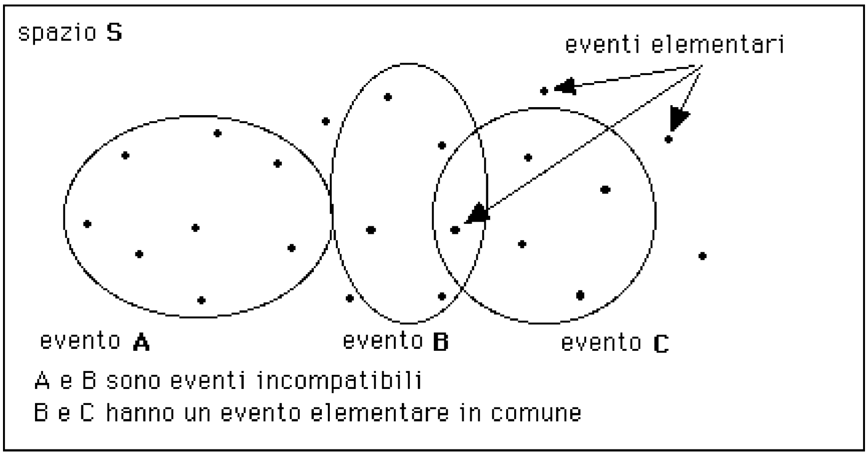
ELEMENTI DI TEORIA DELLE PROBABILITÀ

(come ragionare quantitativamente in condizioni di incertezza)

LO SPAZIO DEGLI EVENTI

Prima di introdurre il concetto di probabilità analizziamo la **definizione di evento**: consideriamo tutte le possibili modalità con cui si può presentare un qualsiasi fenomeno, per esempio i 37 risultati possibili nel gioco della roulette: ognuno di essi può essere considerato un elemento di uno spazio S che li contenga tutti (diagrammi di Venn).

I sottoinsiemi di S (p.es. tutti i numeri pari; tutti i numeri verdi; tutti i numeri superiori a 24) sono eventi composti determinati dal concorso di uno o più eventi elementari (consistenti ciascuno in un solo elemento dello spazio S); p.es. nel gioco della roulette l'evento "dispari" è rappresentato dagli eventi elementari corrispondenti all'uscita dei numeri 1, 3, 5, ..., 35.
Eventi composti possono avere elementi in comune (il numero 28 è pari e è "superiore a 24" e "non verde"); in caso contrario si dicono incompatibili o mutuamente esclusivi.
Analogamente, gli eventi elementari possono appartenere a più eventi composti.

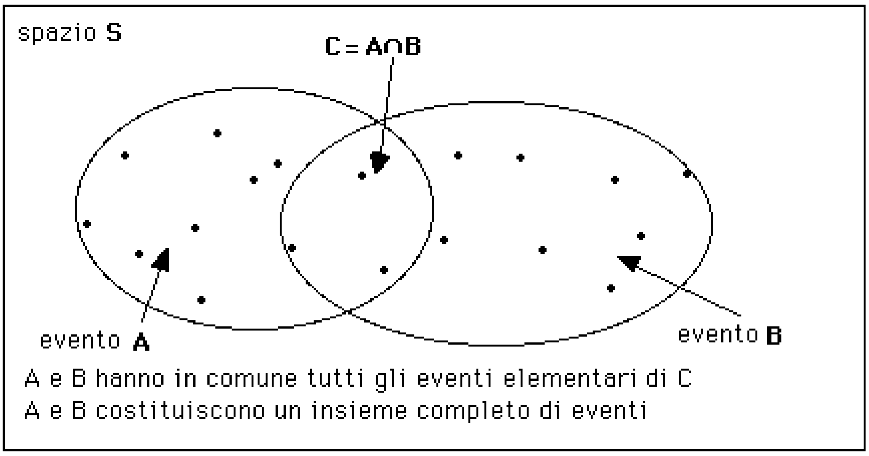


Se un insieme di eventi contiene tutti gli eventi elementari, tale insieme si dice completo.

Lo spazio S può contenere anche un numero infinito di eventi elementari come nel caso in cui questi rappresentino tutti i possibili risultati di una misura; p.es. di una lunghezza, di una massa, di un tempo.

Di tutti i possibili eventi definibili nello spazio S, in questi cenni elementari di teoria delle probabilità, siamo interessati solo alla categoria dei sottoinsiemi disgiunti che definiscono gli eventi incompatibili o mutuamente esclusivi (p.es. "pari" e "dispari" sono eventi mutuamente esclusivi: non hanno nessun elemento in comune; invece "rosso" e "dispari" hanno eventi elementari in comune).

Qualora gli eventi non siano incompatibili è possibile, almeno in linea di principio, ricondurli ad eventi incompatibili. Per chiarire il concetto consideriamo la figura:



Gli eventi A e B non sono incompatibili perché hanno degli eventi elementari in comune: tutti gli eventi appartenenti all'intersezione $C = A \cap B$.
Per riportarli ad eventi incompatibili consideriamo i tre eventi:
$A' = A - C = A - (A \cap B)$
$B' = B - C = A - (A \cap B)$
$C = A \cap B$.

Nel caso di eventi non incompatibili può avere importanza la domanda: qual è la probabilità che si verifichi l'evento A subordinata al fatto che si è verificato l'evento B? Questo tipo di calcolo è leggermente più complesso (probabilità condizionate) e diventa indispensabile per una trattazione avanzata dell'analisi dei dati sperimentali che non affronteremo in questo corso elementare.

EVENTI MUTUAMENTE ESCLUSIVI, INDIPENDENTI, O ...

Va prestata molta attenzione alla definizione di eventi incompatibili e indipendenti:

- i possibili risultati di un lancio di un dado sono eventi incompatibili: o esce l'1 o il 2 o...il 6;

- il risultato di un lancio e il successivo sono indipendenti;

- nel lancio di un dado l'uscita di un numero pari e del numero 3 sono eventi incompatibili;

- l'uscita di un numero pari e del numero 2 non sono né eventi indipendenti (anzi, se esce il 2 sicuramente il risultato è pari) né eventi esclusivi (se esce pari può uscire il numero 2); la trattazione di questo tipo di eventi richiede l'introduzione del concetto di probabilità condizionata.

PROBABILITÀ

Definizione soggettivistica di probabilità (de Finetti)

Grado di fiducia che si ha nel verificarsi di un evento

Quanto credo nel risultato che ho ottenuto? La risposta è solo apparentemente soggettiva: la stima del *degree of belief* è soggettiva ma se equa (quanto si punterebbe in una scommessa in un cui nessuno dei partecipanti vuole perdere?) diventa oggettiva e quindi utilizzabile per ragionare quantitativamente anche in condizioni di incertezza. Chi, infatti, in tutta onestà, giudicherebbe diversa dal 50% la probabilità di ottenere testa nel lancio di una moneta non truccata?

A partire dal concetto di scommessa equa si può costruire un insieme di regole in grado di calcolare la probabilità di ogni tipo di evento, eventualmente subordinato a qualsiasi condizione.

Un elemento centrale della teoria rimane però il soggetto che, a partire dalle conoscenze acquisite precedentemente, deve valutare la probabilità che un particolare evento si possa verificare.

Per chiarire il concetto si rifletta sul fatto che anche nell'esperienza quotidiana il ragionamento col quale si valuta la possibilità che si verifichi un qualche evento parte dalle conoscenze che già si hanno (qual è la probabilità che piova se vedo che il cielo è nuvoloso?) ma deve essere continuamente aggiornato man mano che si acquisiscono altri elementi relativi al problema (se sento tuonare la probabilità che piova aumenta! Se poi mi sento bagnato …). Allo stesso modo la probabilità che la massa di un corpo sia negativa deve essere nulla anche in presenza di una misura che, a causa degli errori misura, risulti negativa: la conoscenza, in questo caso di leggi fisiche, deve modificare il calcolo ottenuto (probabilità condizionata di Bayes).

Si possono dare altre definizioni di probabilità che corrispondono a differenti scuole di pensiero e che per applicazioni elementari sono sostanzialmente equivalenti a quella data ma più semplice utilizzo. Analizziamo quelle che possono semplificarci alcune schematizzazioni utili nel seguito.

Definizione classica di probabilità (Laplace)

Nel caso di eventi elementari equiprobabili la probabilità di un evento è pari al rapporto fra il numero di casi favorevoli al verificarsi dell'evento e il numero totale di casi possibili:

$$P = \frac{\text{n. casi favorevoli}}{\text{n. casi possibili}}$$

Sorvoliamo per il momento sul fatto che questa definizione di probabilità si fondi sul concetto di equiprobabilità e analizziamo come esempio il lancio di un dado a 6 facce. Se il dado non è truccato nessun numero verrà preferito all'altro (**equiprobabilità**).

In queste condizioni possiamo applicare la definizione e calcolare, per esempio, la probabilità che esca un numero pari[(1)].
I casi favorevoli (numero pari) sono rappresentati da 2, 4, 6 e i casi totali da 1, 2, 3, 4, 5, 6; abbiamo 3 casi favorevoli su 6 possibili: quindi la probabilità di ottenere un numero pari è $\frac{3}{6} = \frac{1}{2} = 50\%$ come potevamo aspettarci intuitivamente in base alla nostra esperienza.

Dalla definizione si può notare come la probabilità sia un numero compreso fra 0 (nessun caso favorevole) e 1 (tutti gli eventi possibili sono favorevoli): $\mathbf{0 \le p \le 1}$.

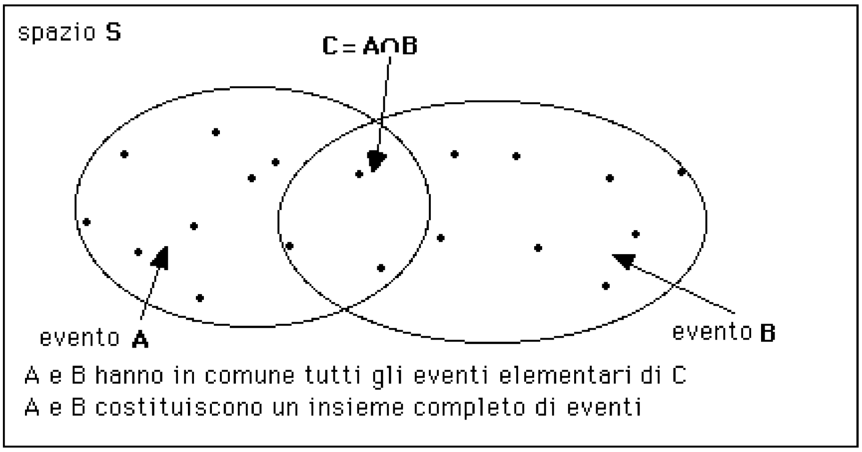
Anche se è impossibile definire la probabilità nel caso di eventi elementari non equiprobabili (per esempio un dado truccato) la definizione classica è spesso molto utile.
Possiamo per esempio calcolare facilmente la probabilità di un evento elementare: nel caso del dado a 6 facce l'uscita di un particolare numero corrisponde ad un evento elementare mentre abbiamo 6 casi possibili; la probabilità che esca un numero particolare è $\frac{1}{6}$. Quindi: $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$.

Come altro esempio consideriamo il lancio di 3 monete e chiediamoci quale sia la probabilità di ottenere 2 e solo 2 facce uguali.
Indichiamo con T testa e C croce. Tutti i possibili risultati sono costituiti da:

TTT TTC TCT TCC CTT CTC CCT CCC.

Abbiamo quindi 6 casi favorevoli su 8 possibili e la probabilità del nostro evento è pari a $\frac{6}{8}$ (bisogna prestare molta attenzione nel contare i casi favorevoli e quelli possibili; per esempio TTC, TCT e CTT sono 3 eventi diversi anche se contengono lo stesso numero di T e C).

In generale, utilizzando la definizione classica di probabilità è immediato calcolare la probabilità di verificarsi di eventi composti: esaminiamo nuovamente la figura e calcoliamo la probabilità che si verifichi l'evento A o l'evento B (unione).



[1] I possibili risultati del lancio {1, 2, 3, 4, 5, 6} sono eventi elementari (non sono scomponibili in eventi più semplici); un risultato pari è invece un evento non elementare perché è rappresentato da più eventi elementari {2}, {4}, {6}.

Supponiamo che l'evento A sia costituito da n_A eventi elementari, l'evento B da n_B eventi elementari, che l'intersezione di A e B contenga n_C eventi elementari e che lo spazio S sia costituito da n_T eventi elementari.

$$P(AoB) = P(A \cup B) = \frac{n_A + n_B - n_C}{n_T} = \frac{n_A}{n_T} + \frac{n_B}{n_T} - \frac{n_C}{n_T} = P(A) + P(B) - P(A \cap B).$$

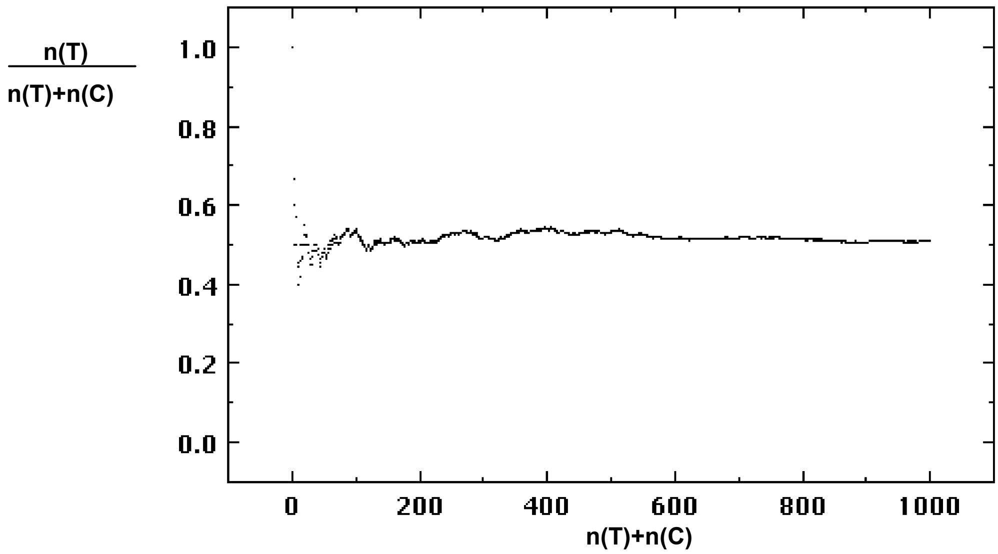
Nel caso quindi di eventi incompatibili, cioè quando $P(A \cap B)=0$, la probabilità di A o B è semplicemente $P(A \cup B) = P(A) + P(B)$.

Anche se è impossibile definire la probabilità nel caso di eventi elementari non equiprobabili (per esempio un dado truccato) la definizione classica è spesso molto utile: la tecnica si riduce a un opportuno conteggio di eventi; per questo motivo sarà spesso utile ricorrere al calcolo combinatoriale[2].

Definizione frequentistica di probabilità (Venn)

Eseguiamo più volte lo stesso esperimento e, per semplificare l'analisi del fenomeno, dopo ogni misura riportiamo il sistema nelle condizioni iniziali. In queste condizioni ogni risultato è indipendente da tutti gli altri. Analizziamo ora la frazione di volte in cui si presenta un particolare risultato.

Come esempio consideriamo il lancio di una moneta[3] e riportiamo in un grafico in ascissa il numero di lanci e in ordinata il rapporto fra il numero di volte (**frequenza**) in cui esce testa e il numero di lanci effettuati (**frequenza relativa**):



Come si può notare, nei primissimi lanci la frequenza relativa varia notevolmente perché per diversi lanci di seguito può uscire sempre testa o sempre croce (chiameremo questa variazione **fluttuazione statistica**).

Dopo circa una cinquantina di lanci, tuttavia, s'inizia a vedere una tendenza verso il valore 50% mentre l'ampiezza delle fluttuazioni diminuisce (possiamo facilmente immaginare che una lunga sequenza di sole teste o sole croci accada molto raramente).

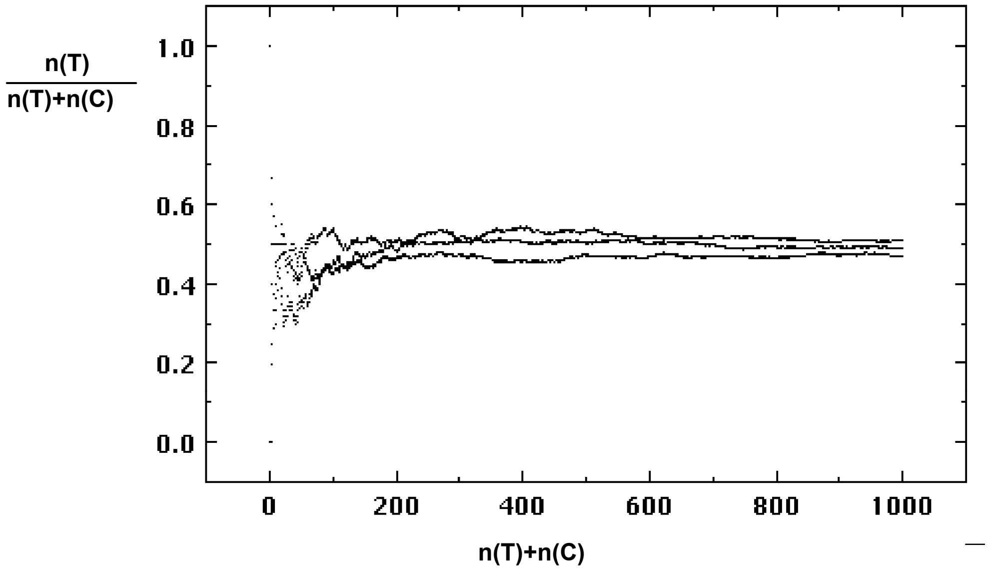
---

[2] Per alcuni richiami al calcolo combinatoriale vedere l'appendice

[3] Gli esempi di questo paragrafo sono stati ottenuti da una simulazione al calcolatore.

Intuitivamente possiamo ritenere che la frequenza relativa del 50% corrisponda alla probabilità che esca testa.

Ripetiamo l'esperienza altre 2 volte e grafichiamo nuovamente i risultati:



si può notare che l'andamento che avevamo osservato è riproducibile: al crescere del numero di lanci la frazione di teste differisce sempre meno dal valore 50% e l'ampiezza delle fluttuazioni diminuisce.
Da qui si può intuire come al crescere verso infinito del numero di lanci la frazione di teste tenderà asintoticamente al 50%; nella definizione frequentistica questo limite viene chiamato probabilità.

Questo limite non va però inteso in senso matematico stretto perché la natura casuale del fenomeno non impedisce che da un certo lancio in poi si ottenga una sequenza di molte teste o molte croci: indipendentemente dalla storia del fenomeno ad ogni lancio la probabilità di testa è la stessa (*il caso non ha memoria*).

Formalizziamo quanto abbiamo notato finora enunciando la:

legge empirica del caso (o legge debole dei grandi numeri):
consideriamo un sistema fisico suscettibile di misura; quando il numero delle prove è molto grande la frequenza relativa "praticamente" coincide con la probabilità.

Più rigorosamente il **teorema di Bernoulli** (legge forte dei grandi numeri) enuncia:

- se un evento E ha probabilità p di verificarsi
- se si eseguono N prove indipendenti e si ottengono k successi
- scelto un $\varepsilon > 0$ piccolo a piacere;

allora: $$\lim_{N\to\infty} P\left(\left|p-\frac{k}{N}\right|<\varepsilon\right)=1$$

Il teorema afferma, più correttamente, che al crescere di N la frequenza relativa non tende alla probabilità p ma che ciò diventa probabilisticamente sempre più certo.

La definizione frequentistica di probabilità consente di risolvere il problema di eventi non equiprobabili (dado truccato) ma obbliga ad eseguire, qualora sia possibile, un numero elevato di misure (metodo a posteriori).

Come esercizio si provi a lanciare almeno una ventina di volte due monete e si osservi l'andamento della frequenza relativa con la quale si osserva testa in almeno una delle due monete. Si confronti poi il risultato ottenuto con quanto è prevedibile in base alla definizione classica della probabilità.

Definizione assiomatica di probabilità (Kolmogorov)

Seguendo la recente teoria assiomatica la probabilità viene definita sulla base dei seguenti assiomi e proprietà:

1) al verificarsi di ogni evento E viene associato un numero non negativo $0 \leq p(E) \leq 1$ che viene definito probabilità

2) se $p(E) = 1$ l'evento E è certo

3) se due eventi E_1 e E_2 rispettivamente con probabilità $p(E_1)$ e $p(E_2)$ di realizzarsi sono **mutuamente esclusivi** (o incompatibili: se si verifica uno non si verifica l'altro [4]) **la probabilità che si verifichi l'evento E_1 o E_2** è:
$\boxed{p(E_1 \textbf{ o } E_2) = p(E_1) + p(E_2)}$ (probabilità semplice o evento **OR**)

4) se due eventi E_1 e E_2 rispettivamente con probabilità $p(E_1)$ e $p(E_2)$ di realizzarsi sono **indipendenti statisticamente** (l'accadere o meno di E_1 non altera la probabilità di verificarsi di E_2) **la probabilità che si verifichino sia E_1 sia E_2** è:
$\boxed{p(E_1 \textbf{ e } E_2) = p(E_1) \text{ x } p(E_2)}$ (probabilità composta o evento **AND**)

L'assioma 3) e la proprietà 4) consentono di calcolare la probabilità in molti casi.
In generale: se le $p(E_i)$ sono uguali (eventi equiprobabili), il calcolo coincide con quello della definizione classica.

Come esempio si calcoli la probabilità che nel lancio di un dado si ottenga un particolare risultato E_1.
Poiché tale evento è incompatibile con gli altri 5 casi (eventi E_2 E_3 E_4 E_5 E_6), per il terzo assioma risulta: $p(E_1 \text{ o } E_2 \text{ o } E_3 \text{ o } E_4 \text{ o } E_5 \text{ o } E_6) = p(E_1) + p(E_2) + p(E_3) + p(E_4) + p(E_5) + p(E_6)$
Per il secondo assioma si ha che: $p(E_1 \text{ o } E_2 \text{ o } E_3 \text{ o } E_4 \text{ o } E_5 \text{ o } E_6) = 1$ [5].
Nell'ipotesi che le $p(E_i)$ siano uguali (equiprobabilità): $6\, p(E_i) = 1$ da cui si ottiene $p(E_i) = \frac{1}{6}$.

[4] Ciò significa che gli eventi E_1 e E_2 o sono eventi elementari o sono eventi composti che non hanno eventi elementari in comune.
[5] E_1, E_2, E_3, E_4, E_5 e E_6, rappresentando tutti gli eventi elementari dello spazio dei risultati del lancio di un dado, costituiscono un insieme completo di eventi; è certo che si verifichi uno di tali risultati.

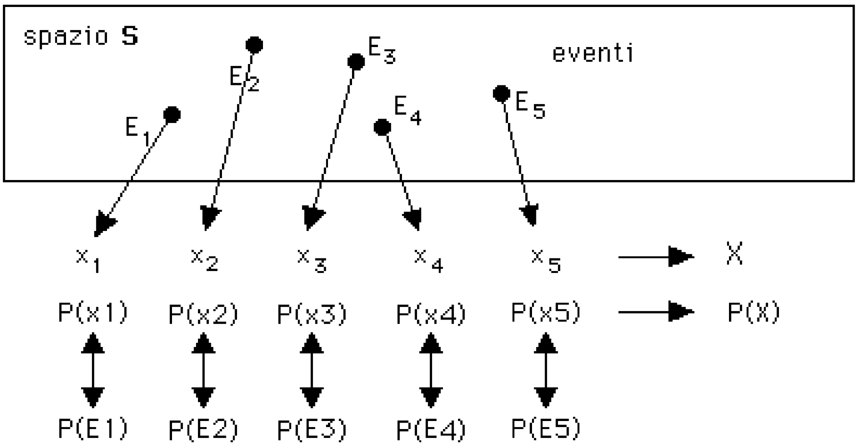
VARIABILI ALEATORIE

Consideriamo nuovamente lo spazio S degli eventi; in base a quanto visto finora siamo in grado di associare ad ogni evento, elementare o non, una probabilità. Parallelamente può essere possibile associare ad ogni evento dello spazio S un numero: se gli eventi rappresentano delle misure, i numeri da associare possono essere i risultati numerici di tali misure; se gli eventi elementari di S rappresentano le 6 facce di un dado è naturale associare ad ogni faccia il numero di punti che vi sono riportati - ma potremmo associarvi anche il quadrato di tale numero o la superficie totale dei punti che sono riportati su ogni faccia; nel caso delle lotterie a certe combinazioni di numeri si associano dei valori in denaro; …).

Se è possibile mettere in corrispondenza tutti gli eventi relativi ad un particolare fenomeno con dei numeri in modo tale che agli eventi E_1, E_2, ..., E_N corrispondano i numeri x_1, x_2, ..., x_N allora:

si definisce **variabile aleatoria** (v.a.) la variabile $X = \{x_1, x_2, ..., x_N\}$.

È quindi ovvio il significato di variabile aleatoria: si tratta di una variabile che assume particolari valori non in base a leggi deterministiche ma al caso.



Se la corrispondenza è tra gli elementi di S e una parte o tutti i numeri naturali si parla di variabile aleatoria discreta.
Se invece è necessario un insieme continuo di numeri reali la variabile aleatoria è detta continua (come spesso accade qualora S rappresenti lo spazio dei risultati di misure).

Poiché ad ogni elemento di S si può associare una probabilità, data la corrispondenza fra gli eventi E_i di S e i valori x_i assunti da X, si può parlare tanto di probabilità associata dell'evento E_i quanto di probabilità associata alla possibilità che X assuma il valore x_i: $P(E_i) = P(X=x_i)$.

VARIABILI ALEATORIE DISCRETE

In base alle definizioni precedenti, avendo associato all'evento E_i il valore x_i assunto dalla v.a. X, si ha: $P(E_i) = P(X = x_i)$. La funzione P(X) definita per i valori $X=\{x_1, x_2, ..., x_N\}$ viene detta distribuzione di probabilità (distribuzione) della variabile aleatoria discreta X. Essa è una probabilità e come tale rispetta, ad esempio, le condizioni della teoria assiomatica. Anzi, in questa teoria una qualsiasi funzione P(X) con $X = \{x_1, x_2, ..., x_N\}$ è una distribuzione di probabilità se:

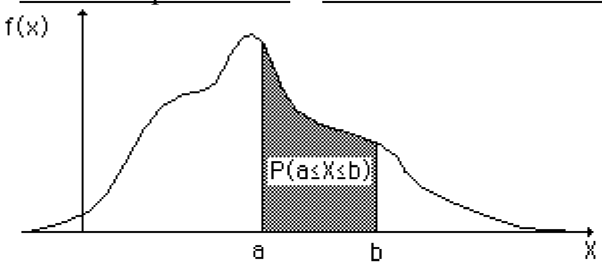
- per ogni x_i risulta $1 \geq P(x_i) \geq 0$ e se
- risulta $P(x_1) + P(x_2) + ... + P(x_N) = \sum_{i=1,N} P(x_i) = 1$ (proprietà di chiusura)

Infatti, poiché gli eventi E_i associati ai valori $\{x_1, x_2, ..., x_N\}$ costituiscono un insieme completo di eventi incompatibili, la probabilità che si verifichi almeno un evento (e quindi che si ottenga almeno un valore tra quelli possibili: $\{x_1, x_2, ..., x_N\}$) corrisponde alla certezza (P=1); inoltre, poiché gli eventi cui sono associati i valori assunti da X sono incompatibili, la probabilità che si verifichi x_1 o x_2 o ... x_N è pari a $P(x_1) + P(x_2) + ...P(x_N) = \sum_{i=1,N} P(x_i)$

VARIABILI ALEATORIE CONTINUE

Se gli eventi elementari sono infiniti è ancora possibile associare ad essi una variabile aleatoria X che risulterà continua. In questo caso non ha più senso chiedersi quale sia la probabilità che X assuma uno dei suoi infiniti valori. Sarà invece utile definire la probabilità infinitesima che X assuma un valore compreso fra x e x+dx: $dP(x) = f(x)\, dx$ [6].

Si definisce la f(x) come densità di probabilità o funzione di distribuzione (distribuzione)



Graficamente la probabilità che X assuma un particolare valore all'interno di un intervallo altri non è che l'area racchiusa al di sotto della densità di probabilità in quell'intervallo.

Se la f(x) è una funzione di distribuzione allora, considerando la **proprietà di chiusura**, si ha: $\int f(x)\, dx = 1$ qualora l'integrale venga calcolato su tutto l'insieme di definizione della variabile aleatoria X.
Infatti $\int f(x)\, dx = P(X = \text{un qualsiasi valore } x) = P(\text{evento certo}) = 1$.

Come esempio consideriamo una variabile aleatoria X uniformemente distribuita fra i valori **a** e **b**. Dato che la probabilità è uniforme, la densità di probabilità f(x) sarà nulla al di fuori dell'intervallo [**a**; **b**] e costante all'interno. Detto K il valore di tale costante, dalla proprietà di chiusura si ottiene: $1 = \int_{-\infty}^{\infty} f(x)\, dx = \int_a^b K\, dx = K\,(b-a)$. Pertanto nell'intervallo [**a**; **b**] la distribuzione di probabilità vale $f(x) = K = \dfrac{1}{b-a}$.

---

[6] Essendo una probabilità, dP(x) è una quantità adimensionale; pertanto f(x) ha le dimensioni dell'inverso di X.

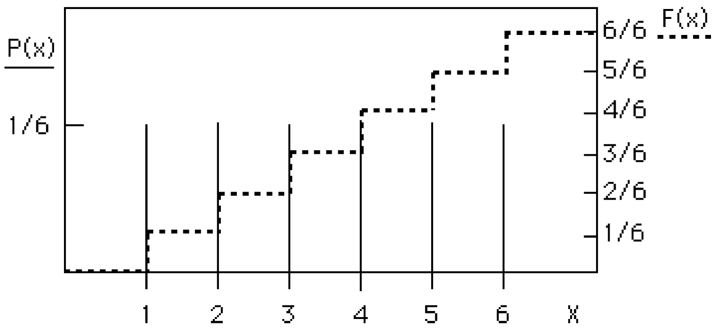
FUNZIONE CUMULATIVA

Può tornare utile definire, sia nel caso discreto sia in quello continuo, una funzione cumulativa della variabile aleatoria X che rappresenta la probabilità che la v.a. X sia minore o uguale ad un particolare valore x:

- v.a. discreta $$F(x) = \sum_{x_i \le x} P(x_i) = P(X \le x).$$

- variabile aleatoria continua $$F(x) = \int_{-\infty}^{x} f(x')\,dx' = P(X \le x).$$

Per esempio, nel caso della distribuzione uniforme di una variabile aleatoria discreta la F(x) e' una funzione a gradini (di altezza uguale, pari a P) che vale 0 per ogni X minore del più piccolo valore possibile della variabile aleatoria e vale 1 per ogni X maggiore o uguale al più grande valore possibile.



Per esempio nel caso della distribuzione uniforme di una variabile aleatoria continua definita fra **a** e **b**, la F(x) è una funzione triangolare che vale 0 per ogni valore di X minore di **a**, cresce linearmente fra **a** e **b** con pendenza $\frac{1}{b-a}$ e vale 1 per $x \ge$ **b**.

Dalle definizioni della funzione cumulativa si ha che:
$F(-\infty) = 0$ (la probabilità che la X non assuma nessun valore è nulla);
$F(\infty) = 1$ (la probabilità che la X assuma un qualsiasi valore è 1).

Tramite la funzione cumulativa è molto semplice calcolare la probabilità che la variabile aleatoria sia compresa in un certo intervallo.
Nel caso continuo, per esempio, la probabilità che X sia compresa fra x_1 e x_2 vale:

$$P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} f(x')\,dx' = \int_{-\infty}^{x_2} f(x')\,dx' - \int_{-\infty}^{x_1} f(x')\,dx' = F(x_2) - F(x_1);$$

analogamente nel caso di v.a. discrete:

$$P(x_1 \le X \le x_2) = \sum_{x_1 \le x_i \le x_2} P(x_i) = \sum_{-\infty \le x_i \le x_2} P(x_i) - \sum_{-\infty \le x_i \le x_1} P(x_i) = F(x_2) - F(x_1).$$

Dall'esempio è chiara l'utilità della funzione cumulativa: una volta definita non è più necessario distinguere fra v.a. discrete e continue.

RIASSUNTI DI UNA DISTRIBUZIONE

Spesso di una distribuzione di probabilità di una variabile aleatoria discreta o di una funzione di distribuzione di una variabile aleatoria continua (in generale diremo di una distribuzione) non è tanto importante conoscere la probabilità per ogni valore o intervallo di valori di X quanto alcuni valori caratteristici che riassumono tale distribuzione: per esempio il valore più probabile, un valore che indichi quanto la distribuzione sia simmetrica rispetto ad un valore centrale, ecc.).

VALOR MEDIO

Data una variabile aleatoria X discreta con distribuzione di probabilità P(X) si definisce valore atteso o speranza matematica o valore medio o **media** (da non confondersi con la media aritmetica [7]) della variabile aleatoria X la somma:

$$\sum_{\text{ogni}\,x_i} x_i P(x_i)$$

Analogamente, nel caso di una v.a. continua con densità di probabilità f(x), si definisce valore atteso (o media) l'integrale esteso a tutto il campo S di definizione di X:

$$\int_S x\, f(x)\, dx$$

Il significato di tali espressioni è quello di una media effettuata pesando ogni valore della variabile aleatoria con la probabilità che essa assuma quel valore (analogia con il baricentro in meccanica dove le distanze vengono pesate con le masse).

Simbolicamente il valore atteso della v.a. X si indica con **E(X)** (da "Expectation value", cioè valore previsto)[8].

Avremo in seguito bisogno di calcolare anche il valor medio non della variabile X ma di una sua funzione g(X). Tale valor medio E[g(X)] è pari alla somma su tutti i valori di X della quantità $g(x_i)\,P(x_i)$:

$$E[g(X)] = \sum_{\text{ogni}\,x_i} g(x_i)\, P(x_i)$$

e nel caso continuo

$$E[g(X)] = \int_S g(x)\, f(x)\, dx$$

In generale indicheremo con **m** il valore atteso della variabile aleatoria X:

$\boxed{\mathbf{m = E(X)}}$ è un riassunto della distribuzione di X

---

[7] La media è un riassunto di una distribuzione che è possibile calcolare quando sia nota la distribuzione di probabilità (calcolo delle probabilità); la media aritmetica è il risultato di un calcolo statistico che si effettua sui dati di un campione di risultati; solo per popolazione infinita la media aritmetica tende (probabilisticamente) alla media.

[8] Dal punto di vista dimensionale E(X) è il prodotto di X per una probabilità, quindi E(X) ha le dimensioni di X.

Ad esempio calcoliamo il valor medio della distribuzione dei risultati nel lancio di un dado:

$$E(X) = \sum_{i=1,6} x_i\ P(x_i) = \tfrac{1}{6}+\tfrac{2}{6}+\tfrac{3}{6}+\tfrac{4}{6}+\tfrac{5}{6}+\tfrac{6}{6} = \tfrac{21}{6} = \tfrac{7}{2} = 3{,}5$$

Nel caso di una distribuzione uniforme continua fra **a** e **b** si ha invece:

$$E(X) = \int_{-\infty}^{\infty} x\, f(x) dx = \int_{-\infty}^{a} x\ 0\, dx + \int_{a}^{b} \frac{x}{b-a} dx + \int_{b}^{\infty} x\ 0\ dx = \frac{b^2 - a^2}{2\,(b-a)} = \frac{a+b}{2}\,.$$

Analizziamo alcune proprietà del valore di aspettazione E(X) ricordando che è un'operazione lineare (sommatoria o integrale); per semplicità non viene esplicitato l'intervallo di sommatoria o di integrazione che è esteso a tutto il campo di definizione della v.a.(9) :

- $\mathbf{E(a\ X)} = \sum a\ x_i\ P(x_i) = a \sum x_i\ P(x_i) = \mathbf{a\ E(X)}$ [X discreta]
- $\mathbf{E(b)} = \int b\ \ f(x)\ dx = b \int f(x)\ dx = b\ 1 = \mathbf{b}$ [X continua]
- $\mathbf{E(a\ X + b)} = E(aX) + E(b) = \mathbf{a\ E(X) + b}$ [X qualsiasi]

SCARTO

Si definisce **scarto** (dal valore medio) la variabile aleatoria **X - m** cioè la differenza tra il valore della variabile aleatoria e il valore medio della sua distribuzione.

Vediamo se è possibile definire il valore atteso dello scarto (scarto medio) come un ulteriore riassunto di una distribuzione: lo scarto medio è una funzione della variabile X e quindi, applicando la definizione di valore atteso e la sua linearità si ottiene:

$$E(X - m) = E(X) - E(m) = E(X) - m = m - m = 0$$

cioè il valore atteso dello scarto dal valore atteso è sempre nullo indipendentemente dalla distribuzione e quindi non può essere utilizzato per riassumere la distribuzione di X.

VARIANZA

Si definisce **varianza** della variabile aleatoria X il valore atteso del quadrato(10) dello scarto dal valore atteso:

$$\sigma^2(X) = E\{[X - E(X)]^2\}$$

pertanto si ha: $\sigma^2(X) = \sum \{x_i - E(X)\}^2 P(x_i)$ [v.a. discreta]

$$\sigma^2(X) = \int_S \{x - E(X)\}^2 f(x)\, dx \quad \text{[v.a. continua]}$$

analogamente la varianza di una funzione g(X) è:

$$\sigma^2[g(X)] = E\{(g(X) - E[g(X)])^2\}$$

---

9 In questi esempi **a** e **b** sono costanti e non variabili aleatoria

10 Poiché dal punto di vista dimensionale $\sigma^2(X)$ è pari al prodotto di x^2 per una probabilità, essa ha le dimensioni del quadrato di X.

In generale indicheremo[11] con $\boldsymbol{\sigma}^2$ la varianza della v.a. X :

$$\boxed{\boldsymbol{\sigma^2 = \sigma^2\ (X)}}\ \text{è un riassunto della distribuzione di X}$$

Spesso sarà utile ricordare che:

$$\boldsymbol{\sigma^2(X)} = E\{[X - E(X)]^2\} = E[X^2 - 2\,X\,E(X) + E(X)^2] = E(X^2 - 2\,X\,m + m^2) =$$

$$= E(X^2) - 2\,m\,E(X) + m^2 = \boldsymbol{E(X^2) - m^2}.$$

e analogamente

$$\boldsymbol{\sigma^2\ [g(X)]} = E\{[g(X)^2 - 2\,g(X)\,E[g(X)] + E[g(X)]^2\} =$$

$$= E[g(X)^2] - 2\,E[g(X)]^2 + E[g(X)]^2 = \boldsymbol{E[g(X)^2] - E[g(X)]^2}.$$

- come esempio calcoliamo la varianza della distribuzione di probabilità nel caso del lancio di un dado: $\sigma^2 = E\{(X-m)^2\} = \frac{(-2{,}5)^2}{6} + \frac{(-1{,}5)^2}{6} + \frac{(-0{,}5)^2}{6} + \frac{0{,}5^2}{6} + \frac{1{,}5^2}{6} + \frac{2{,}5^2}{6} = \frac{17{,}5}{6} = \frac{35}{12}$

o alternativamente $\sigma^2 = E(X^2) - m^2 = \frac{1}{6} + \frac{4}{6} + \frac{9}{6} + \frac{16}{6} + \frac{25}{6} + \frac{36}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}$

- e nel caso di una distribuzione uniforme continua fra **a** e **b**:

$$\boxed{\sigma^2(X) = \int_{-\infty}^{\infty} x^2 f(x)\,dx - m^2 = \int_a^b \frac{x^2}{b-a}\,dx - m^2 = \frac{b^3 - a^3}{3(b-a)} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}}$$

Analizziamo alcune proprietà della $\sigma^2(X)$ ricordando che non è un'operazione lineare (sommatoria o integrale del quadrato di uno scarto):

- $\boldsymbol{\sigma^2(a)} = E\{[a - E(a)]^2\} = E[(a-a)^2] = \mathbf{0}$

- $\boldsymbol{\sigma^2(a\,X+b)} = E\{[(a\,X+b) - E(a\,X+b)]^2\} = E\{[(a\,X+b) - (a\,m+b)]^2\} = a^2\,E[(X-m)^2] = \boldsymbol{a^2\,\sigma^2(X)}$

DEVIAZIONE STANDARD o scarto quadratico medio

Spesso è più comodo utilizzare al posto della varianza σ^2 di una distribuzione la sua radice quadrata σ detta **deviazione standard** o <u>scarto quadratico medio</u> che ha le stesse dimensioni fisiche della variabile aleatoria[12].

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

---

11 Non è stata fissata una convenzione per i simboli da utilizzare per la media e la varianza: altri autori utilizzano E(X) e Var(X)

12 Come vedremo in seguito la deviazione standard è legata alla larghezza della distribuzione intorno alla media: più è grande la deviazione standard e maggiore è la quantità di risultati che distano dalla media.
Perché una distribuzione di probabilità non può avere deviazione standard nulla?

VARIABILE RIDOTTA

Si definisce variabile ridotta o standardizzata o scarto standardizzato la variabile:

$$\boxed{Z = \frac{X - m}{\sigma}}$$

che ha la caratteristica di essere adimensionale,

di avere valore atteso nullo: $E(\frac{X-m}{\sigma}) = \frac{E(X-m)}{\sigma} = 0$

e varianza unitaria: $\sigma^2(Z) = E[(Z\text{-}0)^2] = E(Z^2) = E\left[\left(\frac{X-m}{\sigma}\right)^2\right] = E\left[\frac{(X-m)^2}{\sigma^2}\right] = \frac{E[(X-m)^2]}{\sigma^2} = 1$

Per esempio nel caso di una v.a. X con distribuzione uniforme fra **a** e **b** è: $z = \dfrac{x - \frac{a+b}{2}}{\frac{b-a}{\sqrt{12}}}$.

••• Come **esempio dell'uso dei riassunti di una distribuzione** consideriamo nuovamente la distribuzione uniforme di una variabile aleatoria continua compresa fra **a** e **b** e chiediamoci quale sia la probabilità che X assuma un valore compreso fra $m - \sigma$ e $m + \sigma$.

Si tratta di calcolare $\int_{m-\sigma}^{m+\sigma} f(x)dx$ con $m = \frac{a+b}{2}$ e $\sigma = \frac{b-a}{\sqrt{12}}$;

poiché nell'intervallo in esame f(x) vale $\frac{1}{b-a}$,

$$\int_{m-\sigma}^{m+\sigma} f(x)dx = \frac{2\,\sigma}{b-a} = \frac{1}{\sqrt{3}} = 57{,}7\ \%.$$

Questo risultato è valido in generale: indipendentemente dai valori assunti da **a** e **b**, nel caso di una distribuzione uniforme, la variabile aleatoria X è compresa nel 57,7 % dei casi nell'intervallo $[m - \sigma; m + \sigma]$.

Ovviamente lo stesso risultato si può ottenere utilizzando la variabile ridotta:

$$\int_{m-\sigma}^{m+\sigma} f(x)dx = \int_{m-\sigma}^{m+\sigma} \frac{1}{b-a} dx = \int_{-1}^{1} \frac{\sigma}{b-a} dz = \int_{-1}^{1} \frac{1}{\sqrt{12}} dz = \frac{1}{\sqrt{3}}.$$

Nel seguito chiameremo intervallo di confidenza un intervallo, per esempio $[m - \sigma; m + \sigma]$, della variabile aleatoria e livello di confidenza la probabilità $P(m - \sigma < X < m + \sigma)$ ad esso associato. Vedremo che, come nel caso della distribuzione uniforme, in un intervallo di semiampiezza σ centrato intorno al valore medio è compreso più del 50% della probabilità.

DISUGUAGLIANZA DI CHEBYCHEV

Molto spesso la distribuzione di una variabile aleatoria non è nota ma eseguendo una serie di misure se ne possono stimare media e varianza. A partire dalla sola conoscenza di **m** e $\boldsymbol{\sigma^2}$ è possibile ricavare molte informazioni:

	disuguaglianza di Chebychev:
v.a. discreta	$\sum_{m-K\sigma\le x_i\le m+K\sigma} P(x_i) \ge 1-\frac{1}{K^2}$
v.a. continua	$\int_{m-K\sigma}^{m+K\sigma} f(x)\,dx \ge 1-\frac{1}{K^2}$

Dimostriamo l'uguaglianza nel caso discreto: consideriamo la varianza di una variabile aleatoria discreta:

$$\sigma^2(X) = \sum (x_i - m)^2 P(x_i);$$

- scomponiamo la sommatoria in tre parti:

$$\sum_{x_i<m-K\sigma}(x_i-m)^2 p(x_i) + \sum_{m-K\sigma\le x_i\le m+K\sigma}(x_i-m)^2 p(x_i) + \sum_{x_i>m+K\sigma}(x_i-m)^2 p(x_i)$$

dove K è una costante positiva;

- eliminando $\sum_{m-K\sigma\le x_i\le m+K\sigma}(x_i-m)^2 P(x_i) \ge 0$ si ha: $\sigma^2(X) \ge \sum_{x_i<m-K\sigma}(x_i-m)^2 P(x_i) + \sum_{x_i>m+K\sigma}(x_i-m)^2 P(x_i)$
- Negli intervalli $x_i < m - K\sigma$ e $x_i > m + K\sigma$ risulta $(x_i-m)^2 \ge K^2\sigma^2$, quindi:

$$\sigma^2(X) \ge K^2\sigma^2\Big[\sum_{x_i<m-K\sigma} P(x_i) + \sum_{x_i>m+K\sigma} P(x_i)\Big] \qquad \textbf{(1)}$$

- Dalla proprietà di chiusura:

$$1 = \sum_{x_i<m-K\sigma} P(x_i) + \sum_{m-K\sigma\le x_i\le m+K\sigma} P(x_i) + \sum_{x_i>m+K\sigma} P(x_i)$$

e quindi

$$\sum_{x_i<m-K\sigma} P(x_i) + \sum_{x_i>m+K\sigma} P(x_i) = 1 - \sum_{m-K\sigma\le x_i\le m+K\sigma} P(x_i)$$

E, sostituendo in **(1)**, si ha: $\sigma^2 \ge K^2\sigma^2\Big[1 - \sum_{m-K\sigma\le x_i\le m+K\sigma} P(x_i)\Big]$ cioè: $\frac{1}{K^2} \ge 1 - \sum_{m-K\sigma\le x_i\le m+K\sigma} P(x_i)$

da cui segue la disuguaglianza di Chebychev:

$$\sum_{m-K\sigma\le x_i\le m+K\sigma} P(x_i) \ge 1 - \frac{1}{K^2} \qquad \textbf{(2)}$$

Analogamente nel caso di una variabile aleatoria continua: $\sigma^2(X) = \int (x-m)^2 f(x)\,dx;$

- scomponiamo analogamente l'integrale in tre parti:

$$\sigma^2(X) = \int_{-\infty}^{m-K\sigma}(x-m)^2 f(x)dx + \int_{m-K\sigma}^{m+K\sigma}(x-m)^2 f(x)dx + \int_{m-K\sigma}^{\infty}(x-m)^2 f(x)dx$$

- Con lo stesso procedimento già usato risulta: $\sigma^2(X) \ge K^2\sigma^2\Big[1 - \int_{m-K\sigma}^{m+K\sigma} f(x)dx\Big]$

da cui segue la disuguaglianza di Chebychev: $\int_{m-K\sigma}^{m+K\sigma} f(x)\,dx \ge 1 - \frac{1}{K^2}$ **(3)**

Le quantità **(2)** e **(3)** $\sum_{m-K\sigma \le x_i \le m+K\sigma} P(x_i)$ e $\int_{m-K\sigma}^{m+K\sigma} f(x)\,dx$ rappresentano la probabilità che la variabile aleatoria X ha di essere compresa nell'intervallo [m - K σ ; m + K σ].
La disuguaglianza di Chebychev può essere quindi letta nel seguente modo:

$$\boxed{P\left(|x-m| \le K\sigma\right) \ge 1 - \frac{1}{K^2}}$$

Dalla disuguaglianza segue, cioè, che **pur non conoscendo la distribuzione della X è possibile calcolare dei limiti inferiori per i livelli di confidenza**:

K	$P\left(\lvert x-m \rvert \le K\sigma\right) \ge$
1	0
1,5	55,6 %
2	75,0 %
2,5	84,0 %
3	88,9 %
4	93,7 %
5	96,0 %

Da qui segue l'importanza di conoscere almeno la media e la varianza di una distribuzione: sono sufficienti a fornire, anche da sole, notevoli informazioni.

Qualora fosse nota la distribuzione di X sarebbe possibile calcolare esattamente i livelli di confidenza e il risultato sarebbe superiore ai valori ottenuti da Chebychev (per esercizio si confronti la tabella con quanto ricavabile da una distribuzione uniforme e dalla distribuzione di Gauss).

IL CASO DI PIÙ VARIABILI ALEATORIE

Nell'elaborazione dei dati relativi ai risultati di misure molto spesso si avrà bisogno di considerare funzioni di più variabili aleatorie. Estendiamo quindi alcuni dei concetti finora acquisiti al caso di più variabili aleatorie.

Per semplicità ci limiteremo al caso in cui queste sono tutte variabili continue (l'estensione al caso di variabili discrete e a quello di variabili sia continue sia discrete è ovvia). Sempre per semplicità consideriamo solo il caso di due v.a.; l'estensione a più variabili è immediata.

Si definisce <u>distribuzione di probabilità congiunta</u> (evento AND) di due variabili aleatorie (e analogamente nel caso di più v.a.) la probabilità che, contemporaneamente [13], ciascuna v.a. assuma un particolare valore:

$$\boxed{d^2P = d^2P(x,y) = P[(x \le X < x + dx), (y \le Y < y + dy)] = f(x,y)\, dx\, dy}$$

Nota la f(x,y) è possibile calcolare la <u>distribuzione marginale della v.a. X</u> cioè la distribuzione di probabilità della v.a. X senza nessuna richiesta sui valori assunti dalla v.a. Y:

$$f(x) = \int_{S_Y} f(x,y)\, dy\ ;$$

analogamente la distribuzione marginale della Y è $\quad f(y) = \int_{S_X} f(x,y)\, dx\ .$

Calcoliamo ora il valore atteso di X:

$$\mathbf{E(X)} = \iint_{S_X S_Y} x\ f(x,y)\, dx\, dy = \int_{S_X} x \left(\int_{S_Y} f(x,y)\, dy \right) dx = \int_{S_X} x\ f(x)\, dx = \mathbf{m_X}.$$

Analogamente per la varianza di X:

$$\boldsymbol{\sigma^2(X)} = \iint_{S_X S_Y} (x - m_X)^2 f(x,y)\, dx\, dy = \int_{S_X} (x - m_X)^2 \left(\int_{S_Y} f(x,y)\, dy \right) dx = \int_{S_X} (x - m_X)^2 f(x)\, dx = \boldsymbol{\sigma^2_X}$$

Esaminiamo alcune proprietà delle operazioni di valore atteso e varianza.
Per semplicità considereremo solo il caso di 2 v.a. (X e Y con, rispettivamente, media m_X e m_Y e varianza σ^2_X e σ^2_Y) essendo ovvia l'estensione al caso della combinazione lineare di più variabili.

$$\begin{aligned}
- \mathbf{E(aX+bY+c)} &= \iint (a\, x + b\, y + c)\, f(x,y)\, dx\, dy = \\
&= a \iint x\, f(x,y)\, dx\, dy + b \iint y\, f(x,y)\, dx\, dy + c \iint f(x,y)\, dx\, dy = \\
&= a \int x \left(\int f(x,y)\, dy\right) dx + b \int y \left(\int f(x,y)\, dx\right) dy + c \iint f(x,y)\, dx\, dy = \\
&= a \int x\ f(x)\, dx + b \int y\, f(y)\, dy + c\ 1 = a\, E(X) + b\, E(Y) + c = \mathbf{a\, m_X + b\, m_Y + c} \qquad \mathbf{(1)}
\end{aligned}$$

---

[13] La contemporaneità non va intesa in senso temporale stretto ma col significato di valori assunti dalle v.a. nello stesso evento.

- **E(XY)** = $\iint x\, y\, f(x,y)\, dx\, dy$; in generale questo calcolo si complica perché non è possibile integrare separatamente le diverse variabili.

  Tuttavia, se le v.a. sono statisticamente indipendenti [14], è possibile scomporre la $f(x, y)$: $f(x,y) = f(x)\, f(y)$. Infatti $f(x,y)\, dx\, dy$ rappresenta la probabilità infinitesima che contemporaneamente $x \le X < x + dx$ e $y \le Y < y + dy$ (evento **AND**).
  Trattandosi di v.a. indipendenti la densità di probabilità congiunta è pari al prodotto delle densità di probabilità. Pertanto, in questo caso particolare (ma frequentissimo):
  $\iint x\, y\, f(x, y)\, dx\, dy = \iint x\, y\, f(x)f(y)dx\, dy = \int x\, f(x)dx \int y\, f(y)dy = \mathbf{E(X)\, E(Y)}$

A partire da questo risultato calcoliamo

- $\boldsymbol{\sigma^2(aX+bY+c)} = E\{[(a\, X + b\, Y + c) - (a\, m_X + b\, m_Y + c)]^2\} = E\{[a\, (X-m_X) + b\, (Y-m_Y)]^2\}=$

$$= E\{[a\, (X-m_X)]^2 + 2\, a\, b\, (X-m_X)\, (Y-m_Y) + [b\, (Y-m_Y)]^2\} =$$
$$= E\{[a\, (X-m_X)]^2\} +2\, a\, b\, E[(X-m_X)\, (Y-m_Y)] + E\{[b\, (Y-m_Y)]^2\} =$$
$$= a^2\, E\{[(X-m_X)]^2\} + 2\, a\, b\, E[(X-m_X)]\, E[(Y-m_Y)]^{(15)} + b^2\; E\{[(Y-m_Y)]^2\}=$$
$$= a^2\, \sigma^2(X) + 2\, a\, b\, 0\, 0 + b^2\, \sigma^2(Y) = \mathbf{a^2\, \sigma^2_X + b^2\, \sigma^2_Y}. \qquad \textbf{(2)}$$

La generalizzazione della proprietà **(1)** è di notevole importanza:

il valor medio di una combinazione lineare di variabili aleatorie è uguale alla combinazione lineare dei valori medi con gli stessi coefficienti:

$$\mathbf{E(aX+bY+c) = a\, m_X + b\, m_Y + c}$$

Altrettanto importante è la generalizzazione della proprietà **(2)**:

la varianza di una combinazione lineare di v.a. indipendenti è uguale alla combinazione lineare delle varianze con i coefficienti al quadrato:

$$\boldsymbol{\sigma^2(aX+bY+c) = a^2\, \sigma^2_X + b^2\, \sigma^2_Y}$$

• • • APPLICAZIONI

A) una combinazione lineare notevole: la media aritmetica

Utilizziamo i risultati appena ottenuti per valutare valore atteso e varianza della media aritmetica.
Consideriamo le x_i determinazioni di una v.a. X (la grandezza fisica in misura) nel seguente modo:
x_1 (prima determinazione di X) può essere vista come la prima determinazione di una v.a. X_1,
x_2 (seconda determinazione di X) può essere vista come prima determinazione della v.a. X_2,
…
x_N (N-ma determinazione di X) può essere vista come la prima determinazione di X_N.

[14] Come spesso accade e per questo ci limiteremo a considerare solo questo caso elementare nel nostro corso.
[15] Se X e Y sono v.a. indipendenti E(XY)= E(X) E(Y) e quindi $E[(X-m_X)(Y-m_Y)]=E[X-m_X]E[Y-m_Y]$; verificare per esercizio

Se il procedimento di misura non ha alterato la grandezza fisica in misura (p.es. errore di inserzione trascurabile) le X_i rappresentano N v.a. indipendenti con media m e varianza σ^2.
La media aritmetica delle x_i quindi altri non è che la combinazione lineare delle N v.a. indipendenti X_i con coefficienti tutti uguali a $\frac{1}{N}$.

Pertanto, applicando i risultati ottenuti nel caso di una generica combinazione lineare, si ottengono le relazioni:

$$E(\overline{X}) = E\left(\frac{\sum x_i}{N}\right) = E\left(\sum \frac{x_i}{N}\right) = \sum \frac{E(X_i)}{N} = \sum \frac{m}{N} = N\frac{m}{N} = m$$

il valore atteso della media aritmetica di N valori di X coincide col valore atteso di X

$$\sigma^2(\overline{X}) = \sigma^2\left(\frac{\sum x_i}{N}\right) = \sigma^2\left(\sum \frac{1}{N} x_i\right) =^* \sum \frac{1}{N^2}\sigma^2(x_i) = \sum \frac{1}{N^2}\sigma^2(X) = N\frac{1}{N^2}\sigma^2(X) = \frac{\sigma^2(X)}{N}$$

la varianza della media aritmetica è N volte più piccola della varianza di X.

B) la varianza di una funzione di più v. a. indipendenti

Analizziamo una qualsiasi funzione di più v.a. indipendenti $Y = g(X_1,X_2,...,X_N)$ e calcoliamone la varianza $\sigma^2(Y)$ in funzione delle $\sigma^2(X_i)$.

Poiché utilizzeremo il risultato in modo approssimativo, calcoliamo lo sviluppo in serie di Taylor di Y nell'intorno dei valori medi m_i di X_i

$$Y = g(X_1, X_2, ..., X_N) =$$

$$= g(m_1, m_2, ..., m_N) + \sum_{i=1,N} \left.\frac{\partial g}{\partial X_i}\right|_m (X_i - m_i) + \frac{1}{2}\sum_{i,j=1,N} \left.\frac{\partial^2 g}{\partial X_i \partial X_j}\right|_m (X_i - m_i)(X_j - m_j) + ...$$

dove le derivate sono calcolate nel punto m: $\{X_1 = m_1, X_2 = m_2, ..., X_N = m_N\}$.

Arrestando lo sviluppo al primo termine (cosa lecita se la funzione è linearizzabile nell'intorno del punto m), si ottiene:

$$\boldsymbol{\sigma^2(Y)} \approx \sigma^2\left[g(m_1, m_2, ..., m_N) + \sum_{i=1,N} \left.\frac{\partial g}{\partial X_i}\right|_m (X_i - m_i)\right] =^{(16)} \sum_{i=1,N} \left(\left.\frac{\partial g}{\partial X_i}\right|_m\right)^2 \sigma^2(X_i - m_i) =$$

$$\sum_{i=1,N} \left(\left.\frac{\partial g}{\partial X_i}\right|_m\right)^2 \sigma^2(X_i)$$

Se, invece, la funzione $Y = g(X_1, X_2, ..., X_N)$ è lineare si ottiene in modo esatto:

$$\sigma^2(Y) = \sum_{i=1,N} \left(\left.\frac{\partial g}{\partial X_i}\right|_m\right)^2 \sigma^2(X_i).$$

---

[16] Qui è stato sfruttato il fatto che la varianza di una combinazione lineare di v.a. indipendenti è uguale alla combinazione lineare delle varianze con i coefficienti al quadrato

C) la formula di propagazione delle incertezze (incertezza standard combinata)

Ogni volta che eseguiamo misure derivate di grandezze fisiche abbiamo bisogno di stimare come l'incertezza nella determinazione dei valori delle grandezze fisiche X_i misurate direttamente si ripercuota sull'indeterminazione della misura derivata $Y = f(X_1, X_2, \ldots, X_N)$.

In generale, per ogni grandezza X_i, avremo calcolato la media aritmetica di una serie di misure. Avremo anche stimato la deviazione standard di X_i calcolando la deviazione standard sperimentale della media aritmetica (incertezza di tipo A). A volte, invece, avremo a disposizione una sola misura e quindi la stima dell'incertezza sarà limitata all'incertezza di tipo B.
Per semplicità di notazione in entrambi i casi chiamiamo $\overline{X}_i$ la migliore stima di X_i e $\sigma_S(\overline{X}_i)$ la migliore stima della loro incertezza: $X_i = \overline{X}_i \pm \sigma_S(\overline{X}_i)$

Incertezze assolute

Utilizziamo per la funzione $Y = f(X_1, X_2, \ldots, X_N)$ la relazione ricavata in precedenza $\sigma^2(Y) \approx \sum_{i=1,N} \left(\left. \frac{\partial f}{\partial X_i} \right|_m \right)^2 \sigma^2(X_i)$ e sostituiamo ai valori attesi e alle varianze delle v.a. X_i le migliori stime che possiamo ricavare sperimentalmente:

$$Y = f(\overline{X}_1, \overline{X}_2, \cdots, \overline{X}_N) \pm \sqrt{\sum_{i=1,N} \left(\left. \frac{\partial f}{\partial X_i} \right|_{\overline{X}} \right)^2 \sigma_S^2(\overline{X}_i)}$$

dove con $\overline{X}$ abbiamo indicato il punto $\overline{X} = \{\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_N\}$.

propagazione delle incertezze assolute

Incertezze relative

Supponiamo, e in laboratorio vedremo che è un caso frequentissimo, che la funzione $Y = f(X_1, X_2, \ldots, X_N)$ sia espressa sotto forma di un monomio: $Y = c\, X_1^{p_1}\, X_2^{p_2} \ldots X_N^{p_N}$.
Applichiamo la formula di propagazione delle incertezze assolute iniziando a calcolare i coefficienti di sensibilità $\frac{\partial f}{\partial X_i} = c\, X_1^{p_1}\, X_2^{p_2} \ldots p_i X_i^{p_i - 1} \ldots X_N^{p_N} = p_i \frac{Y}{X_i}$.

Pertanto $\sigma_s(Y) = \sqrt{\sum_{i=1,N} \left(p_i \frac{Y}{X_i} \right)^2 \sigma^2(X_i)} = Y \sqrt{\sum_{i=1,N} p_i^2 \left(\frac{\sigma(X_i)}{X_i} \right)^2}$.

Passando alle incertezze relative: $\frac{\sigma_s(Y)}{Y} = \sqrt{\sum_{i=1,N} p_i^2 \left(\frac{\sigma(X_i)}{X_i} \right)^2}$ e sostituendo ai valori attesi e alle varianze delle v.a. X_i le migliori stime che possiamo ricavare sperimentalmente:

$$\frac{\sigma_s(Y)}{Y} = \sqrt{\sum_{i=1,N} p_i^2 \left(\frac{\sigma_s(\overline{X}_i)}{\overline{X}_i} \right)^2}$$

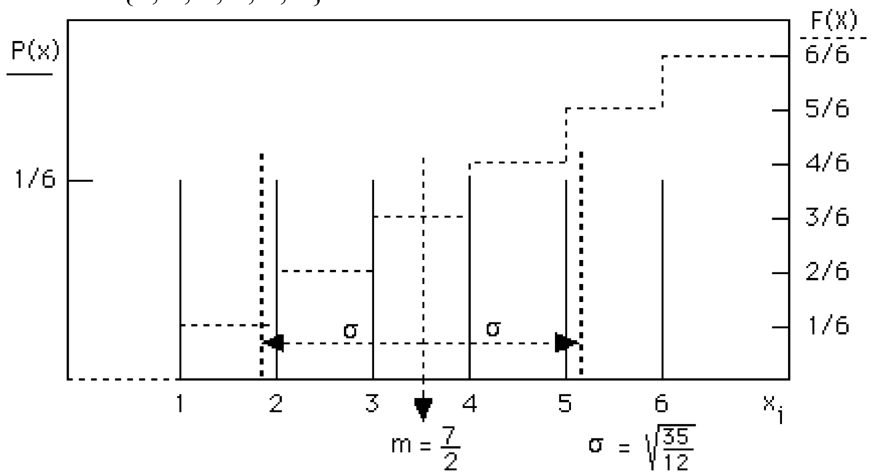
propagazione delle incertezze relative
(valida solo per monomi)

DISTRIBUZIONI DI PROBABILITÀ

Analizziamo alcune delle distribuzioni che incontreremo più frequentemente.

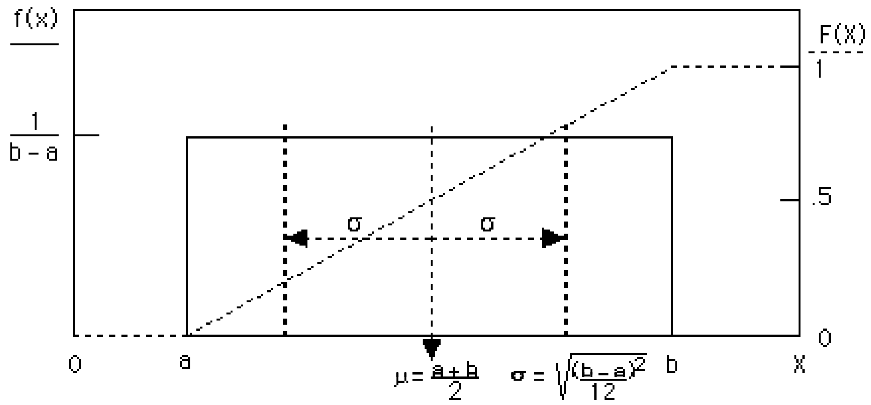
DISTRIBUZIONE UNIFORME o RETTANGOLARE

Consideriamo il caso del lancio di un dado: la distribuzione uniforme della variabile aleatoria discreta X = {1, 2, 3, 4, 5, 6} e la funzione cumulativa avranno l'andamento:



Invece, come abbiamo visto, una v.a. continua con distribuzione uniforme fra **a** e **b** è caratterizzata da una densità di probabilità: $f(x) = \frac{1}{b-a}$ che ha come riassunti media $m = \frac{a+b}{2}$ e varianza $\sigma^2 = \frac{(b-a)^2}{12}$.

Graficamente:



Questa distribuzione è spesso utilizzata per la stima di incertezze di tipo B: si ipotizza che la misura ottenuta sia la determinazione di una v.a. (con valore vero pari al risultato ottenuto) uniformemente distribuita all'interno di un particolare intervallo [a;b] pari, a seconda dei casi, alla divisione sulla scala di uno strumento analogico, ad un certo numero di digit in uno strumento digitale, alla larghezza dell'intervallo definito da una tolleranza o da un valore minimo e uno massimo al di fuori del quale non è ragionevole aspettarsi dei risultati.

In questi casi l'incertezza viene stimata con la deviazione standard $\boxed{\sigma_B = \frac{b-a}{\sqrt{12}}}$

DISTRIBUZIONE BINOMIALE o di BERNOULLI

È la distribuzione di probabilità di una variabile aleatoria discreta K che conta il numero di successi in una serie di **N** tentativi indipendenti, ciascuno con probabilità **p** di successo e probabilità q = 1 - p di insuccesso:

$$\boxed{P_{N,p}(k) = \binom{N}{k} p^k q^{N-k}}$$

Valore medio: $\mathbf{m = N\,p}$ **Varianza:** $\boldsymbol{\sigma^2 = N\,p\,q}$

La distribuzione binomiale è completamente caratterizzata dai due parametri **N** e **p**.
La variabile aleatoria K può assumere tutti i valori discreti compresi fra 0 (nessun successo nelle N prove) e N (solo successi).

DERIVAZIONE DELLA DISTRIBUZIONE

1) Consideriamo uno spazio S contenente solo due eventi E_1 e E_2, anche non elementari ma incompatibili [17], e supponiamo di conoscere $p = P(E_1)$ e $q = P(E_2)$; poiché E_1 e E_2 costituiscono un insieme completo, dalla proprietà di chiusura segue p + q = 1.

2) Eseguiamo N prove indipendenti e supponiamo che l'evento E_1 si presenti k volte; ovviamente l'evento E_2 si presenterà nelle rimanenti N-k volte.
Per esempio se N = 5 e k = 3, una possibile sequenza sarà : $E_1\,E_2\,E_2\,E_1\,E_1$.
- Chiediamoci quale sia la probabilità che eseguendo 5 prove indipendenti si ottenga quella particolare sequenza.
Ciò equivale alla probabilità che la prima volta si verifichi l'evento E_1 la cui probabilità è p; poi si verifichi l'evento E_2 la cui probabilità è q e così di seguito. Dobbiamo cioè valutare la probabilità congiunta di ottenere quegli eventi (che sono indipendenti).
Tale probabilità sarà quindi pari al prodotto delle probabilità dei singoli eventi:

$P(E_1)\,P(E_2)\,P(E_2)\,P(E_1)\,P(E_1) = p\,q\,q\,p\,p$; in generale avremo $p^k q^{N-k}$.

3) Poiché la sequenza in esame è solo una di tutte le possibili sequenze di N eventi con k volte E_1, non essendo interessati a quale sia l'ordine con il quale si presentano i k successi, dobbiamo calcolare quale sia la probabilità di ottenere una sequenza o un'altra o un'altra.
Le diverse sequenze sono incompatibili (il verificarsi dell'una esclude il verificarsi dell'altra) perciò la probabilità cercata sarà la somma su tutte le sequenze della loro probabilità $p^k q^{N-k}$.
- Dato che il numero totale delle possibili sequenze in cui E_1 compare k volte è pari al numero di combinazioni di N elementi in k posizioni, cioè $\binom{N}{k}$, ne segue che la distribuzione di probabilità di Bernoulli [18] è pari alla somma di $\binom{N}{k}$ termini uguali a $p^k q^{N-k}$:

$$\boxed{P_{N,p}(k) = \binom{N}{k} p^k q^{N-k}}$$

[17] Ad esempio l'evento E_1 può essere una particolare modalità di presentarsi di un fenomeno e l'evento E_2 è l'evento complementare che rappresenta tutti gli altri possibili risultati .
[18] Detta anche binomiale per l'uso del coefficiente binomiale; vedi anche l'appendice.

La quantità $\frac{N!}{k!(N-k)!}$ è rappresentata dal simbolo $\binom{N}{k}$ che si legge "N su K".

Essa è detta coefficiente di Newton o binomiale perché è utilizzata nella formula di espansione della potenza di un binomio:

$$(a+b)^N = \sum_{k=0,N} \binom{N}{k} a^k b^{N-k}$$

CARATTERISTICHE E RIASSUNTI

- Verifichiamo la proprietà di chiusura: $\sum_{k=0,N} P_{N,p}(k) = 1$.

Essendo $\sum_{k=0,N} P_{N,p}(k) = \binom{N}{k} p^k q^{N-k} = (p+q)^N$, dal fatto che $p + q = 1$, risulta $\sum_{k=0,N} P_{N,p}(k) = 1$.

- **Valore medio** : $m = \mathbf{Np}$

Lo possiamo ricavare applicando la definizione : $E(K) = \sum_{k=0,N} kP_{N,p}(k)$.

Con qualche passaggio:

$$E(K) = \sum_{k=0,N} kP_{N,p}(k) = \sum_{k=1,N} kP_{N,p}(k) = \sum_{k=1,N} k\binom{N}{k} p^k q^{N-k} =$$

$$= \sum_{k=1,N} k \frac{N(N-1)!}{k(k-1)!\,[(N-1)-(k-1)]!} p p^{k-1} q^{(N-1)-(k-1)} =$$

$$= Np \sum_{k=1,N} k \frac{(N-1)!}{(k-1)!\,[(N-1)-(k-1)]!} p^{k-1} q^{(N-1)-(k-1)}$$

Poniamo ora $j = k-1$ e $M = N-1$ (in questo modo $k = 1,N$ corrisponde a $j = 0,M$):

$$E(K) = Np \sum_{j=0,M} \frac{M!}{j!(M-j)!} p^j q^{M-j} = Np \sum_{j=0,M} P_{M,p}(j) = N\,p$$

Un metodo più efficiente per calcolare la media della bernoulliana consiste nel considerare la v.a. K che conta il numero di successi come somma di N variabili k_i, ognuna delle quali vale 1 in caso di successo della prova relativa e 0 in caso contrario.

Per esempio se $N = 4$ e $k = 3$ una particolare disposizione sarà:

$E_1 \quad E_2 \quad E_1 \quad E_1$ alla quale corrispondono:

$k_1=1 \quad k_2=0 \quad k_3=1 \quad k_4=1$ e quindi $k = k_1 + k_2 + k_3 + k_4 = 1+0+1+1 = 3$.

Analizziamo le v.a. ki : $P(k_i=1) = P(E_1) = p$ e
$P(k_i=0) = P(E_2) = q$;

quindi $E(k_i) = \sum_{k_i=0,1} k_i P(k_i) = 0\,P(0) + 1\,P(1) = 0\,q + 1\,p = p$;

e poiché $k = \sum_{i=1,N} k_i$, si ha: $E(K) = E\left(\sum k_i\right) = \sum E(k_i) = \sum p = Np$.

- **Varianza** : $\sigma^2 =$ **Npq**.

Si potrebbe calcolare applicando direttamente la definizione di varianza: $\sigma^2(K) = \sum_{k=0,N} (k - Np)^2 \binom{N}{k} p^k q^{N-k}$ e sfruttando l'uguaglianza: $\sigma^2(K) = E(K^2) - [E(K)]^2$ si potrebbe sviluppare il calcolo solo per ricavare $E(K^2)$. Da questo si otterrebbe, con qualche ulteriore passaggio, $\sigma^2 = Npq$.

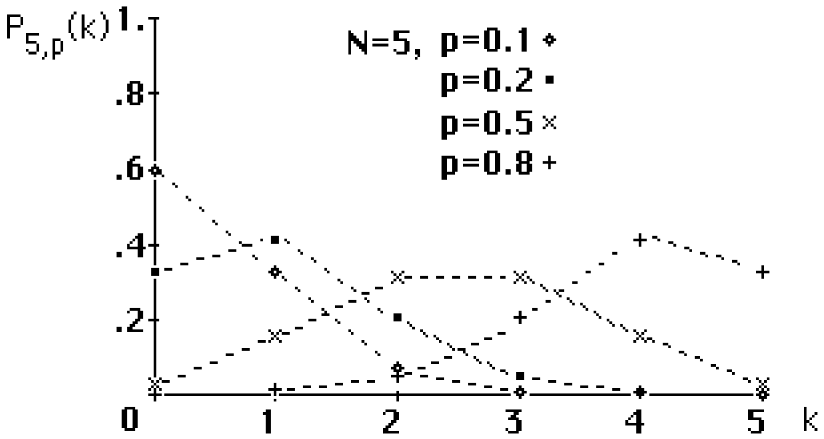
Utilizzando invece la precedente scomposizione $k = \sum_{i=1,N} k_i$ si ottiene più rapidamente:

$$\sigma^2(k_i) = \sum_{k_i=0,1} [k_i - E(k_i)]^2 P(k_i) = (0-p)^2 P(0) + (1-p)^2 P(1) = p^2 q + q^2 p = pq(p+q) = pq$$

e quindi: $\sigma^2(K) = \sigma^2\left(\sum k_i\right) =$ [19] $\sum \sigma^2(K_i) = \sum pq = N\,p\,q$.

ESEMPI

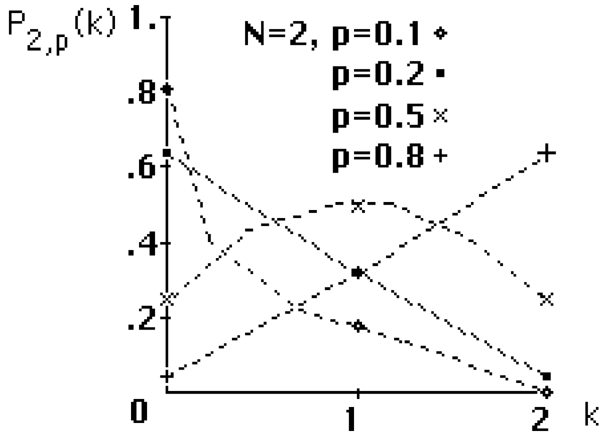
Analizziamo ora alcuni esempi di distribuzione binomiale (anche se la v.a. è discreta, al solo scopo di meglio visualizzare gli andamenti, sono state tratteggiate delle linee congiungenti i diversi punti):

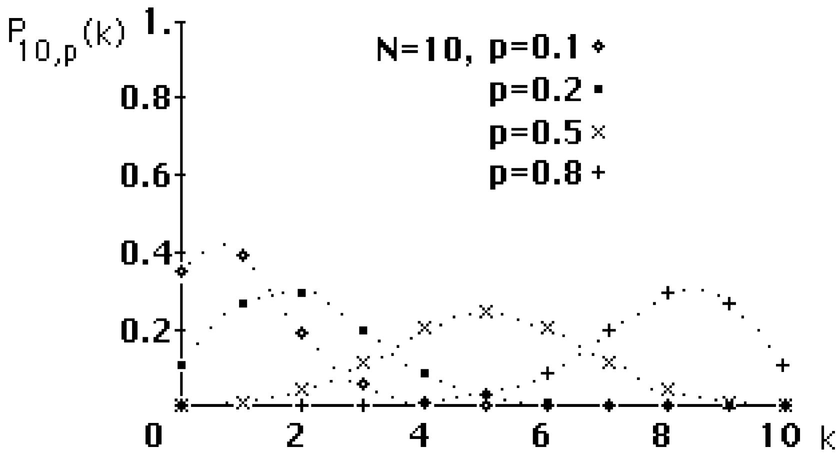


Fissiamo per il momento l'attenzione sul caso p=0,5 (simbolo "x") che può per esempio rappresentare il caso di un lancio di N = 5 monete (o 5 lanci di una moneta).
La variabile K può rappresentare il numero di teste ottenuto che varierà, ovviamente, fra 0 e 5. La $P_{5;0,5}(K)$ è la probabilità di ottenere esattamente 0,1,2,3,4,5 teste.
Negli altri casi:
- si può notare come per p = 0,2 e p = 0,8 gli andamenti siano invertiti: p = 0,2 corrisponde a q = 0,8 e viceversa;
- p = q = 0,5 ha un andamento simmetrico;
- il massimo coincide o è nelle vicinanze della media Np;

[19] Tale risultato è stato ottenuto sfruttando il fatto che le v.a. k_i sono indipendenti.

Questi altri grafici si riferiscono a diversi valori di N:





Per confrontarli fissiamo p (per esempio p = 0,2); al crescere di N da 2 a 5 a 10 il valore della massima probabilità diminuisce (conseguenza della proprietà di chiusura e del maggior numero di possibili valori k) e la distribuzione diventa più simmetrica intorno al valor medio.

Riesaminiamo ora, alla luce della distribuzione binomiale, l'esempio del lancio di una moneta utilizzato nel dare la definizione frequentistica di probabilità.

Poiché p = 0,5 e q = 0,5, al variare del numero N di lanci si avrà $m = \frac{N}{2}$ e $\sigma^2 = \frac{N}{4}$.

La frequenza relativa di teste $f = \frac{k}{N}$ avrà quindi un valore medio $E\left(\frac{k}{N}\right) = \frac{1}{N}\frac{N}{2} = \frac{1}{2}$;

la σ(f) sarà invece: $\sigma\left(\frac{k}{N}\right) = \frac{1}{N}\sqrt{\frac{N}{4}} = \frac{1}{2\sqrt{N}}$.

È quindi chiaro perché ci si attende che al crescere di N si abbia $f \to \frac{1}{2}$ e l'ampiezza delle fluttuazioni si riduca: $\sigma(f) \to 0$.

DISTRIBUZIONE DI POISSON

È la distribuzione di probabilità di una v.a. discreta K che conta il numero di volte in cui si verifica un <u>evento raro</u> (cioè di piccola probabilità: $p \to 0$) quando esso si presenta mediamente **m** volte:

$$\boxed{P_m(k) = e^{-m}\frac{m^k}{k!}}$$

Valore medio: $\mu = \mathbf{m}$ **Varianza** : $\sigma^2 = \mathbf{m}$ [20]

La distribuzione è caratterizzata dal solo parametro **m**.
La variabile aleatoria K può assumere tutti i valori discreti compresi fra k = 0 (nessun successo) e ∞ (solo successi!).

DERIVAZIONE DELLA DISTRIBUZIONE

Si può derivare questa distribuzione da quella binomiale considerando il limite delle N prove tendente ad infinito e la probabilità p di successo tendente a zero.
Tali limiti vanno però calcolati con la condizione che il valor medio Np si mantenga costante e finito: Np = m (con questa posizione $N \to \infty$ implica che $p \to 0$ e, viceversa, $p \to 0$ implica $N \to \infty$).

$$P_m(k) = \lim_{N\to\infty, Np=m} P_{N,p}(k) = \lim_{N\to\infty, Np=m}\binom{N}{k}p^k q^{N-k} = \lim_{N\to\infty}\frac{N!}{k!(N-k)!}\left(\frac{m}{N}\right)^k\left(1-\frac{m}{N}\right)^{N-k} =$$

$$= \lim_{N\to\infty}\frac{N(N-1)\cdots(N-k+1)(N-k)!}{k!(N-k)!}\frac{m^k}{N^k}\frac{\left(1-\frac{m}{N}\right)^N}{\left(1-\frac{m}{N}\right)^k} =$$

$$= \lim_{N\to\infty}\frac{N}{N}\frac{N-1}{N}\cdots\frac{N-k+1}{N}\frac{m^k}{k!}\frac{\left(1-\frac{m}{N}\right)^N}{\left(1-\frac{m}{N}\right)^k} = \frac{m^k}{k!}\lim_{N\to\infty}\frac{1}{1-\frac{m}{N}}\frac{1-\frac{1}{N}}{1-\frac{m}{N}}\cdots\frac{1-\frac{k-1}{N}}{1-\frac{m}{N}}\left(1-\frac{m}{N}\right)^N$$

passando al limite per $N \to \infty$ il prodotto dei primi k termini della sommatoria tende a 1 mentre il limite di $\left(1-\frac{m}{N}\right)^N$ vale e^{-m}; pertanto si ottiene la distribuzione di Poisson:

$$\boxed{P_m(k) = e^{-m}\frac{m^k}{k!}}$$

20 A prima vista il fatto che nella poissoniana risulti σ^2 = m potrebbe far pensare ad un errore di calcolo dimensionale; in realtà la v.a. della distribuzione di Poisson è un numero puro !

CARATTERISTICHE E RIASSUNTI

- Verifichiamo se la proprietà di chiusura $\sum_{k=0,\infty} P_m(k) = 1$ è soddisfatta.

La cosa è evidente se si riflette sul fatto che $\sum_{k=0,\infty} e^{-m}\frac{m^k}{k!} = e^{-m}\sum_{k=0,\infty}\frac{m^k}{k!} = e^{-m}e^m$

- **Valore medio**: m = **m**

Lo possiamo ricavare applicando la definizione:

$$E(K) = \sum_{k=0,\infty} kP_m(k) = \sum_{k=1,\infty} ke^{-m}\frac{m^k}{k!} = me^{-m}\sum_{k=1,\infty}\frac{m^{k-1}}{(k-1)!};$$

e ponendo $k-1=j$ si ottiene $E(K) = me^{-m}\sum_{k=0,\infty}\frac{m^j}{j!} = m\,e^{-m}e^m = m$

Non potevamo aspettarci un risultato diverso vista la definizione di **m** nella distribuzione di Poisson (**m** numero medio di successi).

- **Varianza** : σ^2 = **m**

Applichiamo la definizione di varianza ricordando che $\sigma^2(K) = E(K^2) - [E(K)]^2$:

$$E(K^2) = \sum_{k=0,\infty} k^2e^{-m}\frac{m^k}{k!} = \sum_{k=1,\infty} k^2e^{-m}\frac{m^k}{k!} = \sum_{k=1,\infty} ke^{-m}\frac{m^k}{(k-1)!} =$$

$$= m\sum_{k=1,\infty} ke^{-m}\frac{m^{k-1}}{(k-1)!};$$

poniamo k-1 = j

$$= m\sum_{j=0,\infty}(j+1)e^{-m}\frac{m^j}{j!} = m\sum_{j=0,\infty} je^{-m}\frac{m^j}{j!} + m\sum_{j=0,\infty} e^{-m}\frac{m^j}{j!} = m\,E(J) + m\,1 = m^2 + m.$$

Quindi $E(K^2) = m^2 + m$; mentre $[E(K)]^2 = m^2$ da cui $\sigma^2(K) = m^2 + m - m^2 = m$.

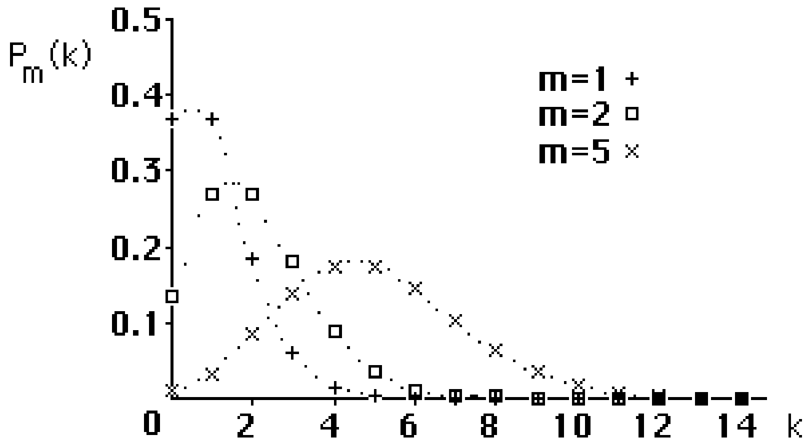
Avremmo potuto ottenere m e σ^2 molto più rapidamente considerando che la distribuzione di Poisson è derivata dalla binomiale imponendo Np = m e calcolando il limite $N \to \infty$ (o $p \to 0$) e quindi applicare tali posizioni direttamente alla media e varianza della binomiale:

$$m = Np = m;$$

$$\sigma^2 = \lim_{p\to 0} Npq = \lim_{p\to 0}(Np)(1-p) = m\lim_{p\to o}(1-p) = m$$

ESEMPI

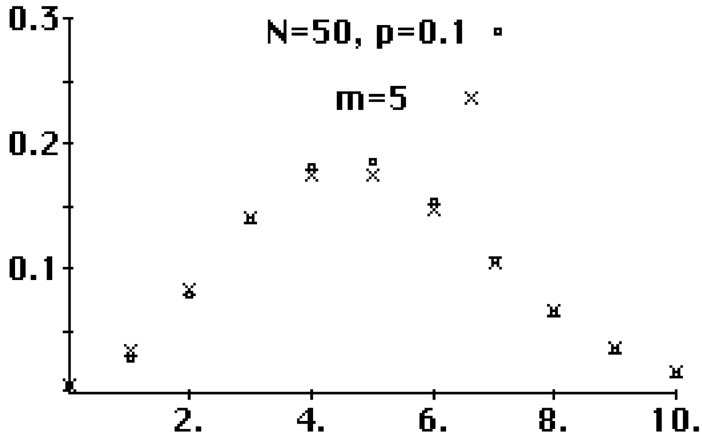
Analizziamo alcuni esempi di distribuzione di Poisson (anche in questo caso la linea che congiunge i punti non ha alcun significato dato che la funzione è definita solo per valori interi):



Dal confronto con i grafici per diversi valori di **m** si nota anche come al crescere di **m** la distribuzione diventi simmetrica rispetto alla media.

Come esempio calcoliamo la probabilità che k sia compreso in un intervallo di semiampiezza 2σ intorno a **m**. Limitiamoci a considerare il caso **m** = 2. I valori di K (interi) che cadono nell'intervallo:$[2 - 2\sqrt{2}\,;2 + 2\sqrt{2}\,] = [-0,8; 4,8]$ sono K = {0, 1, 2, 3, 4}.
Pertanto, $P = \left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} + \frac{2^4}{4!}\right)e^{-2} = 94{,}7\ \%$

È possibile ricavare la distribuzione degli eventi rari a partire dalla distribuzione binomiale ponendo Np = m e facendo tendere $p \to 0$. Quello che segue è il confronto fra una binomiale di parametri N = 50 e p = 0,1 con una poissoniana di valor medio m = Np = 5:



Si può notare quanto sia buona l'approssimazione. Essa è tanto migliore quanto più è elevato N e quanto più piccola è la probabilità p. Questo fatto consente di sostituire la binomiale con la poissoniana di più rapido calcolo.

A questo proposito consideriamo il seguente problema: se la percentuale di componenti difettosi prodotti da un'industria è pari allo 0,2 %, qual è la probabilità di trovare k componenti guasti in una fornitura di 1000 pezzi?
In questo caso N = 1000 e p = 0,002 quindi m = Np = 2: la soluzione è stata graficata precedentemente col simbolo " ". Anche se il numero di componenti guasti potrebbe essere pari a 1000 (nel nostro limite pari a ∞) con grande probabilità non si discosterà molto dal valore medio **m**: nella distribuzione di Poisson (come nella binomiale) il massimo della probabilità si ha per valori di k prossimi alla media; nell'esempio precedente avevamo visto che nell'intervallo $0 \leq k \leq 4$ il livello di confidenza è pari al 94,7 %.

DISTRIBUZIONE ESPONENZIALE

Spesso interessa conoscere dopo quanto tempo t si verifica un particolare evento raro sapendo che la sua frequenza (numero di eventi che si manifestano nell'unità di tempo) è ν. In questo caso la v.a. continua t segue la distribuzione esponenziale [21]

$$\boxed{f(t) = \nu\, e^{-\nu t}}$$

Valore medio : $m = \frac{1}{\nu}$ **Varianza :** $\sigma^2 = \frac{1}{\nu^2}$

La distribuzione esponenziale è quindi caratterizzata dal solo parametro ν [22].

La variabile aleatoria t può assumere tutti i valori compresi fra 0 (nessun evento al tempo iniziale) e ∞ (nessun evento mai!).

DERIVAZIONE DELLA DISTRIBUZIONE

Quanti eventi di un particolare tipo si verificano in un intervallo temporale Δt sapendo che la loro frequenza temporale media ν (numero di eventi nell'unità di tempo) ha un valore molto piccolo? In questo caso ν Δt rappresenta il valor medio di una distribuzione di Poisson:

$$P(\text{k eventi nell'intervallo } \Delta t) = e^{-\nu\Delta t}\frac{(\nu\Delta t)^k}{k!}$$

In particolare la probabilità che non si verifichi nessun evento nell'intervallo temporale Δt:

$$P(\text{0 eventi nell'intervallo } \Delta t) = e^{-\nu\Delta t}\frac{(\nu\Delta t)^0}{0!} = e^{-\nu\,\Delta t} \qquad \text{(a)}$$

Calcoliamo, invece, la probabilità di avere almeno un evento nell'intervallo Δt. Ricorrendo alla proprietà di chiusura si ottiene che tale probabilità è:

$$P(k \geq 1) = 1 - P(k = 0) = 1 - e^{-\nu\Delta t} \qquad \text{(b)}$$

La domanda: "dopo quanto tempo t mi aspetto che si verifichi l'evento raro?" corrisponde a: "qual è la probabilità che non si verifichi nessun evento per tutto il tempo t e che se ne realizzi almeno uno nell'istante successivo dt?"

In questo caso la probabilità infinitesima (dt è infinitesimo) vale:

$dP = f(t)\,dt = P(0 \text{ nel tempo } \Delta t{=}t) \times P(\geq 1 \text{ nel tempo } \Delta t{=}dt)$ =[23] $e^{-\nu t} \times (1 - e^{-\nu dt})$
=[24] $e^{-\nu t}\, \nu\, dt$ e quindi:

$$\boxed{f(t) = \nu\, e^{-\nu t}}$$

---

21 Quanti fenomeni fisici conoscete che seguono una legge di tipo esponenziale.
22 Le cui dimensioni sono quelle dell'inverso delle dimensioni di t (che è generalmente un tempo)
23 Utilizzando rispettivamente la (a) e la (b) e il fatto che i 2 eventi "0 in t" e "almeno 1 in dt" sono indipendenti
24 Dallo sviluppo in serie dell'esponenziale

CARATTERISTICHE E RIASSUNTI

- La proprietà di chiusura è verificata immediatamente dal calcolo: $\int_0^{\infty} \nu\, e^{-\nu t}\, dt = 1$

- **Valore medio** : $m = \frac{1}{\nu}$.

Lo possiamo ricavare applicando la definizione :

$$E(t) = \int_0^{\infty} t\nu\, e^{-\nu t}\, dt = \frac{1}{\nu}\int_0^{\infty} \nu t\, e^{-\nu t}\, d(\nu t) = \frac{1}{\nu}\int_0^{\infty} x\, e^{-x}\, dx = \frac{1}{\nu}$$ ([25]).

- **Varianza** : $\sigma^2 = \frac{1}{\nu^2}$.

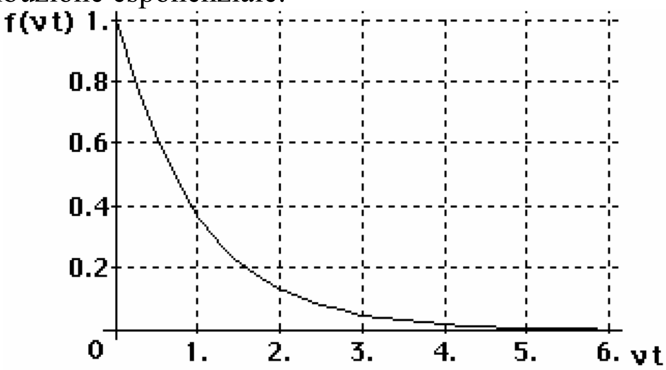
Applichiamo la definizione di varianza ricordando che $\sigma^2(K) = E(K^2) - [E(K)]^2$:

$$E(K^2) = \int_0^{\infty} t^2\, \nu\, e^{-\nu t}\, dt = \frac{1}{\nu^2}\int_0^{\infty} (\nu t)^2\, e^{-\nu t}\, d(\nu t) = \frac{1}{\nu^2}\int_0^{\infty} x^2\, e^{-x}\, dx = \frac{2}{\nu^2}$$

Quindi $\sigma^2(K) = E(K^2) - [E(K)]^2 = \frac{2}{\nu^2} - \frac{1}{\nu^2} = \frac{1}{\nu^2}$

ESEMPI

Analizziamo la distribuzione esponenziale:



Se mediamente sono costretto a far riparare la mia auto una volta l'anno, qual è la probabilità che essa si guasti k volte in 2 anni?

$$\nu = 1/y; \qquad \Delta t = 2\ y; \qquad P(k) = e^{-2}\frac{2^k}{k!}$$

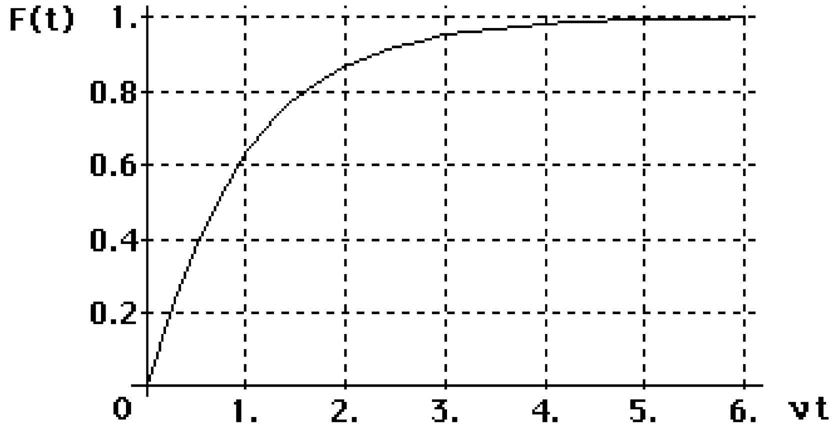
La distribuzione esponenziale risulta poi utile quando si voglia calcolare la probabilità di non aver guasti in un certo periodo di tempo, conoscendo la frequenza di tali guasti.
Per esempio, se in media si verifica un guasto l'anno in un dispositivo, la probabilità di non avere guasti (probabilità di sopravvivenza) si riduce esponenzialmente e dopo 5 anni diventa praticamente nulla $P \leq 0{,}67\%$.
La stessa probabilità si avrebbe dopo 10 anni qualora la frequenza di guasti fosse di 0,5 guasti/anno

---

25 $\frac{1}{\nu}$ rappresenta il tempo mediamente necessario perché si verifichi un evento.

La <u>funzione cumulativa</u> della distribuzione esponenziale è:

$$F(t) = \int_{-\infty}^{t} f(t)dt = \int_{0}^{t} \nu\, e^{-\nu t}\, dt = 1 - e^{-\nu t} = 1 - f(t):$$



Se in media si verifica un guasto l'anno, la probabilità di avere almeno un guasto nell'intervallo $[m-\sigma;m+\sigma] = [0;2]$ è pari a: $F(2) - F(0) = 86{,}5\%$.

Talora può essere utile la seguente tabella ricavata dalla distribuzione esponenziale:

νt	**f(νt)**	**F(νt)**
0	100,0 %	0,0 %
1	36,8 %	63,2 %
2	13,5 %	86,5 %
3	4,98 %	95,0 %
4	1,83 %	98,2 %
5	0,67 %	99,3 %