}{3}$.

Sia dunque $A \subset S$ un evento con $p(A) > 0$. La probabilità che si verifichi B supposto che A si sia verificato è detta probabilità condizionata di B relativamente ad A ed è indidicata con $p(B|A)$.

Se $A \cap B = \emptyset$ allora $p(B|A) = 0$ perchè i due eventi sono incompatibili e quindi la probabilità che si verifichi B una volta verificato A è 0, perchè tale eventualità è impossibile.

Se invece $B \subset A$ allora $p(B|A) = p(B)/_{p(A)}$ perchè il verificarsi di B è del tutto subordinato al verificarsi di A. Così nel lancio di un dado l'evento "esce il 2" (evento B) è contenuto nell'evento "esce un numero pari" (evento A). E' evidente infatti che non può uscire il 2 se non esce un numero pari.

Perciò una volta certi che è uscito un numero pari la probabilità che esso sia un 2 è $\frac{1}{3}$ ossia $\frac{1/6}{1/2}$, cioè la probabilità che esca un 2 ($\frac{1}{6}$) divisa per la probabilità che esca un numero pari ($\frac{1}{2}$).

In generale quando si richiede $p(B|A)$ è come se il nuovo insieme di eventi fosse diventato A, perchè tale evento si suppone verificato, e in questo nuovo insieme vogliamo la probabilità di quella parte di B che vi è interessata.

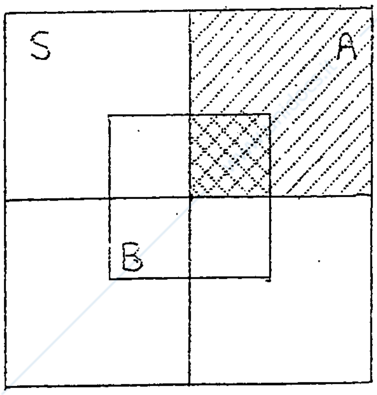
Avremo quindi il seguente

Teorema (delle probabilità composte)

La probabilità di B relativamente ad A, supposto $p(A) > 0$, è il quoziente fra la probabilità dell'evento intersezione, $A \cap B$, e la probabilità di A.

$$(1) \qquad \boxed{p(B|A) = \frac{p(B \cap A)}{p(A)}}$$

A titolo di esempio supponiamo che S sia il quadrato in figura, A sia la regione tratteggiata e B sia il quadrato al centro



E' chiaro che rispetto ad S $p(A \cap B) = \frac{1}{16}$, ma rispetto ad A la sua probabilità vale $\frac{1}{4}$. Tale probabilità è quella che abbiamo indicato con $p(B|A)$. Si verifica facilmente che

$$\frac{1}{4} = p(B|A) = \frac{p(A \cap B)}{p(A)} = \frac{1}{16} : \frac{1}{4}$$

La formula (1) può anche essere scritta nella forma

$$(2) \qquad \boxed{p(B|A) \cdot p(A) = p(B \cap A)}$$

e, scambiando i ruoli di A e B,

$$(3) \qquad \boxed{p(A|B) . p(B) = p(A \cap B)}$$

Se poi si hanno n insiemi $A_1, A_2, \ldots, A_n$ la (2) e la (3) si generalizzano nel seguente modo

$$(4)\quad \boxed{p(A_1 \cap A_2 \cap \ldots \cap A_n) = p(A_1)\cdot p(A_2|A_1)\cdot p(A_3|A_1 \cap A_2)\cdot\ldots \cdot p(A_n|A_1 \cap A_2 \ldots \cap A_{n-1})}$$

Esempio - Da una scatola contenente 12 articoli, quattro dei quali sono difettosi, se ne scelgono a caso 3. Determinare la probabilità che nessuno dei tre sia difettoso.

1° modo. Le possibili terne di articoli sono $\binom{12}{3} = \frac{12!}{3!9!} = 220$. Le terne non difettose sono $\binom{8}{3} = \frac{8!}{3!5!} = 56$. Perciò la probabilità cercata è $\frac{56}{220} = \frac{14}{55}$

2° modo (usando il risultato (4)).

La probabilità che il primo oggetto scelto sia non difettoso è $\frac{8}{12}$ perchè si hanno 8 articoli perfetti su un totale di 12. Se il primo articolo è perfetto la probabilità che anche il secondo lo sia è $\frac{7}{11}$ perchè sono rimasti 7 articoli perfetti su un totale di 11. Se i primi due articoli scelti non sono difettosi la probabilità che anche il terzo non lo sia è $\frac{6}{10}$ perchè sono rimasti 6 articoli perfetti su un totale di 10. Avremo dunque, per il teorema delle probabilità composte nella forma (4), che la probabilità cercata è

$$\frac{8}{12}\cdot\frac{7}{11}\cdot\frac{6}{10} = \frac{14}{55}$$

Eventi indipendenti

Quando si considera $p(B|A)$ può capitare che

$$p(B|A) > p(B)$$

$$p(B|A) < p(B)$$

$$p(B|A) = p(B)$$

Nel primo caso la conoscenza che A si è verificato aumenta la probabilità attribuibile a B. Nel secondo caso invece la diminuisce.

Nel terzo caso si dice che B è indipendente da A perchè sia che A risulti verificato sia che A non risulti verificato la probabilità di B non cambia.

Siano A e B due eventi con probabilità positiva.

Se B è indipendente da A avremo

$$\boxed{p(A \cap B) = p(A).p(B|A) = p(A).p(B)} \tag{5}$$

D'altra parte è anche

$$p(A \cap B) = p(B)\ p(A|B)$$

e, confrontando con la (5), si deduce che

$$p(A|B) = p(A)$$

Cioè anche A è indipendente da B.

Molti considerano la (5) come definizione di eventi indipendenti, cioè definiscono indipendenti due eventi A, B se

$$\boxed{p(A \cap B) = p(A).p(B)}$$

Si osservi che eventi indipendenti e eventi incompatibili sono due cose molto diverse. Per gli eventi incompatibili si ha

$$\boxed{p(A \cup B) = p(A) + p(B)}$$

mentre per quelli indipendenti si ha

$$\boxed{p(A \cap B) = p(A).p(B)}$$

<u>Esempio 1</u> - Quale probabilità ha un uomo di 70 anni di arrivare a 72 anni?

La seguente tabella riporta le probabilità che un uomo di 70 anni arrivi a 71 anni, che uno di 71 anni arrivi a 72 etc.

età	probabilità
70	$P_{70} = 0{,}9492$
71	$P_{71} = 0{,}9444$
72	$P_{72} = 0{,}9391$

Detto A l'evento "un uomo di 70 anni arriva a 71" e B l'evento "un uomo di 71 anni arriva a 72" in base alla formula (2) si ha

$$p(A \cap B) = p(A) . p(B|A) = 0{,}9492 \times 0{,}9444 = 0{,}8964,$$

ossia 89,64%. (Si noti che in questo caso $B|A$ è lo stesso evento che B).

<u>Esempio 2</u> - Si vuole sapere se il daltonismo e la sordità sono o non sono indipendenti. La tabella seguente riporta le percentuali di sordi-daltonici, sordi-non daltonici, non sordi-daltonici e non sordi-non daltonici.

	sordi	non sordi	totale
daltonici	0,0004	0,0796	0,0800
non daltonici	0,0046	0,9154	0,9200
totale	0,0050	0,9950	1,0000

Avremo $p(D_a|S_o) = \dfrac{p(D_a \cap S_o)}{p(S_o)} = \dfrac{0{,}0004}{0{,}0050} = 0{,}08 = p(D_a)$

D'altra parte $p(D_a^c|S_o) = \dfrac{0{,}0046}{0{,}0050} = 0{,}92 = p(D_a^c)$

Perciò i due difetti sono indipendenti.

La definizione di eventi indipendenti può essere generalizzata nella stessa misura come è stato fatto per quella di eventi incompatibili. Così tre eventi A, B, C sono detti indipendenti se sono indipendenti a due a due e inoltre

$$\boxed{p(A \cap B \cap C) = p(A).p(B).p(C)} \tag{6}$$

Esempio - Supponiamo di lanciare un dado 3 volte. Vogliamo la probabilità che nel primo lancio esca numero pari, nel secondo lancio esca il 2 e nel terzo lancio escano il 5 o il 6.

Sia A l'evento: esce un numero pari al primo lancio

B l'evento: esce il 2 al secondo lancio

C l'evento: escono il 5 o il 6 al terzo lancio.

E' chiaro che i tre eventi sono indipendenti. Quindi la probabilità cercata è

$$\frac{1}{2} \cdot \frac{1}{6} \cdot \frac{1}{3} = \frac{1}{36}$$

Avremmo anche potuto procedere così. Tutte le possibili terne ordinate di numeri compresi fra 1 e 6 sono $6^3 = 216$. Di queste la metà, cioè 108, cominciano con un numero pari. Di queste 108 solo $\frac{1}{6}$, cioè 18, hanno come seconda cifra il 2. Di queste 18 solo $\frac{1}{3}$, cioè 6, terminano con il 5 o con il 6. Perciò si tratta di scegliere una di queste 6 terne fra 216. La probabilità è quindi

$$\frac{6}{216} = \frac{1}{36}$$

Il teorema di Bayes.

Siano $A_1, A_2, \dots, A_n$ n eventi a due a due disgiunti e tali che $A_1 \cup A_2 \cup \dots \cup A_n = S$ (si suole dire che essi formano una partizione di S).

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

Sia poi $B \subset S$. Avremo che

$$B = B \cap S = B \cap (A_1 \cup A_2 \dots \cup A_n) =$$

$$(B \cap A_1) \cup (B \cap A_2) \cup \dots \cup (B \cap A_n)$$

dove $A_i \cap B$, $i = 1,2,\dots n$, formano una partizione di B. Avremo quindi

$$(*) \quad p(B) = p(A_1 \cap B) + p(A_2 \cap B) + \dots + p(A_n \cap B)$$

Ricordando che $p(A_i \cap B) = p(A_i).p(B|A_i)$ avremo

$$(**) \quad p(B) = p(A_1).p(B|A_1) + \dots + p(A_n)p(B|A_n)$$

Inoltre si ha $p(A_i|B) = \dfrac{p(A_i \cap B)}{p(B)} = \dfrac{p(A_i)p(B|A_i)}{p(B)}$

sostituendo in tale eguaglianza $p(B)$ dato dalla (**) otteniamo

<u>Teorema</u> (di Bayes).

$$\boxed{p(A_i|B) = \frac{p(A_i)p(B|A_i)}{p(A_1).p(B|A_1) + \dots + p(A_n).p(B|A_n)}}$$

<u>Esempio 1</u> - Tre macchine A, B, C producono rispettivamente il 50%, il 30% e il 20% degli articoli di una fabbrica. La percentuale di quelli difettosi è, nell'ordine, il 3%, il 4% e il 5%. Supponiamo di scegliere a caso un articolo. Determinare

(a) la probabilità che sia difettoso

(b) la probabilità che sia stato prodotto dalla macchina A sapendo che è difettoso

Sia D l'evento: articolo difettoso. In base alla (**) avremo

(a) $p(D) = p(A)p(D|A) + p(B)p(D|B) + p(C)p(D|C)$

$$= 0{,}50 \cdot 0{,}03 + 0{,}30 \cdot 0{,}04 + 0{,}20 \cdot 0{,}05 = 0{,}037$$

(b) $$p(A|D) = \frac{p(A)p(D|A)}{p(A)p(D|A)+p(B)p(D|B) + p(C)p(D|C)} = \frac{0{,}50 \cdot 0{,}03}{0{,}037} =$$

$$= \frac{15}{37} .$$

Esempio 2 - Un tecnico altamente qualificato monta il 30% di una serie di apparecchi di precisione. Un altro tecnico, meno qualificato, ne monta il 70%. L'apparecchio funziona nel 90% dei casi se è stato montato dal primo tecnico e nell'80% dei casi se è stato montato dal secondo tecnico. Supponiamo di scegliere a caso uno degli apparecchi e che esso funzioni. Determinare la probabilità che sia stato montato dal tecnico altamente qualificato.

Indichiamo con A l'evento: l'apparecchio funziona e con B_1, B_2 gli eventi: montaggio da parte del tecnico più qualificato, montaggio da parte del tecnico meno qualificato. Dobbiamo determinare $p(B_1|A)$. Per il teorema di Bayes avremo

$$p(B_1|A) = \frac{p(B_1) \cdot p(A|B_1)}{p(B_1)p(A|B_1)+p(B_2)p(A|B_2)}$$

Poichè $p(B_1) = 0{,}3$, $p(B_2) = 0{,}7$ $p(A|B_1) = 0{,}9$ e $p(A|B_2) = 0{,}8$ otterremo

$$p(B_1|A) = \frac{0{,}3 \cdot 0{,}9}{0{,}3 \cdot 0{,}9+0{,}7 \cdot 0{,}8} = \frac{0{,}27}{0{,}83} \cong 0{,}325$$

- Variabili aleatorie discrete.

Nella sezione precedente abbiamo esaminato il problema di valutare le probabilità di un evento. Abbiamo raccolto in un insieme S gli eventi elementari relativi ad un certo fenomeno e abbiamo cercato di definire una funzione di probabilità da $\mathcal{E}$

in $\mathbb{R}$, dove $\mathcal{E}$ indicava l'insieme i cui elementi erano tutti i sottoinsiemi di S che rappresentavano degli eventi.

Gli eventi però, elementari o meno, sono restati fino ad ora degli enti non numerici. Per molte questioni invece è opportuno interpretare gli eventi relativi ad un certo fenomeno come sottoinsiemi di $\mathbb{R}$. E' così che si giunge alla nozione di <u>variabile aleatoria</u>.

<u>Definizione.</u> <u>Una variabile aleatoria è una funzione</u> $X : S \to \mathbb{R}$ <u>tale che la retroimmagine di ogni intervallo è un evento.</u>

Se S è un insieme finito o numerabile ogni funzione da S in $\mathbb{R}$ è una variabile aleatoria, perchè ogni sottoinsieme di S è un evento.

L'introduzione delle variabili aleatorie non deve sembrare un'inutile complicazione perchè in molti casi non è tanto rilevante sapere qual'è l'evento che si verificherà con maggiore probabilità, quanto tener conto della probabilità che un evento ha in relazione alla sua importanza. Spieghiamoci meglio con un

<u>Esempio</u> - Un giocatore lancia ripetutamente 1 dado. I patti sono che egli riceve 1000 lire se esce l'uno, 2000 se esce il 2, 3000 se esce il 3 e 4000 se esce il 4. Deve invece pagare 5000 lire se esce il 5 e 6000 se esce il 6. E' chiaro che qui la probabilità che ogni numero ha di uscire è la stessa, ma il fatto che esca un numero piuttosto che un altro ha molta importanza. E benchè il giocatore abbia a disposizione 4 numeri e il banco 2 è chiaro che il gioco è favorevole al banco.

Questo esempio può illustrare meglio quanto ora diremo.

Sia data una variabile aleatoria $X : S \to \mathbb{R}$ e supponiamo che essa assuma un numero finito di valori $x_1, \dots, x_n$. Indichiamo con p_i la probabilità del sottoinsieme di S i cui elementi hanno per immagine x_i, cioè

$X(S) = \{ x_1, x_2, \dots, x_n \}$

Posto $A_i = \{ s \in S : X(s) = x_i \}$ abbiamo $X(A_i) = x_i$ e dunque $A_i = X^{-1}(x_i)$

$p_i = p\{ s \in S : X(s) = x_i \} = p(A_i)$

Tale probabilità è spesso indicata con $p(X = x_i)$

I valori p_i formano quella che è detta distribuzione di probabilità di X.

Essendo l'insieme $\{s \in S : X(s) = x_i\}$ la retroimmagine o preimmagine di x_i secondo X possiamo anche dire che p_i è la probabilità dell'evento preimmagine di x_i secondo X. Inoltre le preimmagini di tutti gli x_i formano una partizione di S, perchè sono a due a due disgiunte e la loro unione da tutto S. Da ciò segue che indicate con A_i tali preimmagini si ha

$$1 = p(S) = p(A_1 \cup A_2 \ldots \cup A_n) = p(A_1) + p(A_2) + \ldots + p(A_n)$$

$$= p_1 + p_2 + \ldots + p_n$$

Chiameremo valor medio o speranza matematica della variabile aleatoria X e lo indicheremo con μ il numero

$$\boxed{\mu = p_1 x_1 + p_2 x_2 + \ldots + p_n x_n = \sum_{i=1}^{n} p_i x_i}$$

μ è detto anche valore atteso

Come si vede nella speranza matematica si sommano, ciascuno con la sua probabilità, i valori della variabile aleatoria X.

Diremo invece varianza di X e lo indicheremo con σ^2 il numero

$$\boxed{\sigma^2 = p_1(x_1-\mu)^2 + p_2(x_2-\mu)^2 + \ldots + p_n(x_n-\mu)^2}$$

e chiameremo σ deviazione standard o scarto quadratico medio della variabile aleatoria X.

Come si vede nella varianza di X compaiono, ciascuno con la sua probabilità, i quadrati degli scarti $(x_i-\mu)^2$, $i = 1,2,\ldots,n$.

Gli esempi che ora daremo vogliono spiegare il significato che può avere una variabile aleatoria, la sua distribuzione di probabilità, il suo valor medio, la varianza e la deviazione standard.

<u>Esempio 1</u> – Supponiamo che un giocatore lanci due dadi. L'insieme, S, degli eventi elementari è costituito dalle 36 coppie ordinate di numeri

$$S = \{(1,1),\ (1,2),\ (2,1), \ldots, (6,6)\}$$

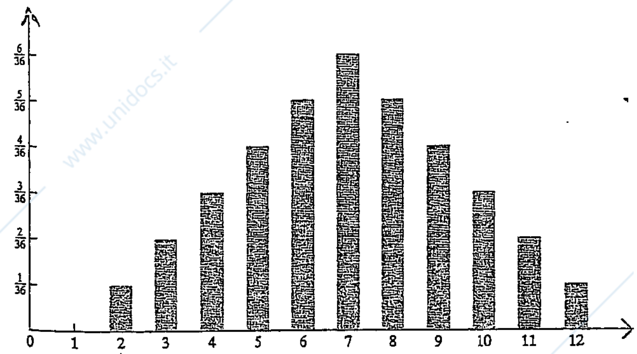
Sia $X : S \to \mathbb{R}$ la funzione che associa ad ogni coppia la somma dei suoi elementi, cioè $X(a,b) = a + b$. Allora X è una variabile aleatoria su S e la sua immagine è

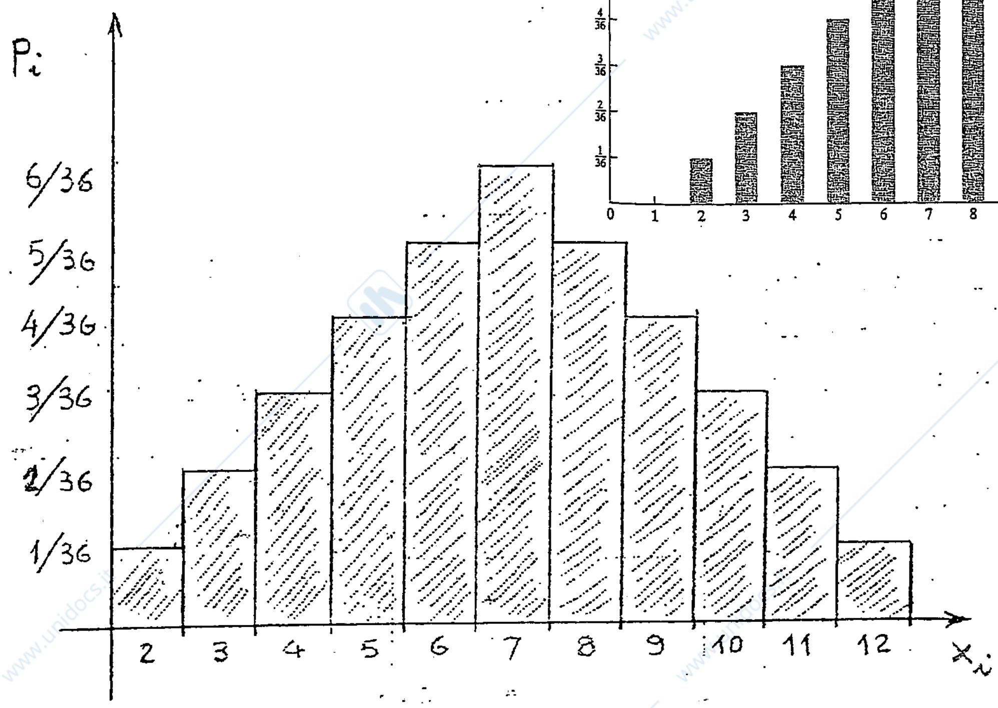
$$X(S) = \{2,3,4,\ldots,11,12\}$$

La distribuzione di probabilità di X è data dalla seguente tabella

x_i	2	3	4	5	6	7	8	9	10	11	12
p_i	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

e, graficamente





Il valore medio di X è dunque

$$\mu = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + \ldots + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} = 7$$

La varianza di X è

$$\sigma^2 = \frac{1}{36}(2-7)^2 + \frac{2}{36}(3-7)^2 + \frac{3}{36}(4-7)^2 + \ldots + \frac{2}{36}(11-7)^2 + \frac{1}{36}(12-7)^2$$

$$= 5{,}8$$

La deviazione standard è dunque $\sigma = \sqrt{5{,}8} \cong 2{,}4$

Esempio 2 - (Distribuzione binomiale). Supponiamo di ripetere n volte un esperimento che ha solo 2 possibili risultati A e B di probabilità a e $b = 1-a$. Supponiamo che le n prove siano indipendenti l'una dall'altra.

Cerchiamo prima di tutto la probabilità che, ad esempio, nelle prime k prove si verifichi A e nelle altre $n-k$ si verifichi B.

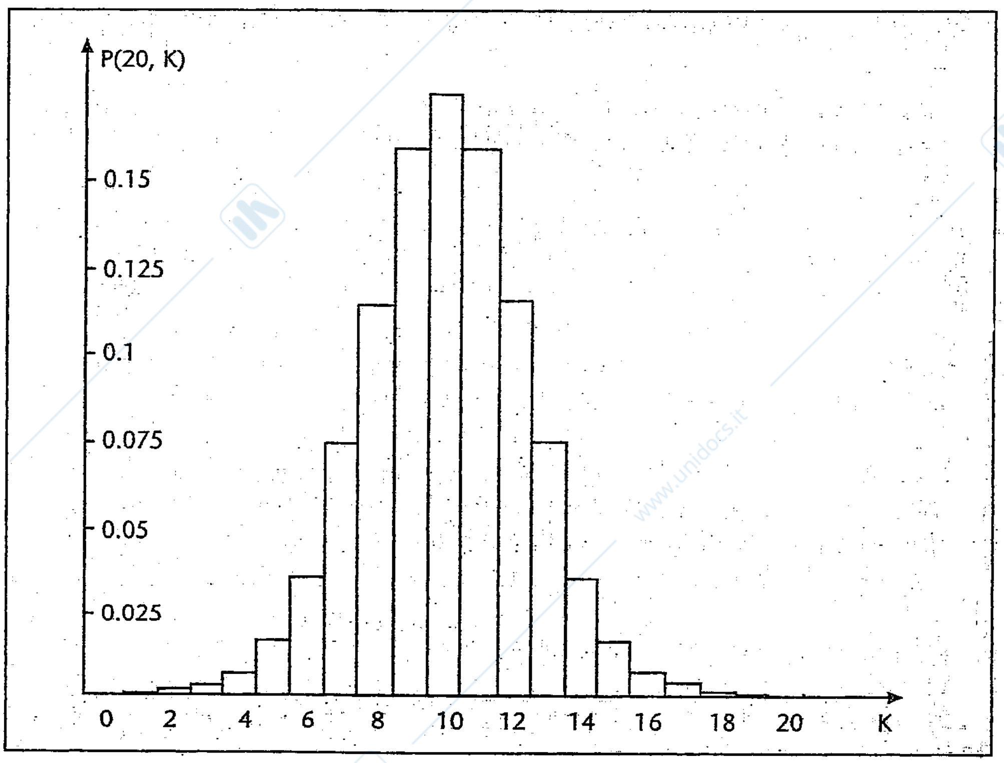
Siccome gli eventi considerati sono indipendenti avremo che la probabilità cercata è

$$\boxed{a^k b^{n-k}}$$

Se ora ci interessa invece che A si verifichi k volte e B $n-k$ volte senza far riferimento all'ordine con cui A e B si verificano troveremo, con un ragionamento analogo a quello fatto per determinare il coefficiente di $a^k b^{n-k}$ nel binomio di Newton, che tale probabilità è

$$\boxed{p(n,k) = \binom{n}{k} a^k b^{n-k}}$$

Fissato n, $p(n, k)$ è una funzione di k (con $k = 0, 1, 2, \dots, n$) detta <u>distribuzione binomiale</u> di parametri n ed a. La seguente figura rappresenta il grafico "a campana" di $p(n, k)$ quando $n = 20$ ed $a = 1/2$ (da P. Marcellini- C. Sbordone, *Elementi di Calcolo*).



In termini di variabili aleatorie:

Sia S l'insieme di tutti i possibili risultati delle n prove e sia $X : S \to \mathbb{R}$ una funzione tale che $X(S) = \{0,1,2,\dots,n\}$ con la condizione che

$X^{-1}(0) = \{$ insieme dei risultati in cui A non compare $\} =$
$\{$ insieme " " " " B compare n volte $\}$

$X^{-1}(1) = \{$ insieme dei risultati in cui A compare 1 volta $\} =$
$\{$ " " " " " B " n-1 volte $\}$

....

$X^{-1}(n) = \{$ insieme dei risultati in cui A compare n volte $\} =$
$\{$ " " " " " B non compare $\}$

Allora X è una variabile aleatoria con la seguente distribuzione di probabilità, detta, per ovvie ragioni, distribuzione binomiale di parametri n ed a.

x_i	0	1	2	3	... n-1	n
p_i	b^n	$\binom{n}{1}a\cdot b^{n-1}$	$\binom{n}{2}a^2 b^{n-2}$	$\binom{n}{3}a^3 b^{n-3}$	$\dots\binom{n}{n-1}a^{n-1}b$	a^n

Il suo valor medio è

$$\mu = 0.b^n + 1.\binom{n}{1}a\,b^{n-1} + 2\binom{n}{2}a^2b^{n-2} + \dots + (n-1)\binom{n}{n-1}a^{n-1}b + n\,a^n \implies$$

$$\mu = na$$

La sua varianza è $\sigma^2 = n\,a\,b$ e la deviazione standard è

$\sigma = \sqrt{nab}$.

Ad esempio se si lancia un dado 180 volte e si indica con A l'evento "esce il 3" e con B l'evento "non esce il 3" avremo $a = \frac{1}{6}$ e $b = \frac{5}{6}$. Perciò il numero dei 3 che ci si aspetta è $180.\frac{1}{6} = 30$.

La deviazione standard è $\sqrt{180.\frac{1}{6}.\frac{5}{6}} = 5$.

E' chiaro che se avessimo considerato la variabile aleatoria $Y : S \to \mathbb{R}$ tale che $Y(S) = \{0,1,2,\dots,n\}$

$Y^{-1}(0) = \{$insieme dei risultati in cui B non compare$\}$

$= \{$insieme dei risultati in cui A compare n volte$\}$

$= X^{-1}(n)$

$Y^{-1}(1) = \{$insieme dei risultati in cui B compare 1 volta$\}$

$= \{$insieme dei risultati in cui A compare n-1 volte$\}$

$= X^{-1}(n-1)$

etc., cioè $Y^{-1}(i) = X^{-1}(n-i)$ avremmo ottenuto la distribuzione

x_i	0	1	...	n-1	n
p_i	a^n	$\binom{n}{1}a^{n-1}b$	...	$\binom{n}{n-1}a\,b^{n-1}$	b^n

il cui valor medio è $\mu_y = nb$ mentre la varianza e l'errore quadratico medio rimangono gli stessi di quelli trovati per X.

<u>Esempio 3</u> - La probabilità che una macchina produca un pezzo difettoso è del 10%. Si determini la probabilità che in un insieme di 3 pezzi ci siano $m = 0,1,2,3$ pezzi difettosi. Si determini poi il valor medio e lo scarto quadratico medio della variabile aleatoria X definita relativamente a questo problema seguendo i criteri dell'esempio 2 .

Avremo (cfr. esempio 2) $a = 0{,}9$ e $b = 0{,}1$ $n = 3$.

Da cui

$$p_0 = \binom{n}{0}a^3 = 0{,}729$$

$$P_1 = \binom{n}{1}a^2 b = 0{,}243$$

$$P_2 = \binom{n}{2}a\,b^2 = 0{,}027$$

$$P_3 = \binom{n}{n}b^3 = 0{,}001$$

La distribuzione della variabile aleatoria X è dunque

x_i	0	1	2	3
p_i	0,729	0,243	0,027	0,001

Il suo valor medio è

$$\mu = 1\cdot 0{,}243 + 2\cdot 0{,}027 + 3\cdot 0{,}001 = 0{,}3 = 3\cdot 0{,}1$$

e l'errore quadratico medio è

$$\sqrt{0{,}729(0{,}3)^2 + 0{,}243(1-0{,}3)^2 + 0{,}027(2-0{,}3)^2 + 0{,}001(3-0{,}3)^2} =$$

$$\sqrt{0{,}06561 + 0{,}11907 + 0{,}07803 + 0{,}00729} = \sqrt{0{,}27} =$$

$$= \sqrt{3\cdot 0{,}1\cdot 0{,}9} \cong 0{,}5196$$

Per calcolare la varianza di una variabile aleatoria X è utile spesso tenere presente la seguente formula

$$\boxed{\sigma^2 = \sum_i p_i x_i^2 - \mu^2}$$

che si dimostra facilmente osservando che

$$\sigma^2 = \sum_i p_i (x_i - \mu)^2 = \sum_i p_i x_i^2 - 2\mu \sum_i p_i x_i + \mu^2 \sum_i p_i$$

$$= \sum_i p_i x_i^2 - 2\mu^2 + \mu^2 = \sum_i p_i x_i^2 - \mu^2$$

Il numero μ è detto spesso centro di dispersione della variabile aleatoria X.

Se, a partire da X, consideriamo la variabile aleatoria $Y = X - \mu_x$, i cui valori sono cioè $y_i = x_i - \mu_x$, avremo che la sua speranza matematica o valor medio è 0. Infatti sarà

$$\mu_y = \sum_i p_i (x_i - \mu_x) = \sum_i p_i x_i - \mu_x = \mu_x - \mu_x = 0 .$$

Si dice in tal caso che Y è una variabile aleatoria centrata o che Y è uno scarto.

Abbiamo visto che se eseguiamo n volte un esperimento con due soli risultati possibili, A e B, di probabilità a e b = 1 - a, la speranza matematica del numero delle realizzazioni di A è eguale al prodotto del numero delle prove, n, per la probabilità di A cioè

$$\boxed{\mu = na}$$

Se in tale formula si conoscono μ e a si può determinare n, cioè quante prove occorrono per avere la speranza matematica richiesta del numero delle realizzazioni di A.

Se la variabile aleatoria X assume, invece che un numero finito di valori, un'infinità numerabile di valori x_1, x_2, x_3, ..., x_n, ... allora per definire il suo valor medio occorre fare ricorso ad una serie

$$\boxed{\mu = \sum_{1}^{\infty} {}_i p_i x_i}$$

Noi consideriamo solo casi in cui tale serie è convergente.

Esempio 1 - (La distribuzione di Poisson). Sia $X : S \to \mathbb{R}$ tale che $X(S) = \{0,1,2,3,\ldots,n,\ldots\}$ e supponiamo che esista una costante m tale che la probabilità che X assuma valore k sia

$$\boxed{p_k = \frac{m^k e^{-m}}{k!}}$$

m si dice parametro della distribuzione di Poisson

Si ha cioè il seguente quadro

x_i	0	1	2	3	4	...
p_i	e^{-m}	$m^1 e^{-m}$	$\frac{m^2 e^{-m}}{2}$	$\frac{m^3 e^{-m}}{3!}$	$\frac{m^4 e^{-m}}{4!}$	...

Tale distribuzione di probabilità è detta distribuzione di Poisson. Non rientra negli scopi del nostro corso fa vedere che il valor medio di una variabile aleatoria X che segna la distribuzione di Poisson è m cioè

$$m = p_1 x_1 + p_2 x_2 + \dots + p_n x_n + \dots = \sum_{0}^{\infty}{}_k \; k \frac{m^k e^{-m}}{k!}$$

La varianza è ancora $\sigma^2 = m$ e lo scarto quadratico medio è $\sqrt{m}$. La distribuzione di Poisson si presenta in molti fenomeni naturali come il numero di chiamate telefoniche che un centralino riceve ogni minuto, il numero di errori per pagine in un testo molto lungo, il numero di particelle α emesse da una sostanza radioattiva in intervalli eguali di tempo, il numero di piante che si riscontrano in un rettangolo dopo aver diviso una regione vasta in tanti rettangoli eguali etc. ... .

Inoltre la distribuzione di Poisson approssima abbastanza bene la distribuzione binomiale per valori di n grandi e a patto che a sia molto piccolo, scegliendo $m = na$. La seguente tabella mostra questo fatto con $n = 300$, $a = \frac{1}{200}$ e quindi $m = 3/2$

	0	1	2	3	4	5
d. binomiale	0,222	0,335	0,251	0,125	0,046	0,013
d. Poisson	0,223	0,334	0,251	0,125	0,047	0,014

I due esempi che seguono mostrano come poter usare la distribuzione di Poisson.

Esempio 2 - Supponiamo che in un libro di 500 pagine ci siano 300 errori. Determinare la probabilità che in una pagina ci siano esattamente 3 errori.

Consideriamo le pagine come altrettanti rettangoli disposti uno accanto all'altro per terra e ogni errore come una palla che cadendo colpisce uno dei rettangoli. La probabilità che una pagina sia colpita è, per ogni tiro, $\frac{1}{500}$. Dunque $a = \frac{1}{500}$. Il numero dei tiri è 300. Dunque $n = 300$.

Usando la distribuzione binomiale avremmo che

$$p(x = 3) = \binom{300}{3}\left(\frac{1}{500}\right)^3\left(1-\frac{1}{500}\right)^{497} = 0,013$$

Usando la distribuzione di Poisson ($m = \frac{3}{5}$) si ha

$$p(x = 3) = \frac{1}{6}\left(\frac{3}{5}\right)^3 e^{-\frac{3}{5}} = 0,019$$

<u>Esempio 3</u> - Supponiamo che il 2% delle analisi che escono da un laboratorio siano sbagliate. Determinare la probabilità che su un campione di 100 analisi ce ne siano 3 sbagliate.

Avremo $n = 100$ e $a = 0,02$.

Usando la distribuzione binomiale si ottiene

$$p(x = 3) = \binom{100}{3}(0,02)^3(0,98)^{97} = 0,182$$

Usando la distribuzione di Poisson ($m = 2$) si ha

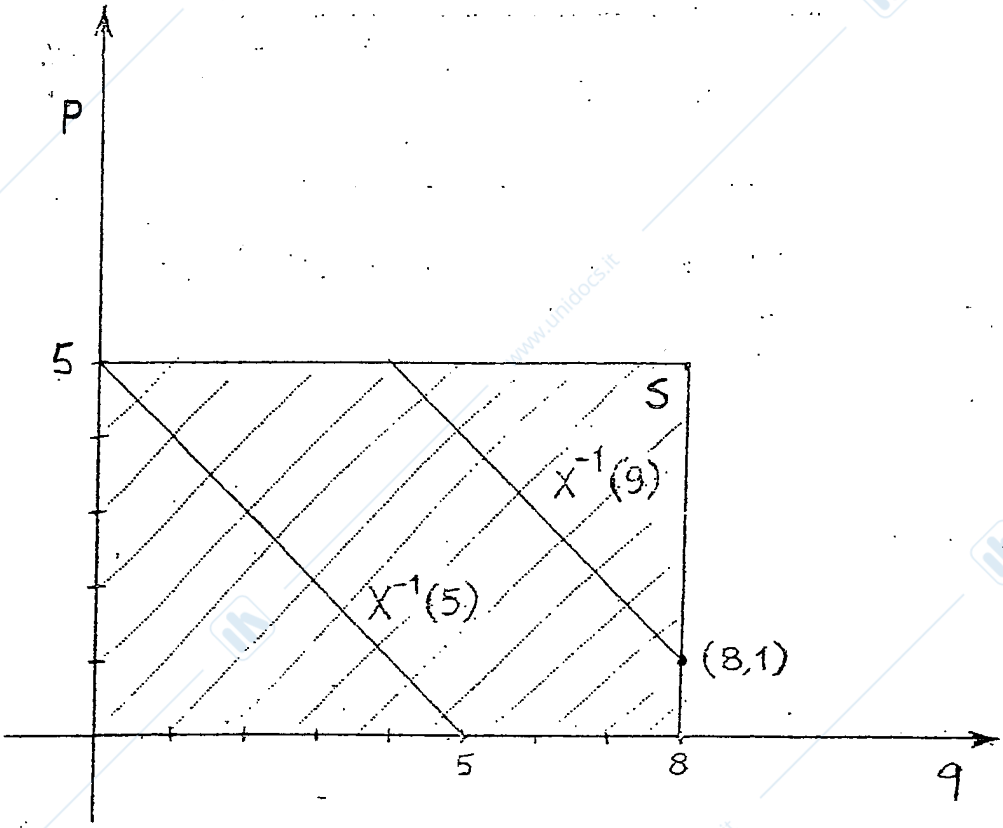
$$p(x = 3) = \frac{1}{6}\, 2^3 e^{-2} \cong 0,180.$$

- <u>Variabili aleatorie continue</u>.

Supponiamo che S sia più che numerabile e che la variabile aleatoria $X : S \to \mathbb{R}$ prenda tutti i valori di un intervallo (a,b) non escludendo che (a,b) sia tutto $\mathbb{R}$ cioè $(-\infty, +\infty)$. Diremo <u>continua</u> tale variabile aleatoria. dove $X(S) = (a, b)$

ossia $X^{-1}(a, b) = S$

<u>Esempio</u> - Supponiamo che p, q soddisfino alle due limitazioni $0 \leq p \leq 5$, $0 \leq q \leq 8$ e sia $S = \{(p,q) \in \mathbb{R}^2 : 0 \leq p \leq 5, 0 \leq q \leq 8\}$. Consideriamo la variabile aleatoria $X : S \to \mathbb{R}$ così definita $X(p,q) = p + q$. E' chiaro che $X(S) = [0,13]$ e $X^{-1}(k) = \{(p,q) \in S : p + q = k\}$. Geometricamente



Supponiamo che esista una funzione non negativa $f:(a,b) \to \mathbb{R}$ tale che per ogni intervallo $[\alpha,\beta] \subset (a,b)$ si abbia

$$\boxed{p(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x)dx}$$

dove $p(\alpha < X < \beta)$ è la probabilità che X assuma valori compresi fra α e β o, in altre parole, la probabilità di $X^{-1}(\alpha,\beta)$.

La funzione f sarà chiamata allora <u>densità di probabilità</u> di X.

Essendo

$$p(X = \alpha) = \int_{\alpha}^{\alpha} f(x)dx = 0 \quad e$$

$$p(X = \beta) = \int_{\beta}^{\beta} f(x)dx = 0$$

avremo anche

$$\boxed{p(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} f(x)dx}$$

in quanto $p(\alpha \leq X \leq \beta) = p(X = \alpha) + p(\alpha < X < \beta) + p(X = \beta)$.
Quindi l'integrale

$$\int_{\alpha}^{\beta} f(x)dx$$

rappresenta anche la probabilità di $X^{-1}[\alpha, \beta]$.

Si osservi poi che dovrà essere

$$\int_{a}^{b} f(x)dx = 1$$

in quanto tale integrale esprime la probabilità di $X^{-1}(a,b) = S$.

La funzione f è definita solo in (a,b) ma noi possiamo considerarla definita su tutto $\mathbb{R}$ prendendola 0 al di fuori di (a,b). Con tale convenzione possiamo scrivere

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

La <u>speranza matematica</u> o valor medio della variabile aleatoria X sarà definita, in analogia a quanto fatto per le variabili a leatorie discrete, dall'integrale

$$\mu = \int_{-\infty}^{+\infty} x\, f(x)dx$$

Se $\mu = 0$ la variabile aleatoria X si dice centrata.
La varianza e l'errore quadratico medio saranno, rispettivamente,

$$\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,dx$$

$$\sigma = \left[\int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,dx\right]^{\frac{1}{2}}$$

La funzione

$$F(x) = \int_{-\infty}^{x} f(s)\,ds = p(-\infty < X \leqslant x)$$

è detta funzione di distribuzione o semplicemente distribuzione.
Essendo $F'(x) = f(x) \geqslant 0$ tale funzione è crescente.
Inoltre

$$\lim_{x \to +\infty} F(x) = \lim_{x \to +\infty} \int_{-\infty}^{x} f(s)\,ds = 1.$$

Due variabili aleatorie continue possono avere la stessa densità e quindi anche la stessa distribuzione. Per questo è opportuno un esame delle densità e relative distribuzioni che si presentano più frequenti nelle applicazioni.

Distribuzione uniforme.

Se la densità $f : \mathbb{R} \to \mathbb{R}$ è del tipo

$$f(x) = \begin{cases} 0 & \text{se } x < a \\ k > 0 & \text{se } a \leqslant x \leqslant b \\ 0 & \text{se } x > b \end{cases}$$

allora la distribuzione $F(x)$ è detta uniforme.

Dovendo essere

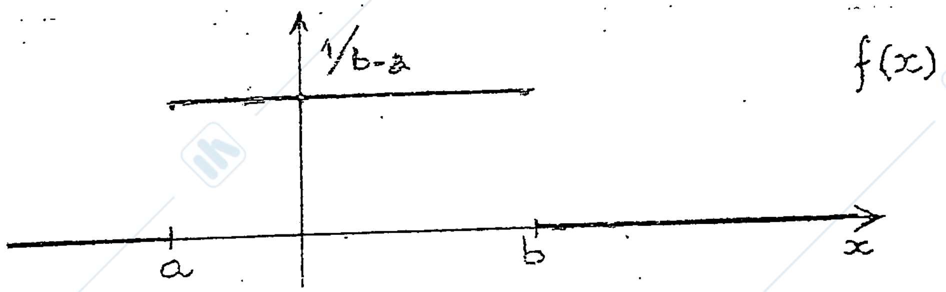
$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_a^b f(x)\,dx = 1$$

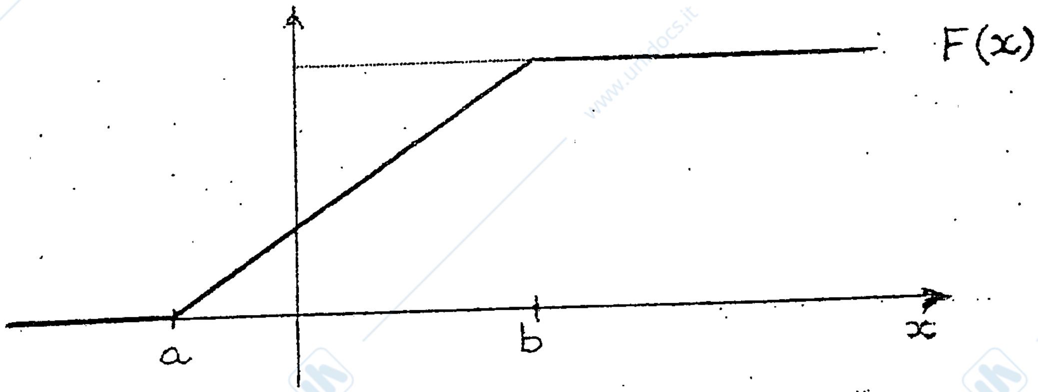
si deduce che $k = \frac{1}{b-a}$

Avremo quindi

$$F(x) = \begin{cases} 0 & \text{se} \quad x < a \\ \int_a^x \frac{1}{b-a}\,dx = \frac{x-a}{b-a} & \text{se} \quad a \leqslant x \leqslant b \\ 1 & \text{se} \quad x > b \end{cases}$$

I grafici di f e F si presentano così





Per una variabile aleatoria X con distribuzione uniforme si hanno i seguenti dati

Valore medio	$\mu = \frac{b+a}{2}$
varianza	$\sigma^2 = \frac{(b-a)^2}{12}$
scarto quadratico medio	$\sigma = \frac{b-a}{2\sqrt{3}}$

Variabili aleatorie con distribuzione uniforme si incontrano di frequente nelle misure di grandezze, nei problemi di orientazione degli animali e in tutti quei problemi il cui risultato appartiene ad un certo intervallo $[a,b]$ senza che si sia capaci di dare una preferenza a certi punti dell'intervallo rispetto ad altri.

Distribuzione normale (o di Gauss).

E' detta normale una distribuzione F di densità

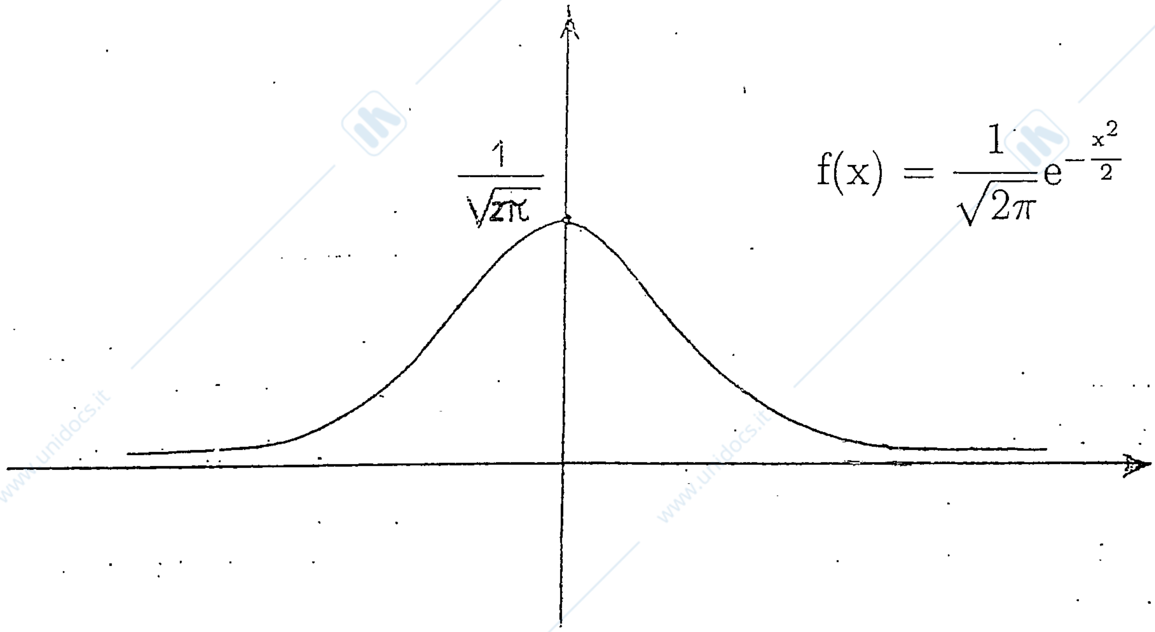
$$(1) \qquad f(x) = \frac{h}{\sqrt{\pi}} e^{-h^2(x-m)^2}, \quad h > 0.$$

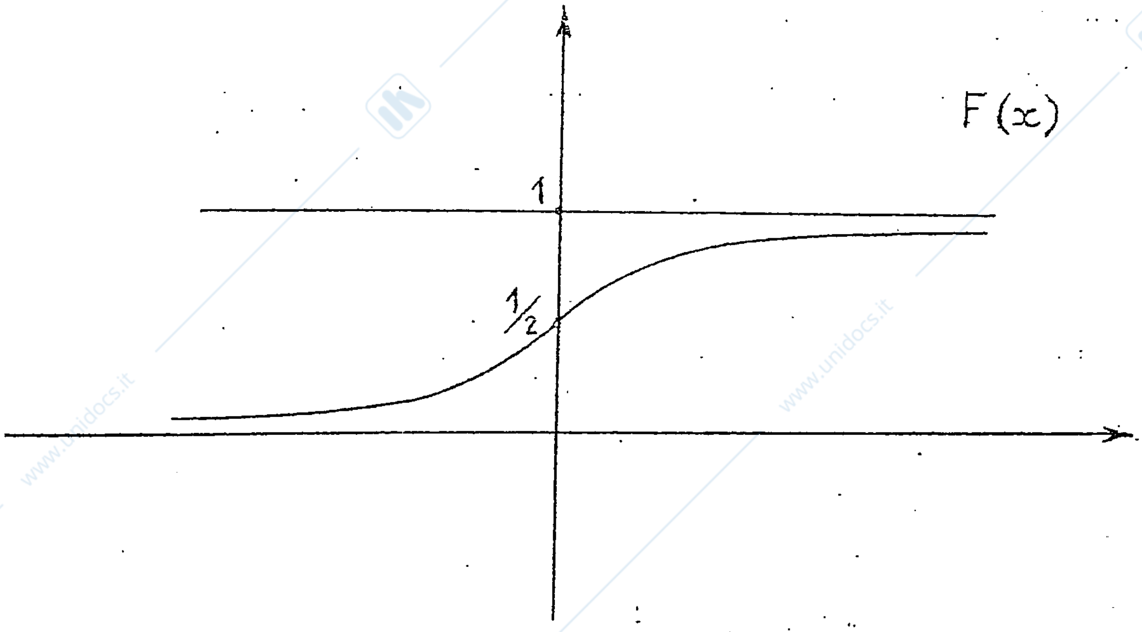
Si può dimostrare che il valore medio, la varianza e lo scarto quadratico medio di una variabile aleatoria con distribuzione normale sono, rispettivamente,

$$\boxed{\begin{aligned} \mu &= m \\ \sigma^2 &= \frac{1}{2} h^{-2} \\ \sigma &= \frac{1}{\sqrt{2}} h^{-1} \end{aligned}}$$

Casi particolari, ma abbastanza interessanti nelle applicazioni, sono quelli con $m = 0$ (distribuzione centrata) e quelli con $m = 0$ e $\sigma = 1$, da cui $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

In quest'ultimo caso i grafici di f(x) e di F(x) si presentano come in figura -

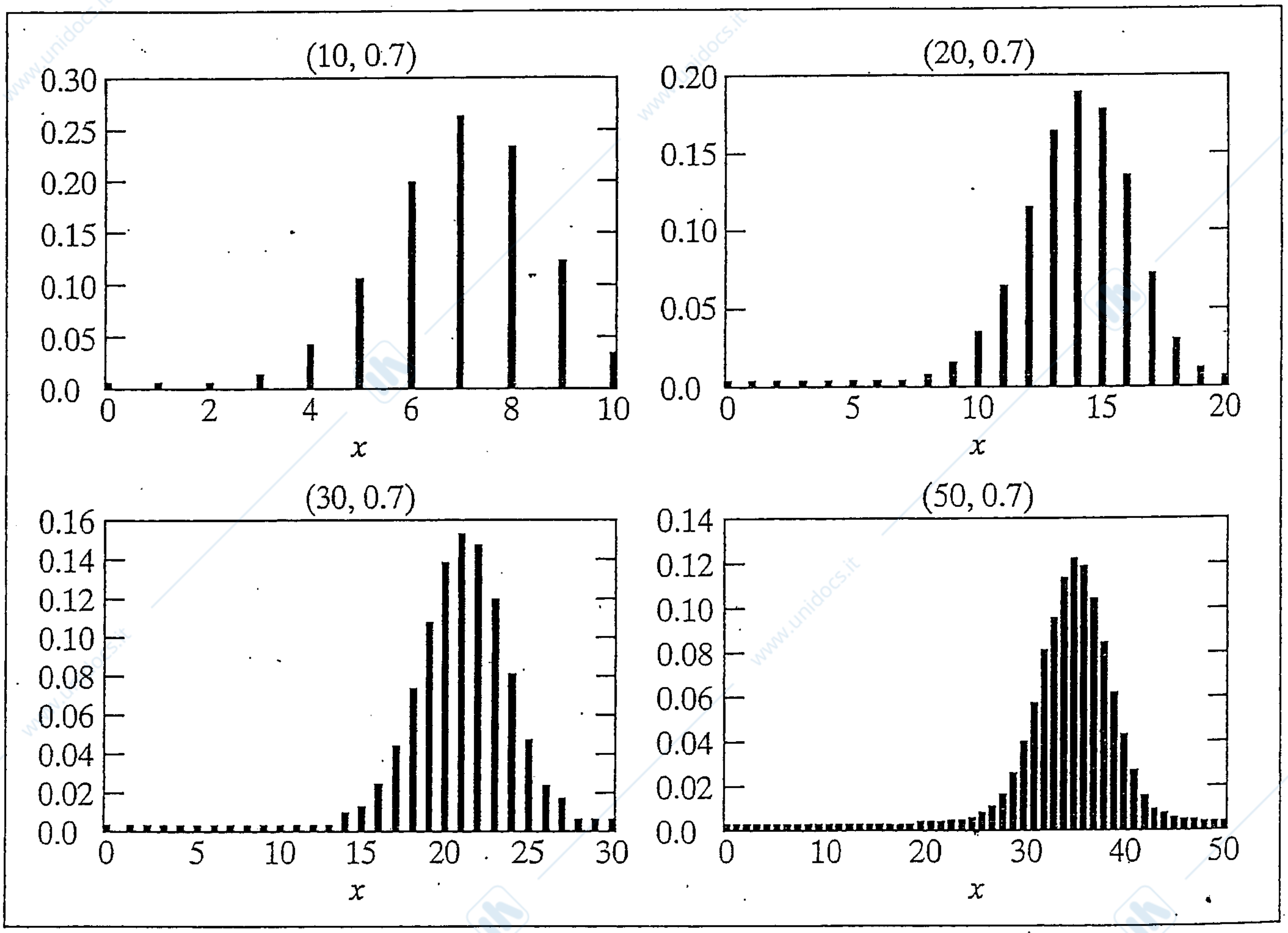




La distribuzione normale ha un'importanza fondamentale nelle scienze applicate perchè le variabili aleatorie continue con distribuzione normale o tali che la loro distribuzione può essere approssimata con successo da quella normale si presentano in moltissimi fenomeni. Inoltre anche variabili aleatorie discrete con distribuzione binomiale possono essere approssimate da variabili aleatorie continue con distribuzioni normale.

Il Teorema di De Moivre-Laplace

Un risultato importante della teoria della probabilità, noto come Teorema di De Moivre-Laplace, afferma che se n è grande, una distribuzione binomiale di parametri n ed a è ben approssimata da una distribuzione normale con la stessa media na e la stessa varianza na(1-a).



In figura a = 0.7 mentre n assume successivamente i valori 10, 20, 30, 50.

La densità discreta di una variabile aleatoria binomiale diventa sempre più "normale" al crescere di n.
(da Sheldon M. Ross, *Calcolo delle probabilità*).

Facciamo presente ora una difficoltà che si presenta nel valutare la probabilità che una variabile aleatoria continua con distribuzione normale assuma valori compresi fra α e β, nel valutare cioè

$$p(\alpha \leq X \leq \beta) = \frac{h}{\sqrt{\pi}} \int_{\alpha}^{\beta} e^{-h^2(x-m)^2} dx$$

La difficoltà deriva dal fatto che non è possibile determinare una primitiva di $e^{-h^2(x-m)^2}$ in termini di funzioni elementari. Si ricorre allora ad una funzione ausiliaria, detta <u>funzione di probabilità</u>, che si trova <u>tabulata</u> in appositi libri. Essa è così definita

$$(2) \qquad \Phi(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-s^2} ds$$

Si può dimostrare che

$$\int_0^{+\infty} e^{-s^2} ds = \frac{\sqrt{\pi}}{2}$$

Perciò $\lim\limits_{x \to +\infty} \Phi(x) = 1.$

Con l'aiuto di tale funzione si verifica che

$$(3) \qquad \boxed{p(\alpha \leq X \leq \beta) = \frac{1}{2}\left[\Phi(h(\beta-m)) - \Phi(h(\alpha-m))\right]}$$

Si osservi che la funzione Φ è dispari.

Se quindi $\alpha - m = -(\beta - m)$ cioè

$$\boxed{m = \frac{\beta+\alpha}{2}}$$

allora si ha

$$\boxed{p(\alpha \leq X \leq \beta) = \Phi(h(\beta - m))}$$

Un caso particolare è quando si ha $m = 0$, $\alpha = -\beta$ e $h = 1$. Si ottiene allora

$$p(-\beta \leq X \leq \beta) = \Phi(\beta)$$

Quel valore di β per cui $\Phi(\beta) = \frac{1}{2}$ è detto scarto (o errore) probabile ed è indicato con E.
Dalle tavole si ricava

$$E \cong 0,4769$$

Una variabile aleatoria con distribuzione normale di

valor medio $\mu = m = 0$ (cioè centrata)

e di varianza $\sigma^2 = \frac{1}{2} h^{-2} = \frac{1}{2}$ (cioè $h = 1$)

ha tanta probabilità di cadere nell'intervallo $[-E, E]$ quanta ne ha di cadere fuori di tale intervallo.

Un altro caso degno di attenzione è quello in cui $m = 0$. Si osservi che allora a parità di intervallo $[-\beta, \beta]$ la probabilità che una variabile aleatoria vi cada dentro è molto vicina a 1 per valori di h molto grandi. Si verifica infatti che

$$p(-\beta \leq X \leq \beta) = \Phi(h\beta) = \frac{2}{\sqrt{\pi}} \int_0^{h\beta} e^{-s^2} ds$$

$$= 1 - \frac{2}{\sqrt{\pi}} \int_{h\beta}^{+\infty} e^{-s^2} ds$$

e siccome e^{-s^2} decresce molto rapidamente si vede che quando h è abbastanza grande il valore dell'integrale che si sottrae è molto piccolo.

Se si pensa ad X come alla variabile aleatoria centrata che esprime gli scarti probabili di una serie di misure effettuate rispetto alla misura vera si vede che al crescere di h la probabilità che tali scarti siano compresi in un intervallo piccolo ($[-\beta, \beta]$) aumenta. Per tale ragione Gauss chiamò h misura o parametro di precisione.

Come abbiamo accennato all'inizio la distribuzione normale approssima quella binomiale. Tale approssimazione è tanto migliore quanto più è grande il numero, n, delle prove eseguite (e quanto più $a, b = 1-a$ si avvicinano a $\frac{1}{2}$).

La distribuzione normale da scegliere, una volta noti n, a (e $b = 1-a$) è quella di densità

$$f(x) = \frac{1}{\sqrt{2\pi n\,a\,b}}\; e^{-\frac{1}{2nab}(x-na)^2}$$

cioè $h = \frac{1}{\sqrt{2nab}}$ e $m = na$.

Ricordando che nella distribuzione binomiale si ha

$$\boxed{\begin{aligned} \mu &= na \\ \sigma^2 &= nab \end{aligned}}$$

e ricordando i significati di h e m nel caso di quella normale si vede che la distribuzione normale da scegliere è quella che ha valor medio e varianza eguali alla distribuzione binomiale che si vuole approssimare.

Esempio 1 - La probabilità che una macchina fornisca una analisi imperfetta è 0,01.

Si vuole sapere qual'è la probabilità che su 1000 analisi quelle difettose siano non più di 20.

Potremmo usare la distribuzione binomiale ragionando così.
L' evento A è "analisi difettosa" e l' evento B è "analisi non difettosa". Si ha $p(A) = a = 0{,}01$, $p(B) = b = 0{,}99$.
Le prove sono $n = 1000$. Allora la probabilità che su 1000 prove A si verifichi $0,1,2,\dots,20$ volte è

$$p(X = 0) = \binom{1000}{0}(0{,}01)^0 \, (0{,}99)^{1000}$$

$$p(X = 1) = \binom{1000}{1}(0{,}01) \, (0{,}99)^{999}$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

$$p(X = 20) = \binom{1000}{20}(0{,}01)^{20} \, (0{,}99)^{980} \, .$$

Perciò $p(0 \leqslant X \leqslant 20) = \sum_{0}^{20}{}_k \binom{1000}{k}(0{,}01)^k \, (0{,}99)^{1000-k}$

e si vede che tale conto non è per nulla semplice.
Approssimando la distribuzione binomiale con quella normale di valor medio $\mu = na = 1000 \cdot 0{,}01 = 10$ e scarto quadratico medio

$$\sigma = \sqrt{n \cdot a \cdot b} = \sqrt{1000 \cdot 0{,}01 \cdot 0{,}99} = \sqrt{9{,}9}$$

avremo, ricordando che $h = \frac{1}{\sigma\sqrt{2}}$ e $m = na$, e usando la formula (3)

$$p(0 \leqslant X \leqslant 20) = \frac{1}{2}\left[\phi\left(\frac{20-10}{\sqrt{2}\sqrt{9{,}9}}\right) - \phi\left(\frac{0-10}{\sqrt{2}\sqrt{9{,}9}}\right)\right] \cong \frac{1}{2}\left[\phi(2{,}25) + \phi(2{,}25)\right]$$

$= \phi(2{,}25)$ (si ricordi che ϕ è dispari).

Dalle tavole si ricava che

$$\phi(2{,}25) \cong 0{,}9985 \, .$$

Perciò la probabilità che le analisi imperfette non siano più di 20 è 0,9985.

Valori della funzione

$$\Phi(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-s^2} ds$$

x	Φ(x)	Δ	x	Φ(x)	Δ
0,00	0,0000	564	1,50	0,9661	55
0,05	0,0564	561	1,55	0,9716	47
0,10	0,1125	555	1,60	0,9736	41
0,15	0,1680	547	1,65	0,9804	34
0,20	0,2227	536	1,70	0,9838	29
0,25	0,2763	523	1,75	0,9867	24
0,30	0,3286	508	1,80	0,9891	20
0,35	0,3794	490	1,85	0,9911	17
0,40	0,4284	471	1,90	0,9928	14
0,45	0,4755	450	1,95	0,9942	11
0,50	0,5205	428	2,00	0,9953	10
0,55	0,5633	406	2,05	0,9963	7
0,60	0,6039	381	2,10	0,9970	6
0,65	0,6420	358	2,15	0,9976	5
0,70	0,6778	334	2,20	0,9981	4
0,75	0,7112	309	2,25	0,9985	3
0,80	0,7421	286	2,30	0,9988	3
0,85	0,7707	262	2,35	0,9991	2
0,90	0,7969	240	2,40	0,9993	2
0,95	0,8209	218	2,45	0,9995	1
1,00	0,8427	197	2,50	0,9996	1
1,05	0,8624	178	2,55	0,9997	1
1,10	0,8802	159	2,60	0,9998	0
1,15	0,8961	142	2,65	0,9998	1
1,20	0,9103	126	2,70	0,9999	0
1,25	0,9229	111			
1,30	0,9340	98			
1,35	0,9438	85			
1,40	0,9523	74			
1,45	0,9597	64	3,00	1	

Da L. Barletti, *Calcolo delle Probabilità e Statistica*

Applicazioni del Teorema di Bayes

TEST DIAGNOSTICI

Una delle applicazioni più utili del calcolo delle probabilità in campo medico è quella relativa ai *test diagnostici.* Un test si dice *positivo* quando rileva (o meglio sembra rilevare) la presenza di una malattia; si dice negativo quando rileva l'assenza di una malattia (o meglio sembra escluderla). Un test ideale dovrebbe essere positivo solo in presenza di un paziente malato e negativo solo in presenza di un paziente sano. In realtà questo non avviene, ma si possono presentare *quattro situazioni distinte*. Si dice che un soggetto è un

1) *vero-positivo* quando il paziente è malato in presenza di un test positivo.
2) *vero-negativo* quando il paziente è sano in presenza di un test negativo.
3) *falso-positivo* quando il paziente è sano in presenza di un test positivo.
4) *falso-negativo* quando il paziente è malato in presenza di un test negativo.

Nei casi 1) e 2) il test è corretto, nei casi 3) e 4) il test è errato. Un buon test dovrebbe avere una bassa probabilità di errore. Nei casi 3) e 4) gli errori si possono misurare mediante i *tassi di errore*, in termini di probabilità condizionate, ossia i seguenti parametri:

I) Il *tasso di falso-positivo*
$\alpha = P\,(test\ positivo|\ paziente\ sano) = P\,(B|A^c)$

II) Il *tasso di falso-negativo*
$\beta = P\,(test\ negativo|\ paziente\ malato) = P\,(B^c|A)$
Ovviamente affinché il test sia attendibile α e β devono essere piccoli.

III) *Specificità del test*
$1 - \alpha = P\,(test\ negativo|\ paziente\ sano) = P\,(B^c|A^c)$

IV) *Sensibilità del test*
$1 - \beta = P\,(test\ positivo|\ paziente\ malato) = P\,(B|A)$

Si dice infine *accuratezza* del test la probabilità che esso fornisca un risultato esatto qualunque sia lo stato del paziente (casi 1) e 2)). Abbiamo che
$accuratezza = 1 - \alpha\,P\,(A^c) - \beta\,P\,(A)$
Notiamo che, mentre la specificità e la sensibilità sono caratteristiche proprie del test, indipendenti dalla popolazione in esame, l'accuratezza dipende dalla popolazione.

In una popolazione scegliamo ora un individuo a caso, lo sottoponiamo al test e supponiamo che questo risulti *positivo.* Indichiamo con A, A^c e B i seguenti eventi:

A = " l'individuo è malato"
A^c = " l'individuo è sano"
B = " il test è positivo"

Vogliamo sapere qual' è la probabilità che il paziente sia effettivamente malato (in presenza di un test positivo) ossia vogliamo determinare la
$P\,(A|B)$= probabilità di avere un *vero-positivo*
Troveremo poi la probabilità che sia sano (anche se il test è positivo) ossia la
$P\,(A^c|B)$= probabilità di avere un *falso-positivo*

I test diagnostici sono un campo di applicabilità del *Teorema di Bayes* in relazione agli eventi A, A^c e B. Dunque

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A)+P(A^c)P(B|A^c)} = \frac{(1-\beta)P(A)}{(1-\beta)P(A)+\alpha P(A^c)} \quad (*)$$

$$P(A^c|B) = 1 - P(A|B)$$

Da queste formule risulta che, se l'incidenza della malattia è bassa (cioè $P(A)$ è piccola: malattia rara), anche un test efficiente (cioè con α e β piccoli) dà una risposta apparentemente paradossale, perché $P(A|B)$ è bassa, mentre $P(A^c|B)$ è alta. Ossia, dopo un test positivo, risulterebbe più probabile essere sani che malati.

Esempio:
Sottoponiamo un paziente ad un test per sapere se ha una malattia. Sappiamo che la malattia (rara) ha un'incidenza dello 0,2%, ossia la probabilità che il paziente abbia la malattia è
$P(A) = 0{,}002$ (Dunque la probabilità che sia sano è $P(A^c) = 1 - 0{,}002 = 0{,}998$)
Sappiamo che il test dà un risultato positivo corretto (ossia in presenza di malattia) del 99% e un risultato negativo corretto (ossia in assenza di malattia) del 93%, cioè
$1 - \beta = P(test\ positivo|\ paziente\ malato) = P(B|A) = 0{,}99$ (Dunque $\beta = 1 - 0{,}99 = 0{,}01$)
$1 - \alpha = P(test\ negativo|\ paziente\ sano) = P(B^c|A^c) = 0{,}93$ (Dunque $\alpha = 1 - 0{,}93 = 0{,}07$)
Quindi

$$P(A|B) = \frac{0{,}99\times0{,}002}{0{,}99\times0{,}002+0{,}07\times0{,}998} \approx 0{,}028 \quad \text{mentre} \quad P(A^c|B) = 1 - P(A|B) \approx 0{,}97$$

Così, in presenza di un test positivo, la probabilità che il paziente sia malato è circa 0,028, mentre la probabilità che sia sano è circa 0,97.
Tale conclusione è solo in apparenza paradossale. Infatti la situazione che è stata descritta è quella che si presenterebbe in un ipotetico "screening" casuale della popolazione su base, ad esempio, nazionale. Nella pratica medica corrente in realtà il medico decide di sottoporre il paziente ad un test solo quando sospetta che siano presenti le condizioni "favorevoli" alla malattia (popolazioni a rischio); in questo caso la probabilità *a priori* che il paziente sia malato, cioè la $P(A)$, è molto più alta rispetto all'incidenza della malattia su base nazionale, e di conseguenza la probabilità *a posteriori* che il paziente sia malato dopo un test positivo, cioè la $P(A|B)$, viene più elevata. Notiamo che, in ogni caso, in presenza di un test positivo, la probabilità a posteriori $P(A|B)$ risulta più grande della probabilità a priori $P(A)$ (nell'esempio $P(A|B)$ è 14 volte superiore alla $P(A)$).
Questo suggerisce di ripetere il test. Supponiamo che sia di nuovo positivo. Ora, sempre con la formula (*), calcoliamo la nuova probabilità a posteriori, che chiameremo $P_1(A|B)$, assumendo come nuova probabilità a priori la $P(A|B)$ al posto di $P(A)$ e con $P(A^c|B)$ al posto di $P(A^c)$. Per cui

$$P_1(A|B) = \frac{0{,}99\times0{,}028}{0{,}99\times0{,}028+0{,}07\times0{,}97} \approx 0{,}29 \quad \text{mentre la nuova} \quad P_1(A^c|B) \approx 1 - 0{,}29 \approx 0{,}71$$

Iterando il procedimento avremmo $P_2(A|B) \approx 0{,}85$ per cui
$P(A) < P(A|B) < P_1(A|B) < P_2(A|B)$ e così di seguito.

Da M. Abate, *Matematica e statistica. Le basi per le scienze della vita*

Media, mediana e moda

3.6 Media, mediana e moda

In numerose situazioni sperimentali è necessario riassumere molti dati con un solo numero. Vediamo due esempi tipici.

Esempio 3.49

Supponiamo di voler misurare la lunghezza dell'assone della cellula neurale di un ratto. Ripetiamo la misura dieci volte, ottenendo i seguenti risultati, espressi in micrometri:

$$70.78,\ 74.22,\ 74.03,\ 71.71,\ 70.97,\ 73.47,\ 69.28,\ 69.62,\ 72.31,\ 72.76\,.$$

La differenza fra i vari risultati è dovuta agli (inevitabili) errori di misurazione; il nostro obiettivo è estrarre da queste misure un unico valore per la lunghezza dell'assone, scelto in modo che sia plausibile che i risultati delle nostre misure siano variazioni casuali attorno a questo valore.

Esempio 3.50

Nell'Esempio 2.19 abbiamo misurato il peso di 15 cavie, ottenendo i seguenti pesi in grammi:

$$28,\ 32,\ 37,\ 29,\ 31,\ 30,\ 32,\ 26,\ 32,\ 27,\ 29,\ 30,\ 28,\ 31,\ 31\,.$$

Vorremmo trovare un numero che rappresenti il peso tipico di questa popolazione di cavie.

In entrambi questi casi abbiamo un numero finito di dati $x_1, \ldots, x_n$ dello stesso tipo (molte misure della stessa quantità, sullo stesso campione o su campioni diversi) e vorremmo trovare un numero $\bar{x}$ che riassuma questi dati.

Errore o scarto

Supponiamo di riassumere i nostri dati con il valore $\bar{x}$. Se $x_i = \bar{x}$, il dato i-esimo è perfettamente rappresentato da $\bar{x}$. Ma se $x_i \neq \bar{x}$, il valore $\bar{x}$ rappresenta il dato x_i solo a meno dell'**errore** (o **scarto**) $\bar{x} - x_i$. Quindi è ragionevole pensare che la scelta migliore di $\bar{x}$ sia quella che *minimizza* gli errori $\bar{x} - x_i$.

Ma gli errori sono a priori tanti, uno per ogni dato, e non possiamo pretendere di minimizzarli tutti indipendentemente l'uno dell'altro; dobbiamo cercare di minimizzarli nel loro complesso.

Una prima possibilità consiste nel cercare di minimizzare la somma

$$(\bar{x} - x_1) + \cdots + (\bar{x} - x_n) = \sum_{i=1}^{n} (\bar{x} - x_i)$$

degli errori, o, più precisamente, il suo valore assoluto. Siccome il numero con valore assoluto minimo è 0, vogliamo vedere se esiste un numero $\bar{x} \in \mathbb{R}$ tale che $\sum_{i=1}^{n} (\bar{x} - x_i) = 0$. Ma

$$\begin{aligned}\sum_{i=1}^{n} (\bar{x} - x_i) &= (\bar{x} - x_1) + \cdots + (\bar{x} - x_n) \\ &= n\bar{x} - (x_1 + \cdots + x_n) = n\bar{x} - \sum_{i=1}^{n} x_i ,\end{aligned}$$

e quindi

$$\sum_{i=1}^{n} (\bar{x} - x_i) = 0 \iff \bar{x} = \frac{x_1 + \cdots + x_n}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i .$$

Il numero $\bar{x}$ così ottenuto si chiama **media** (o *media aritmetica* o *valore medio*) dei dati $x_1, \ldots, x_n$:

Media

$$\bar{x} = \text{Media}(x_i) = \frac{1}{n} \sum_{i=1}^{n} x_i .$$

Esempio 3.51

La media delle misure delle lunghezze dell'assone del ratto dell'Esempio 3.49 è

$$\begin{aligned}\bar{x} = \frac{1}{10}\big(&70.78 + 74.22 + 74.03 + 71.71 + 70.97 \\ &+ 73.47 + 69.28 + 69.62 + 72.31 + 72.76\big) = 71.915 \text{ m} .\end{aligned}$$

Esempio 3.52

La media dei pesi delle cavie dell'Esempio 3.50 è

$$\begin{aligned}\bar{x} = \frac{1}{15}\big(&28 + 32 + 37 + 29 + 31 + 30 + 32 + 26 \\ &+ 32 + 27 + 29 + 30 + 28 + 31 + 31\big) = 30.2 \text{ g} .\end{aligned}$$

Gli errori $\bar{x} - x_i$ si suddividono naturalmente in due categorie: quelli positivi (per cui $x_i < \bar{x}$) e quelli negativi (per cui $x_i > \bar{x}$). La media è l'unico numero reale $\bar{x} \in \mathbb{R}$ per cui gli errori negativi bilanciano esattamente quelli positivi.

Osservazione 3.18 Potresti obiettare che in un certo senso stiamo barando: gli errori sono errori indipendentemente dal segno, per cui dovremmo piuttosto cercare di minimizzare quantità del tipo

$$\sum_{i=1}^{n} (\bar{x} - x_i)^2 , \tag{3.5}$$

in cui ogni errore non nullo contribuisce con un termine positivo. Ma puoi stare tranquillo: come vedremo [illegible] la media minimizza anche (3.5).

La media risente della presenza di valori estremi nei dati: un solo dato di valore sensibilmente diverso dagli altri può spostare la media in modo significativo (vedi l'esempio successivo). Per questo motivo la presenza di dati anomali nei risultati di un esperimento va esaminata con attenzione: il valore medio calcolato escludendo questi dati estremi può essere più significativo del valore medio calcolato includendoli, *a meno che* la loro presenza non indichi l'esistenza di un fenomeno imprevisto (e quindi interessante) comparso solo in quelle misure. In ogni caso, più i dati sono distribuiti in modo uguale su entrambi i lati della media (cioè, più la media si avvicina alla mediana; vedi oltre) più la media è rappresentativa dei dati nel loro complesso. Fortunatamente, come discuteremo nella Sezione 8.8 (vedi in particolare l'Osservazione 8.13), variazioni nei dati prodotte da errori casuali soddisfano questa condizione.

Esempio 3.53 Guardando i pesi delle cavie dell'Esempio 3.50 si nota subito la presenza di un valore anomalo, i 37 g di peso della terza cavia. Se escludiamo questo valore, la media dei pesi diventa

$$\bar{x} = \frac{1}{14}(28 + 32 + 29 + 31 + 30 + 32 + 26 + 32 + 27 + 29 + 30 + 28 + 31 + 31) \simeq 29.71 \text{ g} .$$

Quindi escludere il dato anomalo causa una variazione della media pari a $30.2 - 29.71 = 0.49$ g.

Osservazione 3.19 Nel calcolo della media aritmetica, ogni dato viene contato tutte le volte che compare. Per esempio, nel caso delle cavie i dati assumono 8 valori distinti (26, 27, 28, 29, 30, 31, 32, 37) e le ripetizioni sono ammesse. Se raggruppiamo insieme i valori ripetuti, possiamo calcolare la media con la formula

$$\bar{x} = \frac{1 \cdot 26 + 1 \cdot 27 + 2 \cdot 28 + 2 \cdot 29 + 2 \cdot 30 + 3 \cdot 31 + 3 \cdot 32 + 1 \cdot 37}{1 + 1 + 2 + 2 + 2 + 3 + 3 + 1}$$

Media ponderata

Più in generale, se abbiamo un gruppo di dati $x_1, \dots, x_n$ che devono essere *pesati* in maniera diversa (per esempio, perché compaiono con frequenze diverse) si può usare la **media ponderata** (o *pesata*) definita da

$$\bar{x} = \frac{f_1 x_1 + \cdots + f_n x_n}{f_1 + \cdots + f_n} ,$$

dove $f_1, \dots, f_n$ sono numeri positivi, detti *pesi* dei dati $x_1, \dots, x_n$. La media aritmetica usuale corrisponde a prendere $f_1 = \cdots = f_n = 1$, cioè a pesare tutti i dati nello stesso modo.

Curiosità 3.5 I pesi hanno una naturale interpretazione probabilistica. Se indichiamo con

$$F = f_1 + \cdots + f_n = \sum_{i=1}^{n} f_i$$

la somma dei pesi e poniamo $p_i = f_i / F$, allora $p_1 + \cdots + p_n = 1$, e possiamo interpretare p_i come la probabilità di ottenere il valore x_i facendo la nostra misura.

Usando i p_i possiamo calcolare la media ponderata con la formula

$$\bar{x} = \sum_{i=1}^{n} p_i x_i . \tag{3.6}$$

Più in generale, in un esperimento probabilistico di spazio degli eventi $\Omega = \{x_1, \ldots, x_n\}$ composto da un numero finito di numeri reali la formula (3.6) (dove p_i ora è la probabilità dell'evento semplice x_i) fornisce il *valore atteso* (o *media*) dell'esperimento. Per esempio, il valore atteso della somma del lancio di due dadi a sei facce non truccati è

$$\frac{1}{36}\cdot 2+\frac{2}{36}\cdot 3+\frac{3}{36}\cdot 4+\frac{4}{36}\cdot 5+\frac{5}{36}\cdot 6+\frac{6}{36}\cdot 7+\frac{5}{36}\cdot 8+\frac{4}{36}\cdot 9+\frac{3}{36}\cdot 10+\frac{2}{36}\cdot 11+\frac{1}{36}\cdot 12 = 7\,.$$

Ne riparleremo nella Sezione 8.2 (vedi l'Esempio 8.10).

Curiosità 3.6 Nel caso di dati *positivi,* invece degli errori assoluti $x - x_i$ possiamo considerare gli errori relativi x/x_i. In questo caso, l'assenza di errore corrisponde ad avere errore relativo uguale a 1; quindi possiamo chiederci se esiste un $\hat{x} \in \mathbb{R}$ per cui il *prodotto* $\prod_{i=1}^n (\hat{x}/x_i)$ sia esattamente uguale a 1. È facile rispondere: siccome

$$\prod_{i=1}^{n} \frac{\hat{x}}{x_i} = \frac{\hat{x}^n}{\prod_{i=1}^n x_i}\,,$$

il prodotto degli errori relativi è uguale a 1 se e solo se

$$\hat{x} = \left(\prod_{i=1}^{n} x_i\right)^{1/n}.$$

Questo $\hat{x}$ è detto *media geometrica* di $x_1, \ldots, x_n$. Si può dimostrare che la media geometrica è sempre minore o uguale alla media aritmetica, cioè che

$$(x_1 \cdots x_n)^{1/n} \le \frac{x_1 + \cdots + x_n}{n}\,,$$

con uguaglianza se e solo se $x_1 = \cdots = x_n$.

Quando i dati non sono equamente distribuiti attorno alla media, può essere utile riassumerli con un altro valore che soddisfa invece questa condizione: la mediana. Prendiamo i nostri dati e disponiamoli in ordine crescente, in modo da avere

$$x_1 \le x_2 \le \cdots \le x_n\,.$$

Mediana

Allora la **mediana** dei dati è un numero M tale che esattamente metà dei dati è minore o uguale a M e metà è maggiore o uguale a M: per l'esattezza,

$$M = \begin{cases} x_{(n+1)/2} & \text{se } n \text{ è dispari}\,, \\ \frac{1}{2}(x_{n/2} + x_{(n/2)+1}) & \text{se } n \text{ è pari}\,. \end{cases}$$

In altre parole, M è il dato centrale se n è dispari ed è la media aritmetica dei due dati centrali se n è pari. In particolare, la mediana non è influenzata da dati estremi anomali (ma dipende dal numero dei dati).

Esempio 3.54

Per calcolare la mediana dei pesi delle cavie dell'Esempio 3.50 cominciamo disponendoli in ordine crescente:

$$26 \le 27 \le 28 \le 28 \le 29 \le 29 \le 30$$
$$\le 30 \le 31 \le 31 \le 31 \le 32 \le 32 \le 32 \le 37\,.$$

Abbiamo $n = 15$ dati; quindi la mediana M è l'ottavo – $8 = (15+1)/2$ – dato, $M = 30$ g. Nota che la mediana non cambia anche se aumentiamo il valore estremo 37, in quanto l'ordine degli altri dati non viene modificato. Se togliamo il dato estremo 37, rimangono 14 dati, per cui la mediana M diventa la media aritmetica fra il settimo – $7 = 14/2$ – e l'ottavo – $8 = (14/2) + 1$ – dato, cioè $M = (30 + 30)/2 = 30$ g. Se invece togliamo il dato estremo 26 (senza togliere il dato estremo 37), la mediana diventa la media aritmetica fra il settimo e l'ottavo dato *nel nuovo ordine,* per cui diventa $M = (30 + 31)/2 = 30.5$ g.

Quartili e percentili

Osservazione 3.20 La mediana divide i dati in due insiemi con ugual numero di elementi. Nulla vieta di suddividere l'insieme ordinato dei dati in un maggior numero di sottoinsiemi. Per esempio, i valori che dividono i dati in quattro sottoinsiemi con ugual numero di elementi si chiamano **quartili**; in particolare, la mediana è il secondo quartile. Se ci sono moltissimi dati, a volte si usano i **percentili**, che dividono i dati in cento sottoinsiemi con ugual numero di elementi — e la mediana è il cinquantesimo percentile.

Curiosità 3.7 Anche la mediana minimizza un'opportuna somma di errori. Per l'esattezza, si può dimostrare che la mediana M dei dati $x_1, \dots, x_n$ rende minima la somma

$$\sum_{i=1}^{n} |M - x_i|$$

dei valori assoluti degli errori.

Moda o classe modale

Media e mediana servono per riassumere dati numerici con un unico valore. A volte si vorrebbe operare in modo analogo anche con dati non numerici (colori degli occhi, specie di animale eccetera). In tal caso si suddividono i dati in classi; si chiama **moda** (o **classe modale**) la classe (o le classi, se ce ne sono più d'una) con il maggior numero di elementi.

Esempio 3.55

Supponiamo di avere come insieme di dati l'elenco dei maschi italiani morti nel 2002 e suddividiamoli in classi d'età come indicato nella Tabella 2.8. Allora la moda (cioè la classe con il maggior numero di elementi) è la classe 70–79.

Esempio 3.56

Ovviamente, si può calcolare la moda anche di dati numerici: la moda è il dato (o i dati) con frequenza assoluta maggiore. Per esempio, le classi modali dei pesi delle solite cavie dell'Esempio 3.54 sono 31 g e 32 g, in quanto questi sono i pesi che si ripetono più frequentemente.