METODI STATISTICI, Ing.Inf./El./Tel. 2007/08

CALCOLO COMBINATORIO.
Una **disposizione semplice** di n oggetti a k a k è una $k-$upla ordinata di k oggetti scelti tra gli n dati (ovviamente: $k \leq n$).
Ad es. le disposizioni di 3 oggetti a, b, c a 2 a 2 ($n = 3$, $k = 2$), sono:

$$(a, b), \quad (b, c), \quad (c, a), \quad (b, a), \quad (c, b), \quad (a, c).$$

Si dice anche "disposizione semplice di n oggetti in classe k". L'aggettivo "semplice" vuol dire "senza ripetizioni"; il nome "disposizione" vuol dire che ha importanza l'ordine.

Proposizione. *Il numero di disposizioni semplici di n oggetti a k a k è il prodotto di k numeri naturali decrescenti a partire da n:*

$$D(n; k) = \ n(n-1)...(n-k+1)$$

Infatti: riempiamo k caselle in ordine: nella prima ho n possibilità di scelta, nella seconda $(n-1)$, ..., nella $k-$esima ho $n-k+1$ possibilità di scelta. □

Ad es. in quanti modi 5 amici possono sedersi su tre sedili numerati di un treno? La domanda dà importanza all'ordine (primo sedile, secondo sedile, terzo sedile), e non ci sono ripetizioni della stessa persona. Quindi si tratta delle disposizioni semplici di 5 persone a 3 a 3: $D(5; 3) = 5 \cdot 4 \cdot 3 = 60$ modi diversi di occupare i tre posti!

Una **permutazione** di n oggetti è una $n-$ upla i cui elementi sono tutti gli n oggetti. Detto altrimenti: è una disposizione semplice degli n oggetti: si tratta del caso $k = n$.

Propos. *Il numero di permutazioni di n oggetti è il prodotto dei primi n numeri naturali:*

$$P(n) = \ n(n-1)...4 \cdot 3 \cdot 2 \cdot 1 \quad \equiv \ n!$$

[Il simbolo $n!$ si legge "n fattoriale" e designa il prodotto dei primi n numeri naturali].
Infatti: una permutazione di n oggetti una disposizione semplice di n oggetti ad n ad n. Quindi $P(n) = D(n; n) = n!$ □
Ad es.: in quanti modi 5 clienti possono mettersi in fila allo sportello di banca? Risposta: tanti quanti i modi di mettere in ordine 5 persone; cioè il numero di permutazioni di 5 oggetti: $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$.

Una **combinazione semplice** di n oggetti a k a k è un sottoinsieme -senza struttura d'ordine- di k oggetti scelti tra gli n.
Si dice anche "combinazione semplice di n oggetti in classe k". L'aggettivo "semplice" vuol dire "senza ripetizioni"; il nome "combinazione" vuol dire che non ha importanza l'ordine.

Ad es. le combinazioni dei 3 oggetti a, b, c, a 2 a 2 sono:

$$\{a, b\}, \ \{b, c\}, \ \{a, c\}.$$

Si noti: {a,b}= {b,a}. Per gli insiemi astratti (per i quali si usa la parentesi graffa) non vige alcuna struttura d'ordine.

Propos. *Il numero di combinazioni semplici di* n *oggetti a* k *a* k *è:*

$$C(n;k) \;=\; \frac{n(n-1)...(n-k+1)}{k!}$$

che si suole indicare anche col simbolo $\begin{pmatrix} n \\ k \end{pmatrix}$ *e si suole scrivere in forma più compatta:*

$$\begin{pmatrix} n \\ k \end{pmatrix} \;=\; \frac{n!}{k!(n-k)!}.$$

Infatti: per ciascuna combinazione di k oggetti, esistono $P(k)$ modi di metterli in ordine. Quindi il numero di disposizioni è più grande del numero di combinazioni e precisamente:

$$D(n;k) \;=\; C(n;k)\cdot P(k) \quad\Longrightarrow\quad C(n;k) \;=\; \frac{D(n;k)}{P(k)}$$

dal che segue l'enunciato. □

Ad es. in quanti modi si possono nominare comitati di 4 persone da un gruppo di 9 persone? La domanda non dà importanza all'ordine e inoltre ogni persona nel comitato non ha ripetizione; quindi si tratta delle combinazioni semplici di 9 oggetti a 4 a 4:

$$C(9;4) \;=\; \begin{pmatrix} 9 \\ 4 \end{pmatrix} \;= 9!/[4!(9-4)!] = \;9\cdot 8\cdot 7\cdot 6/4\cdot 3\cdot 2\cdot 1 = \;126.$$

Altro es.: quante colonne posso compilare al Totocalcio con 5 "due", 6 "uno" e 2 "x"? Risposta: prima mi domando quante colonne con 13 caselle hanno 5 "due" e gli altri simboli diversi da "due": $C(13;5) = \begin{pmatrix} 13 \\ 5 \end{pmatrix}$. Poi mi domando quante colonne con 8 caselle riempite solo con "uno" e con "x" hanno esattamente 6 "uno": $C(8;6) = \begin{pmatrix} 8 \\ 6 \end{pmatrix}$. Allora la risposta è il prodotto:

$$\begin{pmatrix} 13 \\ 5 \end{pmatrix}\cdot\begin{pmatrix} 8 \\ 6 \end{pmatrix}$$

Una **disposizione con ripetizione** di n oggetti a k a k è una $k-$ upla i cui elementi sono gli n oggetti dati, con la possibilità di ripetizione.
Si noti che k può anche essere maggiore di n. Si dice anche: "disposizione con ripetizione di n oggetti in classe k". Il nome "disposizione" vuol dire che ha importanza l'ordine; l'espressione "con ripetizione" vuol dire che sono permesse ripetizioni.

Ad esempio ecco le diposizioni con ripetizione dei tre oggetti dati a, b, c a due a due (quindi: $n = 3$, $k = 2$):

$$(a, a),\ (a, b),\ (b, a),\ (b, b),\ (b, c),\ (c, b),\ (a, c), (c, a), (c, c)$$

Proposiz. *Il numero di disposizioni con ripetizione di n oggetti a k a k è:*

$$D^R(n; k) \ = \ n^k.$$

Infatti: ogni disposizione una k-upla. Occupiamo le k posizioni di $k-$upla in ordine: nella prima posizione ho n possibilità, nella seconda ho ancora n possibilità, ecc. Quindi: numero di oggetti elevato al numero di posizioni. O anche: numero di oggetti elevato al numero di classe. □

Ad esempio: quante schedine di totocalcio posso compilare? La domanda dà importanza all' ordine e si ammettono ovviamente ripetizioni. Ogni schedina è una $13-$upla di simboli scelti fra $1, 2$, x. Il numero di schedine possibili è il numero di disposizioni con ripetizione dei 3 simboli $1, 2$,x, a 13 a 13:

$$D^R(3; 13) = \ 3^{13}.$$

Una **combinazione con ripetizione** di n oggetti a k a k si ottiene scegliendo alcuni oggetti fra gli n oggetti dati e contandoli zero o una o più volte in modo che la somma delle molteplicità sia uguale a k..
Si dice anche: "combinazione con ripetizione di n oggetti in classe k". Il nome "combinazione" vuol dire che non ha importanza l'ordine; l'espressione "con ripetizione" vuol dire che sono permesse ripetizioni (quindi, tra l'altro, può essere $k > n$).

Ad es. le combinazioni con ripetizione dei 3 oggetti a, b, c a 2 a 2 sono:

$$aa,\ ab,\ ac,\ bb,\ bc,\ cc.$$

Senza dimostrazione diamo la proposizione:

Prop. *Il numero delle combinazioni con ripetizione di n oggetti a k a k è:*

$$C^R(n; k) = \ \binom{n+k-1}{k}$$

Es.: Il numero di combinazioni con ripetizione di tre oggetti a due a due:

$$C^R(3; 2) = \binom{3+2-1}{2} = \binom{4}{2} = 6.$$

Infine ecco due proposizioni immancabili dovunque si parli di calcolo combinatorio:

Proposizione. *Facendo la convenzione* $0! = 1$ *e chiamando anche in questi casi* $\binom{n}{k}$ *la quantità* $n!/[k!(n-k)!]$, *vale la seguente* **formula binomiale di Newton:**

$$(a+b)^n = \binom{n}{0} a^n + \binom{n}{1} a^{n-1}b + ... + \binom{n}{n-1} ab^{n-1} + \binom{n}{n} b^n =$$

ovvero, in notazione compatta,

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

Dim.

$$(a+b)^n = (a+b)(a+b)...(a+b) \quad \text{[n volte]}$$

è una lunga somma che contiene più volte l'addendo generico $a^{n-k}b^k$. Fissiamo k, con $k \leq n$. Quante volte appare tale addendo? Tante volte quante sono le scelte di k parentesi tra le n date, in cui pescare $b \cdot b \cdot ... \cdot b$ k volte (automaticamente allora si pesca $a \cdot ... \cdot a$ nelle rimanenti $n-k$ parentesi). In altre parole: tante volte quante sono le combinazioni semplici di k oggetti tra gli n dati. Cioè $\binom{n}{k}$ volte. Quindi tale addendo va' moltiplicato per $\binom{n}{k}$ e la somma va fatta rispetto a k come enunciato. $\square$

Esercizio: provare la proprietà dei coefficienti binomiali:

$$\binom{n-1}{k-1} + \binom{n-1}{k} = \binom{n}{k}.$$

[Grazie a questa si costruisce il famoso "triangolo di Tartaglia"]. Per ispezione diretta:

$$\frac{(n-1)!}{(k-1)!(n-k)!} + \frac{(n-1)!}{k!(n-1-k)!} = \frac{(n-1)!k + (n-1)!(n-k)}{k!(n-k)!} =$$

$$= \frac{(n-1)!(k+n-k)}{k!(n-k)!} = \binom{n}{n-k}.$$

SPAZIO DI PROBABILITÀ

Def. *Sia* Ω *un insieme. Una famiglia* Σ *di sottoinsiemi di* Ω *si dice* $\sigma-$*algebra se:*

1) $\Omega, \ \emptyset \ \in \Sigma$

2) $A \in \Sigma \implies \Omega - A \equiv A^C \ \in \Sigma$

3) $A_1, A_2, ... \in \ \Sigma \implies A_1 \cup A_2 \cup ... \in \Sigma$

In altri termini, una famiglia Σ di sottoinsiemi di Ω è una $\sigma-$algebra se comprende Ω stesso e l'insieme vuoto, e inoltre è chiusa rispetto alla complementazione e all'unione numerabile.

Lo scopo del selezionare una tale famiglia di sottoinsiemi di Ω è quello di definire senza incongruenze una misura di sottoinsiemi di Ω: a volte Σ è chiamata famiglia degli insiemi "misurabili".
Se lo spazio Ω è finito, di solito è conveniente usare come $\sigma-$algebra l'insieme delle parti di Ω, cioè l'insieme di tutti i sottoinsiemi di Ω.

Se $\Omega = R$ non risulta conveniente scegliere come Σ l'insieme delle parti. Se $\Omega = R$ oppure $\Omega = R^n$, la $\sigma-$ algebra di gran lunga più utilizzata è la minima $\sigma-$algebra contenente i sottoinsiemi aperti di Ω (naturalmente "aperti" rispetto alla metrica usuale, di R o di R^n). Essa è denominata "$\sigma-$algebra di Borel", o famiglia degli insiemi "Boreliani". Tale famiglia si ottiene anche partendo da sottofamiglie più ristrette, usando le operazioni di unione numerabile e complementazione. Fra tali sottofamiglie notevoli di sottoinsiemi di $\Omega = R$ segnaliamo gli intervalli semichiusi a destra $(a, b]$, con $a < b$, $a, b \in R$. Essa è sufficiente a generare la famiglia dei Boreliani: ad es. ogni intervallo aperto (a, b) si può ottenere come unione numerabile $\cup_n (a, b - \frac{1}{n}]$.

Def. *Uno spazio di probabilità è una terna (Ω, Σ, P) dove: Ω è un insieme non vuoto, Σ è una $\sigma-$algebra di sottoinsiemi di Ω (detta famiglia degli eventi), P è una legge a valori in $[0, 1]$ (detta probabilità) tali che*

1) per ogni evento $A \in \Sigma$ esiste il numero $P(A)$ compreso fra 0 ed 1 detto "probabilità dell' evento A";

2) $P(\Omega) = 1, \quad P(\emptyset) = 0$;

3) la probabilità P è numerabilmente additiva su ogni successione di eventi a due a due disgiunti:

$$\textit{dati gli eventi} \quad \{A_i\}_{i \in N}, \ \textit{con} \quad A_i \cap A_j = \emptyset, \ forall i \neq j, \quad \textit{si ha}$$

$$P[\ \cup_{i=1}^{\infty} A_i\ ] \ = \ \sum_{i=1}^{\infty} P(A_i) \ \equiv \ \lim_{n \to \infty} \sum_{i=1}^{n} P(A_i).$$

[In particolare per ogni coppia di eventi A, B disgiunti si ha $P(A \cup B) = P(A) + P(B)$.]

Teorema (regola di addizione per eventi arbitrari). *Se A, B sono eventi arbitrari in uno spazio di probabilità Ω, allora*

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

Dim. Scriviamo A come unione disgiunta di $A - B$ e $A \cap B$, e analogamente facciamo per B :

$$A = (A - B) \cup (A \cap B), \qquad B = (B - A) \cup (A \cap B).$$

Allora applicando due volte l'additività (3):

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

$$= P[(A - B) \cup (A \cap B) \cup (B - A)] + P(A \cap B) = P(A \cup B) + P(A \cap B)$$

da cui la tesi sottraendo $P(A \cap B)$ al primo e all'ultimo membro.

Teorema (regola di complementazione) *Se $E \subset \Omega$ è un evento ed $E^C \equiv \Omega - E$ è l'evento complementare, si ha*

$$P(E) = 1 - P(E^C).$$

Dim. Si può applicare la (3) perché sono due eventi disgiunti.

Oss. Se lo spazio di probabilità Ω è costituito da un numero finito di elementi, la famiglia Σ degli eventi coincide con la famiglia di tutti i sottoinsiemi di Ω. In tal caso, se gli elementi di Ω sono ω_i, $\quad i = 1, 2, ..., n,$, si fa spesso l'ipotesi di *"eventi elementari equiprobabili",* il che significa attribuire probabilità $\frac{1}{n}$ a ciascun evento elementare $\{\omega_i\}$. E' in questo caso che vale la definizione classica di probabilità come **"numero dei casi favorevoli diviso il numero di casi possibili".**
Esempio tipico è il dado non truccato, dove si definisce:

$$\Omega = \{1, 2, 3, 4, 5, 6\}, \; P(1) = \frac{1}{6}, ..., P(6) = \frac{1}{6}.$$

Cosi' potremo calcolare, ad es., la probabilità degli eventi

$$A: \text{ esce un numero pari}, \quad B: \text{ esce un numero minore di } 3$$

$$P(A) = P(2) + P(4) + P(6) = \frac{1}{2}, \;\; P(B) = P(1) + P(2) = \frac{1}{3}$$

Altro esempio tipico è il lancio, singolo o multiplo, di una moneta non truccata. Ad es. la probabilità che in cinque lanci di una moneta esca "testa" almeno una volta si trova introducendo l'appropriato spazio di probabilità

Ω= insieme delle 5-uple ordinate di lettere "T" o "C"

In questo spazio l'evento "non esce alcuna testa" è costituito dall'unica 5-upla (C, C, C, C, C) per cui l'evento $A =$ "esce almeno una testa" ha probabilità

$$P(A) = 1 - P(A^C) = \frac{31}{32}.$$

Altro esempio: la probabilità che almeno due fra 20 persone abbiano compleanno nello stesso giorno dell'anno è superiore o inferiore a $\frac{1}{2}$?
Basta calcolare la probabilità dell'evento complementare. Per l'evento complementare ("i 20 compleanni sono tutti distinti") il numero di casi favorevoli è il numero di disposizioni

semplici di 365 oggetti a 20 a 20; il numero di casi possibili è il numero di disposizioni con ripetizione di 365 oggetti a 20 a 20:

$$\frac{365 \cdot 364 \cdot ... \cdot 347 \cdot 346}{(365)^{20}} = (\frac{365}{365})(\frac{364}{365})...(\frac{347}{365})(\frac{346}{365})$$

Al gentile lettore vedere se il complemento a 1 di tale numero è superiore o inferiore a 1/2.

PROBABILITÀ CONDIZIONATA. EVENTI INDIPENDENTI

Spesso si usa la probabilità di un evento B sotto la condizione che avvenga un altro evento A.

Def. Si dice probabilità condizionata di B dato A

$$P(B|A) := \frac{P(A \cap B)}{P(A)}, \qquad [\ P(A) > 0\ ].$$

Es.: in una popolazione i genotipi AA, Aa, aa abbiano probabilità rispettivamente

$$P(AA) = \frac{49}{100}, \quad P(Aa) = \frac{42}{100}, \quad P(aa) = \frac{9}{100}.$$

Supponiamo che dopo un certo tempo muoiano sistematicamente gli individui di tipo aa, sicchè gli adulti sono o AA o Aa. Qual è la probabilità di AA sotto la condizione "poter diventare adulti"? Bisogna calcolare la probabilità condizionata di AA dato l'evento $C = AA \cup Aa$:

$$P(AA|\ AA \cup Aa) = \frac{P(AA \cap [AA \cup Aa])}{P(AA \cup Aa)} = \frac{P(AA)}{P(AA \cup Aa)} = \frac{0.49}{0.49 + 0.42}$$

Teorema della probabilità composta. *Se gli eventi A, B hanno entrambi probabilità non nulla*

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B).$$

Def. In uno spazio di probabilità (Ω, Σ, P) due eventi A,B sono *indipendenti* se

$$P(A \cap B) = P(A)P(B).$$

Tre eventi A_1, A_2, A_3 sono indipendenti se

$$P(A_i \cap A_j) = P(A_i)P(A_j), \forall i \neq j, \quad \text{e } P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$$

m eventi $A_1, A_2, ..., A_m$ sono indipendenti se la probablità dell'intersezione di una qualunque sottofamiglia di essi è uguale al prodotto delle probabilità degli eventi di tale sottofamiglia.

Dalla definizione di probabilità condizionata, in questo caso si ha

$$P(A|B) = P(A), \; P(B|A) = P(B)$$

il che significa che la probabilità di A non dipende dal verificarsi o meno di B, e viceversa: ciò motiva l'aggettivo "indipendente".

Es. Un test diagnostico di una malattia è corretto nel 98% dei casi. Ripetendo due volte il test sullo stesso soggetto, qual è la probabilità di un doppio errore? Sia A="errore nel primo uso del test", B="errore nel secondo uso del test". Essendo due eventi indipendenti,

$$P(A \cap B) = P(A)P(B) = (2/100)(2/100) = 4/10000 = 0,04\%.$$

Teorema di probabilità totale. *Data una "partizione" $A_1, A_2, ..., A_n, ...$ di Ω (cioè $\cup_i A_i = \Omega$ con $A_i \cap A_j = \emptyset, \;\; \forall i \neq j,$) si ha*

$$P(B) = \sum_i P(A_i)P(B|A_i)$$

Dim Sia $\{A_i\}_i$ una partizione di Ω (finita o infinita, non importa). Allora $\{(A_i \cap B)\}_i$ è una partizione di B e per l'additività numerabile si ha:

$$P(B) = \sum_i P(A_i \cap B).$$

Per definizione di probabilità condizionata,

$$P(A_i \cap B) = P(A_i)P(B|A_i)$$

e, sommando rispetto all' indice i,

$$\sum P(A_i \cap B) \; \equiv P(B) = \; \sum P(A_i)P(B|A_i).$$

$\square$

Es. Com'è noto, le trasfusioni di sangue sono possibili : dal gruppo O a tutti gruppi; da A ai gruppi A, AB ; da B ai gruppi B, AB; da AB al solo gruppo AB. Supponiamo anche che le frequenze dei gruppi sanguigni siano:

$$P(O) = 52\%, \quad P(A) = 32\%, \quad P(B) = 10\%, \quad P(AB) = 6\%.$$

Qual è la probabilità che un individuo, scelto a caso, possa donare sangue a un individuo pure scelto a caso ? Si usa il teorema della probabilità totale: la probabilità di poter donare da parte di un "A" è una probabilità condizionata appunto al fatto di essere un "A", ... :

$$P(don) = P(O)P(don|essere"O") + P(A)P(don|essere"A")$$
$$+P(B)P(don|essere"B") + P(AB)P(don|essere"AB")$$
$$= (52/100) + (32/100)(32/100 + 6/100)+$$
$$+(10/100)(10/100 + 6/100) + (6/100)(6/100) \simeq 66\%.$$

Formula di Bayes *In forma semplice:*

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

In forma generale la formula di Bayes si enuncia: data una partizione $\{A_i\}_i$ *di* Ω *si ha*

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_k P(B|A_k)P(A_k)}.$$

Dim. In forma semplice segue direttamente dal teorema della probabilità composta. Sia ora lo spazio ripartito in eventi disgiunti A_i, la cui unione sia Ω. La formula di Bayes nella forma semplice applicata ad A_i e B per i fissato dà:

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{P(B)}$$

dove ora basta sostituire nel denominatore la formula di probabilità totale. □

Es. Con la formula di Bayes ottengo la probabilità di A dato B sapendo la probabilità di B dato A. Aiuta nelle diagnosi. Ad es. se teoricamente la probabilità del sintomo B data la malattia A è il 30%, posso calcolare la probabilità che un paziente affetto dal sintomo B abbia la malattia A:

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

Se, ad es., la percentuale della malattia e del sintomo in Emilia è rispettivamente $P(A) = 0,15$ e $P(B) = 0,05$, la probabilità di malattia A dato il sintomo B è:

$$P(A|B) = \frac{(30/100)(15/100)}{5/100} = \frac{90}{100}.$$

Dunque la presenza del sintomo segnala la presenza della malattia nel 90% dei casi.

Es. Lampadine escono per il 60% da una linea di produzione A e per il 40% dalla linea B. Dalla prima linea esce un 2% di difettose, dall'altra esce un 3.8% di difettose. Con quale probabilità una lampadina difettosa è uscita dalla linea A?

Se D è l'evento "difettosa" i dati del problema sono:

$$P(D|A) = 0.02, \quad P(D|B) = 0.038, \quad P(A) = 0.6, \quad P(B) = 0.4$$

Il numero che cerchiamo è la probabilità condizionata di A dato D:

$$P(A|D) = \frac{P(D|A) \cdot P(A)}{P(D|A) \cdot P(A) + P(D|A^C) \cdot P(A^C)} =$$

$$= \frac{(0.02)(0.6)}{(0.02)(0.6) + (0.038)(0.4)} = \frac{0.012}{0.012 + 0.0152} = 0.441 = 44.1\%$$

Es. Si sa che lo 0.5% dei soggetti di una città è AIDS. Si sa che i tests diagnostici dànno diagnosi corretta nell'80% dei sani e nel 98% dei malati. Qual è la probabilità di esser sano posto che il test sia stato positivo?

Consideriamo gli eventi: A= sano, A^C = malato, B=diagnosi di sanità, B^C=diagnosi di malattia. Sappiamo che

$$P(A^C) = 0.005, \; P(B|A) = 0.80, \; P(B^C|A^C) = 0.98.$$

Vogliamo $P(A|B^C)$, che calcoleremo con la formula di Bayes:

$$P(A|B^C) = \frac{P(B^C|A) \cdot P(A)}{P(B^C|A) \cdot P(A) + P(B^C|A^C) \cdot P(A^C)} =$$

$$= \frac{(0.995)(0.20)}{(0.20)(0.995) + (0.98)(0.005)} = 0.976$$

(incredibilmente alta: ma se stiamo dentro una categoria a rischio, avremmo una incidenza di malattia $P(A^C)$ più elevata, e dunque questa probabilità di errore più contenuta).

Ultima osservazione: notiamo che ci sono due modi di "scegliere a caso" da una popolazione: 1) campionamento con reimmissione
2) campionamento senza reimmissione
come si vede da questo esempio.

Es. Una scatola contiene 10 viti, di cui tre difettose. Si estraggono due viti a caso. Con quale probabilità nessuna delle due è difettosa?

Considero gli eventi A = "prima vite estratta non difettosa", B = "seconda estratta non difettosa". Estraendo con reimmissione, prima di estrarre la seconda volta abbiamo nella scatola l'identica situazione di 10 viti di cui tre difettose; pertanto:

$$P(A) = 7/10, \quad P(B) = 7/10, \quad P(A \cap B) = P(A)P(B) = 49\%.$$

Estraendo senza reimmissione l'evento B non è più indipendente da A:

$$P(A) = 7/10, \quad P(B|A) = 6/9, \quad P(A \cap B) = P(A)P(B|A) = (\frac{7}{10})(\frac{6}{9}) \simeq 47\%$$

Se il numero di individui della popolazione e' infinito o molto grande, non c'è differenza apprezzabile tra estrarre con reimmissione ed estrarre senza reimmisione. Allora conviene per semplicità calcolare ogni cosa "come se" si estraesse con reimmissione.

VARIABILI ALEATORIE

Def. *Consideriamo uno spazio di probabilità* (Ω, Σ, P). *Si dice variabile aleatoria una funzione* $X : \Omega \to R$ *che ad ogni elemento* $\omega \in \Omega$ *fa corrispondere un numero* $X(\omega) \in R$, *in modo che ogni insieme* $\{\omega : a < X(\omega) \leq b\}$ *appartenga* $\sigma-$*algebra* Σ *(cioè alla famiglia degli eventi dello spazio* Ω*). Tale evento sarà anche indicato più concisamente "*$a < X \leq b$*".*

Es. Se $\Omega = \{1, 2, ..., 6\}$ è lo spazio di probabilità per il lancio del dado non truccato, definiamo X:= "numero uscente da un lancio", cioè $X(1) := 1, ..., X(6) := 6$. Potremo allora introdurre e calcolare la probabilità che $X = 5$, che $1 < X \leq 4$, ecc.:

$$P(1 < X \leq 4) = P(X = 2) + P(X = 3) + P(X = 4) = \frac{1}{2}$$

$$P(X \geq 3) = P(X = 3) + ... + P(X = 6) = \frac{2}{3}, \ \ P(X \leq 1,5) = \frac{1}{6}.$$

Definiamo un'altra variabile casuale sullo stesso Ω : sia $Y := 0$ se l'esito del lancio è pari, $Y := 1$ se l'esito del lancio è dispari. Così:

$$P(Y = 0) = 1/2, \ P(Y = 1) = 1/2, \ P(1 < Y \leq 4) = 0, ...\text{ecc.}$$

Def. *Ad ogni variabile aleatoria* X *è associata una misura su* R *chiamata* **legge di probabilità** *(o "distribuzione", o semplicemente "legge") relativa ad* X, *indicata* P_X *e così definita:*

$$\forall a, b \in R, \quad a < b, \quad P_X(\ (a, b]) := \ P(a < X \leq b).$$

In tal caso si dice che X *è distribuita secondo la legge* P_X: $X \sim P_X$.

N.B. Conoscendo X conosciamo la legge o distribuzione associata ad X. Al contrario, conoscendo come è distribuita X, non sappiamo necessariamente tutto su X: ad esempio la legge o distribuzione di X non dice da sola che relazioni ha X con altre variabili. Quando due o più v.a. sono distribuite secondo la stessa legge si dicono "identicamente distribuite".

Def. *Una variabile aleatoria* X *è discreta se*
1) c'è un insieme finito o numerabile di valori x_j, ($j = 1, ..., n$ *oppure* $j \in N$) *tali che*

$$P(X = x_j) > 0,$$

2) la somma delle probabilità è uno: $\sum_j P(X = x_j) = \ 1$.

Quindi una variabile aleatoria discreta assume solo un insieme discreto di valori x_j con rispettive probabilità $p_j \equiv P(X = x_j) > 0$. In altre parole, la sua distribuzione o legge è discreta nel senso che è una misura di tipo "atomico": essa si concentra su un insieme al più discreto di valori x_j attribuendo ad essi masse positive $p_j > 0$.

Ecco tre rappresentazioni equivalenti di una legge P_X asociata ad una v.a. X discreta.
Se X è discreta, una prima ovvia rappresentazione di P_X è per elencazione: si elencano i valori possibili di X e le rispettive probabilità:

$$X \sim \begin{pmatrix} x_1, & x_2, & ... \\ p_1, & p_2, & ... \end{pmatrix}$$

Una seconda rappresentazione equivalente di P_X è la "funzione di probabilità" della v.a. X:

$$f(x) = \begin{cases} p_j, & \text{se } x = x_j \ (j = 1, 2, ...) \\ 0, & \text{altrove.} \end{cases}$$

Una terza rappresentazione equivalente di P_X è la "funzione distribuzione" della v.a. X:

$$\forall x \in R, \quad F(x) := P(X \leq x) \ = \sum_{x_j \leq x} f(x_j).$$

Quest'ultima è una funzione a gradini:

$$F(x) = \begin{cases} 0, & x < x_1 \\ p_1, & x_1 \leq x < x_2 \\ p_1 + p_2, & x_2 \leq x < x_3 \\ ... & \\ p_1 + ... + p_{n-1}, & x_{n-1} \leq x < x_n \\ ... & \end{cases}$$

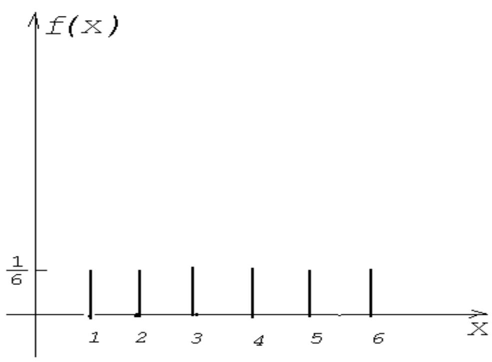
Poiché $f(x_j) = P(X = x_j)$, la funzione distribuzione ha il significato:

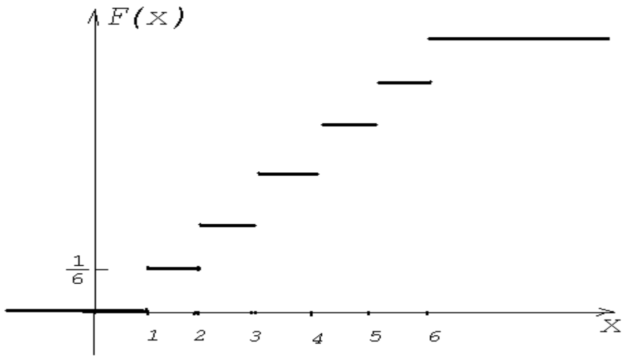
$$F(x) = P(X \leq x), \quad P(a < X \leq b) = F(b) - F(a).$$

Es. Lancio di un dado: la funzione di probabilità è

$$f(x) = \begin{cases} \frac{1}{6} & \text{se x=k} \\ 0, & \text{altrove} \end{cases}$$

ed $F(x) = 0$ per $x < 1$, $F(x) = 1/6$ per $1 \leq x < 2$, ...,$F(x) = 5/6$ per $5 \leq x < 6$, $F(x) = 1 \ \forall x \geq 6$





Es. Sullo spazio di probabilità del lancio di due dadi (i cui elementi sono le 36 coppie $(1,1),(1,2),...,(6,5),(6,6)$) sia $Z :=$ somma dei due numeri uscenti. Quindi:

$$Z : \begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \frac{1}{36} & \frac{2}{36} & \frac{3}{36} & \frac{4}{36} & \frac{5}{36} & \frac{6}{36} & \frac{5}{36} & \frac{4}{36} & \frac{3}{36} & \frac{2}{36} & \frac{1}{36} \end{pmatrix}$$

Per esercizio: descrivere il grafico di $f(x)$ e di $F(x)$.

Def. *Una v.a. X é assolutamente continua se la funzione distribuzione*

$$x \;\rightarrow\; F(x) \;\equiv\; P(X \leq x)$$

è rappresentabile come funzione integrale di una funzione $f(\cdot) \geq 0$:

$$\forall x \in R, \quad F(x) = \int_{-\infty}^{x} f(t)dt$$

La funzione $f(\cdot)$ è supposta almeno integrabile, ed è detta densità di probabilità della v.a. X.

OSS. Qui e altrove si usano integrali "impropri", cioè integrali definiti dove un estremo di integrazione (o entrambi gli estremi) è ∞. Il significato è:

$$\int_{-\infty}^{x} f(t)dt := \lim_{R\to\infty} \int_{-R}^{x} f(t)dt, \quad \int_{-\infty}^{+\infty} f(t)dt := \lim_{R\to\infty} \int_{-R}^{R} f(t)dt, \text{ecc.}$$

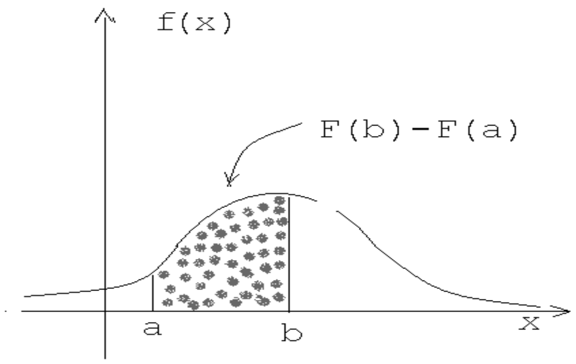
Oss. La funzione f, che appare sotto il segno di integrale, si chiama **densità di probabilità** o semplicemente "densità" della v.a. Derivando ambo i membri, avremo

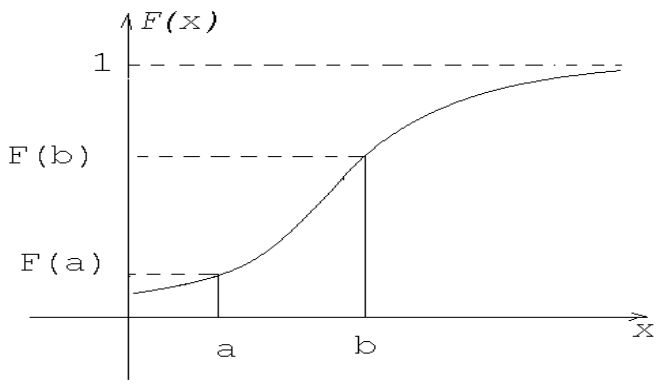
$$F'(x) = f(x), \text{ in ogni } x \text{ dove } f \text{ sia continua.}$$

Oss. importante. Poiché $P(a < X \leq b) = P(X \leq b) - P(X \leq a)$, ecco il modo standard di calcolare le probabilità come integrali della funzione densità:

$$P(a < X \leq b) = F(b) - F(a) = \int_{a}^{b} f(t)dt.$$

Perciò questa probabilità è uguale all'area sotto la curva $x \rightarrow f(x)$ tra gli estremi $x = a$ ed $x = b$.





Ciò comporta, ad es., che per una variabile aleatoria continua si ha sempre $P(X = a) = \int_a^a f(x)dx = 0$. Al contrario, può essere $P(X = a) > 0$ nel caso di variabile X discreta. Analogamente, se X è continua si ha

$$P(a < X < b) = P(a \le X \le b) = P(a \le X < b) = P(a < X \le b).$$

Al contrario, queste stesse probabilità possono differire tra loro nel caso di X discreta.

OSS. Esistono molti altri tipi di v.a.: ad esempio una v.a. è "continua" ma non "assolutamente continua" se la funzione distribuzione F è continua ma non assolutamente continua: in tal caso non esiste una densità f tale che $F(b) - F(a) = \int_a^b f(x)dx$ per ogni a, b. Nel seguito, salvo diversa indicazione, considereremo solo v.a. o discrete o assolutamente continue. Considereremo qualche volta anche v.a. "miste", ma solo quando la funzione distribuzione $x \to F(x) = P(x \le X)$ sia assolutamente continua tranne al più in punti di discontinuità di prima specie. Se un punto $x_\circ$ è discontinuità di prima specie per F, risulta positiva la probabilità

$$P(X = x_\circ) = F(x_\circ^+) - F(x_\circ^-) \equiv \text{ [per cont. da destra] } F(x_\circ) - F(x_\circ^-).$$

In generale, per tutte le possibili v.a.,

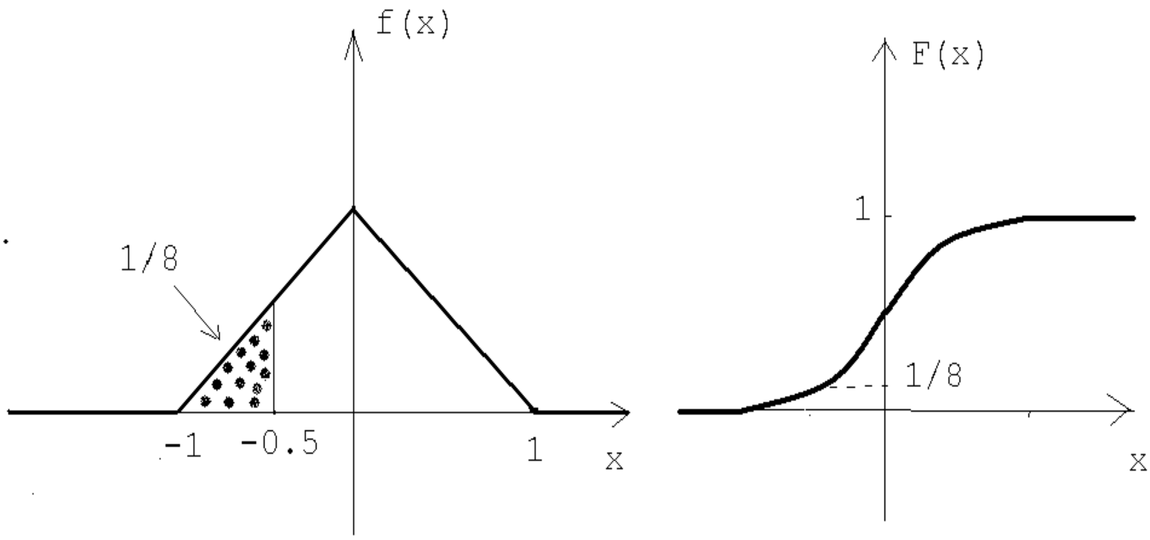
$$P(a < X \le b) = F(b) - F(a), \quad P(a \le X \le b) = F(b) - F(a) + P(X = a).$$

Es. In un processo automatico si riempiono bottigliette di sciroppo. Il contenuto risulta $Y = 100 + X$ ml , dove X è una variabile casuale di densità

$$f(x) = \begin{cases} 1 - |x|, & \text{se } |x| \le 1 \\ 0, & \text{altrove.} \end{cases}$$

Fare il grafico di $f(x)$ e di $F(x)$. In una partita di 1000 confezioni, quante approssimativamente conterranno meno di 99.5 unità di misura?

'area



Il numero di bottigliette cercato è 1000 moltiplicato per $P(Y \leq 99,5) = P(100+X \leq 99,5)$, cioè per

$$P(X \leq -0,5) = \int_{-1}^{-0,5} (1 - |x|)dx = \frac{1}{8}.$$

essendo questa l'area di un triangolo di base $\frac{1}{2}$ e altezza $\frac{1}{2}$.

Def. *Media o speranza matematica* di una variabile casuale X discreta:

$$\mu \equiv E(X) := \sum x_j f(x_j)$$

sotto l'assunzione che sia assolutamente convergente la corrispndente serie numerica: $\sum |x_j| f(x_j) < +\infty$.

Media o speranza matematica di una variabile X continua:

$$\mu \equiv E(X) := \int_{-\infty}^{+\infty} x f(x) dx.$$

sotto l'assunzione che sia assolutamente convergente il corrispondente integrale: $\int_{-\infty}^{+\infty} |x| f(x) dx < +\infty$.

Oss. La media (o valor medio o speranza) appena definita dipende dalla variabile casuale esaminata; essa, nel caso discreto, è la somma dei valori x_j moltiplicati per le rispettive probabilità $f(x_j) \equiv P(X = x_j)$.

Invece, per evitare confusioni, si rammenti che la somma di tutte le probabilità $f(x_j)$ è uno, qualunque sia la variabile casuale X: $\sum f(x_j) = \sum P(X = x_j) = 1$. Nel caso continuo, l'integrale su tutto R della densità è 1:

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

Def. Si dice *varianza* di una variabile casuale X discreta

$$\sigma^2 \equiv Var(X) := \sum_j (x_j - \mu)^2 f(x_j).$$

Varianza di una variabile casuale X continua:

$$\sigma^2 \equiv Var(X) := \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx.$$

La radice quadrata della varianza si dice *deviazione standard* e si indica σ.

Oss. La varianza è nulla solo quando X è una variabile casuale discreta con funzione di probabilità tale che $f(x_1) = 1$ in un certo punto x_1, ed $f(x) = 0$ altrove. Tranne questo unico caso, che interessa la teoria della probabilità solo come caso-limite, si ha sempre $\sigma^2 > 0$.

Es. Se un'epidemia colpisce il 30% della popolazione, la probabilità di contagio per un singolo è di $p = 0.30$. La variabile casuale

$$X : \begin{pmatrix} 0 & 1 \\ 0.7 & 0.3 \end{pmatrix}$$

ha media e varianza rispettivamente $\mu = 0 \cdot 0.7 + 1 \cdot 0.3 = 0.3$ e

$$\sigma^2 = (0 - 0.3)^2 \cdot 0.7 + (1 - 0.3)^2 \cdot 0.3 = 63/1000 + 147/1000 = 0,21$$

Sommando n variabili casuali identiche ad X si ottiene la variabile casuale Z = numero di individui contagiati in un gruppo di n persone. Ad es. se $n = 2$ avremo la variabile casuale che può assumere i valori 0 o 1 o 2 :

$$Z := \begin{pmatrix} 0 & 1 & 2 \\ (7/10)^2 & 2 \cdot (7/10)(3/10) & (3/10)^2 \end{pmatrix}$$

Quindi in un gruppo di due persone il numero atteso di persone contagiate è:

$$\mu = E(Z) = 0 \cdot 0.49 + 1 \cdot 0.42 + 2 \cdot 0.09 = 0.6$$

(non ridere, please: non è detto che la media sia uno dei valori assunti dalla variabile aleatoria), con deviazione standard dalla media:

$$\sigma = \sqrt{(0 - 0.6)^2 \cdot 0.49 + (1 - 0.6)^2 \cdot 0.42 + (2 - 0.6)^2 \cdot 0.09}.$$

Intuitivamente: la media è tanto più rappresentativa della v.a. quanto più piccola è la deviazione standard. La varianza (e anche la deviazione standard) in certo senso misura quanto è dispersa la variabile casuale rispetto alla media.

Es. Sia X distribuita uniformemente nell'intervallo $[0, b]$, cioè X è definita completamente dalla densità:

$$f(x) = \begin{cases} \frac{1}{b} & \text{se } 0 \le x \le b \\ 0 & \text{altrove} \end{cases}.$$

Ecco dunque media e varianza:

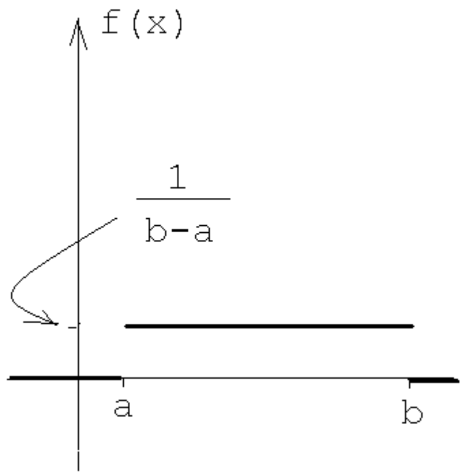
$$\mu = \int_0^b x dx = b/2, \quad \sigma^2 = \int_0^b (x - b/2)^2 b^{-1} dx = b^2/12$$

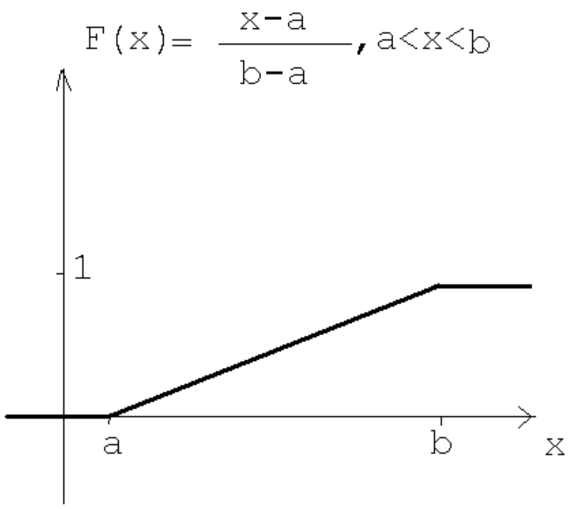
Se poi X è uniforme in $[a, b]$, cioè $f(x) = \begin{cases} \frac{1}{b-a} & \text{se } 0 \le x \le b \\ 0 & \text{altrove,} \end{cases}$

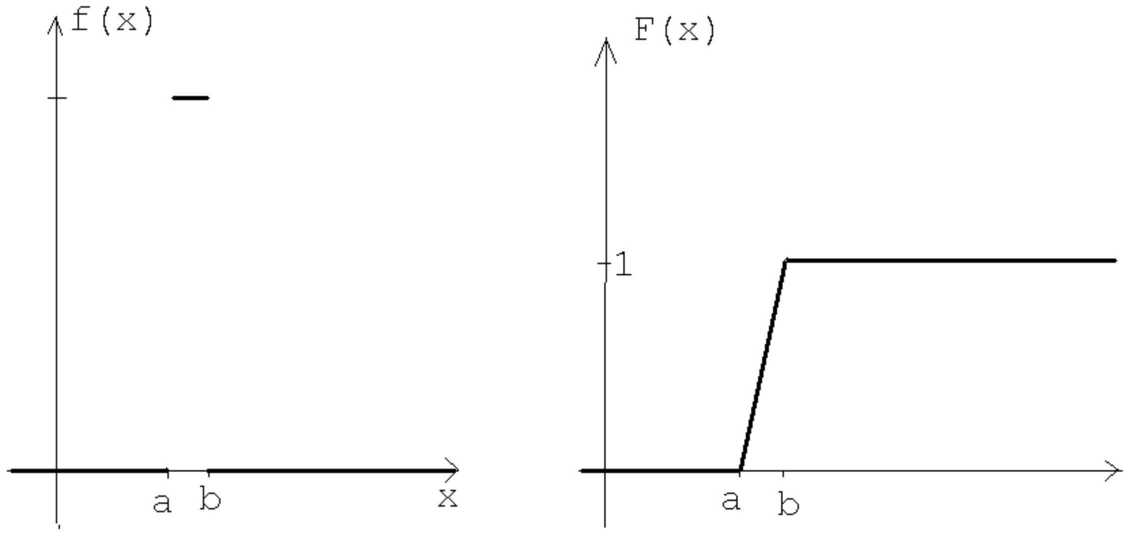
allora abbiamo la seguente

Proposizione La v.a. uniforme nell'intervallo $[a, b]$, $X \sim U([a, b])$, ha media e varianza:

$$E(X) = \frac{a+b}{2}, \qquad Var(X) = \frac{(b-a)^2}{12}$$







Proposizione. Si ha
$$E(X-\mu)=0, \qquad \sigma^2=E(X^2)-\mu^2.$$

Dim. Nel caso discreto:

$$E(X-\mu)\equiv\sum_j (x_j-\mu)f(x_j)=E(X)-\mu=0,$$

$$\sigma^2\equiv\sum_j (x_j-\mu)^2 f(x_j)=\sum_j [x_j^2-2\mu x_j+\mu^2]f(x_j)$$

$$=\sum x_j^2 f(x_j)-2\mu\sum x_j f(x_j)+\mu^2\sum f(x_j)=E(X^2)-2\mu^2+\mu^2.$$

Nel caso continuo è analogo il calcolo, tenendo conto che $\int_{-\infty}^{+\infty} f(x)dx=1,\ \int_{-\infty}^{+\infty} xf(x)dx=\mu$.

Teorema (trasformazione affine di v.a.) Se una v. a. X ha media μ e varianza σ^2, allora la v.a. $X^*=c_1X+c_2,\ \ c_1\neq 0$, ha media e varianza:

$$\mu^*=c_1\mu+c_2, \quad \sigma^{*\ 2}=c_1^2\sigma^2.$$

Dim. Lo proviamo nel caso discreto. La v.a. $X^*=c_1\mu+c_2$ è la seguente:

$$\begin{pmatrix} c_1x_1+c_2 & c_1x_2+c_2 & ... & c_1x_n+c_2 & ... \\ f(x_1) & f(x_2) & ... & f(x_n) & ... \end{pmatrix}$$

Pertanto

$$\mu^*=\sum(c_1x_j+c_2)f(x_j)=c_1\sum x_jf(x_j)+c_2\sum f(x_j)=c_1\mu+c_2$$

$$(\sigma^*)^2 = \sum (c_1 x_j + c_2 - c_1\mu - c_2)^2 f(x_j) =$$
$$= (c_1)^2 \sum (x_j - \mu)^2 f(x_j) = (c_1)^2 \sigma^2$$

Corollario (variabile standardizzata) Se X ha media μ e varianza σ^2, allora la corrispondente variabile aleatoria
$$Z = \frac{X-\mu}{\sigma}$$
ha media 0 e varianza 1.
Dim. Basta prendere $c_1 = 1/\sigma$, $c_2 = -\mu/\sigma$. $c_1 X + c_2 = 1/\sigma\ X - \mu/\sigma$ ha media e varianza rispettivamente:
$$\mu/\sigma - \mu/\sigma = 0, \quad c_1^2\sigma^2 = (\sigma^{-1})^2\sigma^2 = 1.$$

Infine, ecco la funzione generatrice dei momenti, utile a calcolare di fatto media e varianza.

Lemma sulla funzione generatrice. *Sia X una v.a. Se esistono finiti i momenti $E[X^n]$, $\forall n \in N$, e se esiste finita la funzione,*

$$G(t) = E[e^{tX}] = \begin{cases} \sum_j e^{tx_j} f(x_j), & \textit{nel caso discreto} \\ \int_{-\infty}^{+\infty} e^{tx} f(x)\, dx & \textit{nel caso continuo} \end{cases} \qquad (t \in R),$$

allora $G(t)$ soddisfa

$$E(X) \equiv \mu \ = \ G'(0), \qquad E(X^2) \ = \ G''(0), \dots, E(X^n) = G^{(n)}(0).$$

G è detta "funzione generatrice dei momenti".

Dim. Infatti, derivando sotto il segno di serie (o di integrale)

$$G'(t) = E[\ \frac{d}{dt} e^{tX}\ ] \ = \ E[\ X e^{tX}\ ], \qquad G''(t) = E[\ \frac{d^2}{dt^2} e^{tX}\ ]\ = E[\ X^2 e^{tX}\ ].$$

I momenti sono dunque:

$$G'(t)|_{t=0} = E(X), \quad G''(t)|_{t=0} = E(X^2), ..., G^{(n)}(0) = E(X^n).$$

$\square$

LEGGE DI PROBABILITÀ BINOMIALE

Def. Sia $0 < p < 1,$, $n \in N$. Sia $q = 1-p$. Una v.a. discreta X è "binomiale di parametri n e p" (o anche: segue una legge binomiale di parametri n e p) se ha funzione di probabilità:

$$P(X = k) \equiv f(k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, ..., n$$

ossia:

$$X : \begin{pmatrix} 0 & 1 & 2 & ... & n-2 & n-1 & n \\ q^n & npq^{n-1} & \frac{n!}{2!(n-2)!}p^2q^{n-2} & ... & \frac{n!}{(n-2)!2!}p^{n-2}q^2 & np^{n-1}q & p^n \end{pmatrix}$$

Notevole il caso $n = 1$, $0 < p < 1$: una v.a. X che assuma solo i valori 0 e 1 con probabilità $(1-p)$ e p rispettivamente, segue una legge di Bernoulli:

$$X : \quad \begin{pmatrix} 0 & 1 \\ (1-p) & p \end{pmatrix}.$$

Teorema. *Si considerino n prove indipendenti di un esperimento casuale a due esiti. Se p è la probabilità di successo in una singola prova e $q = 1-p$, la probabilità che in n prove indipendenti si abbiano esattamente k successi è*

$$\binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, ..., n.$$

Dim. Abbiamo già definito che cosa si intende per eventi indipendenti e abbiamo sopra costruito lo spazio di probabilità dello schema successo-insuccesso. Definiamo:

$$X = \text{"numero di successi nell'ambito di n prove"}.$$

Vogliamo provare che X è una v.a binomiale. Un particolare evento elementare è

$$\{\omega\} = \{(1, ..., 1, 0, ..., 0)\}$$

il che significa: successo nelle prime k prove e insuccesso nelle rimanenti $n-k$. Esso avrà probabilità

$$P(\{\omega\}) = p \cdot p \cdot ... \cdot p \ \cdot q \cdot q ... \cdot q \ = p^k q^{n-k}$$

in virtù dell'indipendenza. Ma questa n-upla è solo *un particolare* modo di avere k successi. Ora, posso etichettare le n prove con $1, ..., n$, e ci sono $C(n; k)$ modi di scegliere k di queste etichette tra le n date: proprio il numero di combinazioni di n oggetti a k a k. Quindi $P(X = k)$ è semplicemente $p^k q^{n-k}$ moltiplicato per questo numero:

$$P(X = k) \equiv f(k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, ..., n$$

$\square$

Si noti che effettivamente la somma di tutte le probabilità è 1:

$$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1^n = 1,$$

per la formula binomiale di Newton. Inoltre

Proposiz. *Media e varianza di una v.a. binomiale* $X \sim B(n,p)$ *sono:*

$$\mu = np, \qquad \sigma^2 = npq.$$

Dim. La funzione generatrice dei momenti di X è

$$G(t) = E(e^{tX}) = \sum_{k=1}^{n} \; e^{tk} \binom{n}{k} p^k q^{n-k} \;\; = \;\; (pe^t + q)^n$$

Perciò i momenti di grado 1 e 2 sono:

$$E(X) = G'(t)|_{t=0} = n(pe^t + q)^{n-1} \cdot pe^t|_{t=0} \;\; = np,$$

$$E(X^2) = G''(t)|_{t=0} = \{ \; n(n-1)(pe^t + q)^{n-2} \cdot (pe^t)^2 + n(pe^t + q)^{n-1} \cdot pe^t \; \}|_{t=0}.$$

Dunque $\mu = E(X) = G'(0) = np,$, mentre

$$\sigma^2 = E(X^2) - \mu^2 = (n^2 - n)p^2 + np - (np)^2 = -np^2 + np = np(1-p).$$

[Ecco la traccia anche di un altro argomento, più intuitivo: il numero X di successi nell'ambito di n prove può scriversi come somma $X_1 + X_2 + ... + X_n$ dove, per $i = 1, 2, ..., n$,

$$X_i = \text{num. di successi nell'ambito della } i\text{−esima prova} .$$

Ma ciascun

$$X_i : \; \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}, \quad i = 1, ..., n$$

ha media e varianza

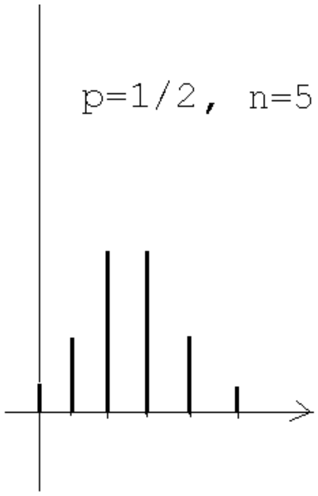
$$E(X_i) = 0 \cdot q + 1 \cdot p = p$$

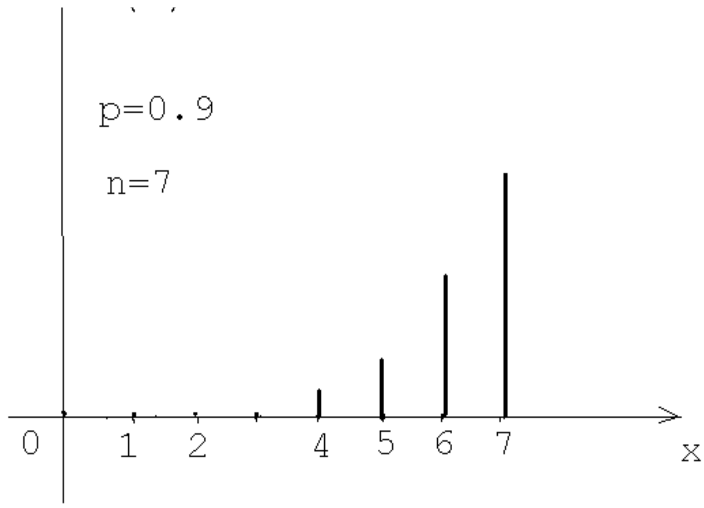
$$\sigma^2 = (0-p)^2 \cdot q + (1-p)^2 \cdot p = p^2 q + q^2 p = pq(p+q) = pq.$$

Poiché, come vedremo, la media di una somma è uguale alla somma delle medie, $E(X) = np$. Vedremo più avanti la definizione di più v.a. "indipendenti", constateremo che le X_i sono indipendenti e vedremo che la varianza di una somma *di v.a. indipendenti* è uguale alla somma delle varianze: dunque $Var(X) = npq$.] $\square$

Es. Se la probabilita' di avere un figlio maschio e' $\frac{1}{2}$, per una famiglia con 5 figli, qual è la probabilità di avere: (i) due maschi (ii) almeno un maschio ? (iii) almeno 3 femmine ? Sia X = "numero di maschi fra $n = 5$ figli:

$$P(X = 2) = \binom{5}{2} (\frac{1}{2})^2 (\frac{1}{2})^3 = 10 \cdot (1/2)^5 = 10/32$$





Es. Sia $p = 90\%$ la probabilità che un test diagnostico dia risposta vera su un individuo. In un gruppo di 7 persone qual è la probabilità che il test dia risposta vera: (i) su tutti e 7; (ii) su almeno 6 ? (iii) su meno della metà ? Qual è il valore atteso di diagnosi veritiere in un gruppo di 75 persone ? con quale deviazione standard ?

Se $X =$ "numero di diagnosi veritiere in un gruppo di n=7 individui"

$$P(X = 7) = \binom{7}{7}(0.9)^7(0.1)^0 = (0.9)^7$$

$$P(X \geq 6) = \binom{7}{6}(0.9)^6(0.1) + \binom{7}{7}(0.9)^7.$$

$$P(X \leq 3) = \sum_{k=0}^{3}\binom{7}{k}(0.9)^k(0.1)^{7-k}.$$

Se $Y =$ "numero di diagnosi veritiere in un gruppo di n=75 individui"

$$E(Y) = np = 75 \cdot (0.9), \quad \sigma = \sqrt{npq} = \sqrt{75 \cdot (0.9)(0.1)}.$$

LEGGE DI PROBABILITÀ DI POISSON

Def. X è una v.a. di Poisson di parametro μ se può assumere gli infinti valori $k = 0, 1, 2, ...$ con probabilità $P(X = k) = f(k) = \frac{\mu^k}{k!}e^{-\mu}$:

$$X : \begin{pmatrix} 0 & 1 & 2 & 3 & ... & k & ... \\ e^{-\mu} & \mu e^{-\mu} & \frac{\mu^2}{2!}e^{-\mu} & \frac{\mu^3}{3!}e^{-\mu} & ... & \frac{\mu^k}{k!}e^{-\mu} & ... \end{pmatrix}$$

Si osservi che effettivamente la somma di tutte le probabilità è 1:

$$\sum_{k=0}^{+\infty}\frac{\mu^k}{k!}e^{-\mu} = e^{\mu} \cdot e^{-\mu} = 1,$$

essendo $\sum_{k=0}^{+\infty}\frac{x^k}{k!} = e^x$ (è la ben nota serie esponenziale). Inoltre (fatto curioso) la media è uguale alla varianza:

Prop. *Se X è una v.a. di Poisson con parametro μ, allora $E(X) = \mu$ e $Var(X) = \mu$.*
Dim. La funzione generatrice dei momenti è:

$$G(t) \equiv E[\ e^{tX}\ ] = \sum_{k=0}^{\infty} e^{tk} e^{-\mu} \frac{\mu^k}{k!} = e^{-\mu} \exp[\mu e^t].$$

Perciò

$$E(X) = G'(t)|_{t=0} = e^{-\mu} \mu e^t \exp[\mu e^t]|_{t=0} = \mu.$$

Inoltre

$$E(X^2) = G''(t)|_{t=0} = \{\ e^{-\mu} \mu e^t \exp[\mu e^t] + e^{-\mu} (\mu e^t)^2 \exp[\mu e^t]\ \}|t = 0$$

. Perciò

$$Var(X) = E(X^2) - \mu^2 = \mu + \ \mu^2 - \ \mu^2 = \mu.$$

□

La v.a. di Poisson è un buon modello per il numero di fenomeni casuali distribuiti con una data densità media μ nell'unità di tempo o nell'unità di volume o nell'unità di superficie: numero di chiamate a un centralino telefonico per minuto; numero di automobili a un casello per ora; numero di infortuni stradali per settimana; numero di stelle per unità di volume abbastanza grande...

Es. Nel 1910 Rutherford e Geiger provarono che il numero di particelle α emesse al secondo da una sostanza radioattiva era una v.a. di Poisson con $\mu = 0.5$. Qual è la probabilità di osservare due o più particelle durante un secondo ?

$$P(X \geq 2) = \sum_{k=2}^{+\infty} \frac{(0,5)^k}{k!}\ e^{-\mu} =$$

$$= 1 - P(X = 0) - P(X = 1) = 1 - e^{-0.5} - 0.5 \cdot e^{0.5} = 1 - 0.91 = 9\%$$

Es. Una certa sospensione batterica contiene 5 batteri per cm^3 (valor medio). Qual è la probabilità che un campione causale di 1 cm^3 contenga (i) nessun batterio (ii) al piú due batteri (iii) almeno 5 batteri ?

$$P(X = 0) = e^{-5}; \quad P(X \leq 2) = e^{-5}(1 + 5 + \frac{25}{2!})$$

$$P(X \geq 5) = 1 - P(X \leq 4) = 1 - e^{-5}(1 + 5 + \frac{25}{2!} + \frac{125}{6} + \frac{625}{24})$$

Es. Per valutare il numero di batteri in una sospensione se ne cerca la diluizione limite alla quale si trova ancora almeno un batterio capace di riprodursi. Ad es. se diluendo acqua di canale con fattore $\frac{1}{10}$, $\frac{1}{100}$, $\frac{1}{10^3}$, $\frac{1}{10^4}$ troviamo, dopo incubazione, sviluppo dei batteri; mentre troviamo sterile la diluizione con fattore $\frac{1}{10^5}$; allora grossolanamente diremo che quel canale conteneva circa 10.000 germi per cm^3. Per raffinare, usiamo la distribuzione di Poisson e inoculiamo in 20 tubi la sospensione diluita con fattore $\frac{1}{10^4}$. Se vi sono in media μ germi per cm^3 di diluito, vi sarà una proporzione $P(X = 0) = e^{-\mu}$ di tubi che non riceveranno alcun germe e perciò saranno sterili. Poniamo di trovare sterili 12 tubi su 20. Ebbene, avremo $e^{-\mu} = \frac{12}{20} = 0.6$ cioè $\mu = -log(0.6) = -(log_e 10) \cdot log_{10}(0.6) = -2.3026 \cdot (-0.222) = 0.51$. Allora la concentrazione di germi nel canale è $0.51 \cdot 10^4 = 5.1 \cdot 10^3$ germi per cm^3.

ALTRE LEGGI DI PROBABILITÀ DISCRETE

Descriviamo la *legge di probabilità ipergeometrica* e poi la *legge geometrica.*

Problema. Da un'urna contenente b palline bianche ed r rosse, se ne estraggono n $(n \leq b+r)$ senza reimmissione. Qual è la probabilità che esattamente k di esse siano rosse?
Risposta. Supponiamo che le palline siano numerate da 1 a $b+r$ e che le palline rosse siano quelle con i numeri $\leq r$. Lo spazio Ω, degli eventi elementari è l'insieme di tutti i sottoinsiemi $\omega = \{\omega_1, ..., \omega_n\}$ di $\{1, 2, ..., b+r\}$. Quindi lo spazio di probabilità Ω è lo spazio delle combinazioni semplici di $b+r$ oggetti ad n ad n :

$$\#\Omega \ = \ C(b+r; n).$$

Se poniamo

$$A_k = \{\omega : \ \omega \text{ ha esattamente } k \text{ elementi con indice } \ \leq r\}$$

la probabilità richiesta è il quoziente

$$P(A_k) = \frac{\# \ A_k}{\# \ \Omega} \ = \ \frac{C(r;k) \cdot C(b; n-k)}{C(b+r; n)} = \frac{\binom{r}{k}\binom{b}{n-k}}{\binom{b+r}{n}}$$

La v.a. $X =$ "numero di palline rosse estratte nell'ambito di n estratte senza restituzione", sapendo che "il numero di rosse è r su un totale inziale di $N = r+b$" è una v.a. *ipergeometrica con parametri* r, N, n. *Essa può assumere i valori* $k = 0, 1, 2, ..., r$, *dove* $n \leq N = b+r$. *La sua funzione massa di probabilità è*

$$P(X=k) = \frac{\binom{r}{k}\binom{b}{n-k}}{\binom{b+r}{n}}, \quad k = 0, 1, ..., r.$$

Es. : qual è la probabilità del trio $\{5, 51, 63\}$ nel gioco del lotto?
Risposta: si dividono gli $N = 90$ numeri in due classi: da una parte gli $r = 3$ numeri indicati, dall'altra parte gli altri $b = 87$. Inoltre ci sono $n = 5$ estrazioni senza reimmissione (quindi non indipendenti, qui non serve la v.a. binomiale). La v.a.

$$X = \ \text{numero di estratti dal primo gruppo nell'ambito di 5 estrazioni}$$

$$P(X=k)=\frac{\binom{r}{k}\binom{b}{n-k}}{\binom{b+r}{n}}=\frac{\binom{3}{3}\binom{87}{5-3}}{\binom{90}{5}}=\frac{1}{11748}=0.0085\%$$

Problema Un dado viene lanciato più volte finché non si ottiene 6. Qual è la probabilità che occorrano esattamente k lanci?
Risposta. È la probabilità che per $k-1$ lanci esca "insuccesso" ed esca "successo" la $k-$esima volta: se T è il numero di lanci necessari ad avere successo,

$$P(T=k)=p(1-p)^{k-1}=\frac{1}{6}(\frac{5}{6})^{k-1}, \quad k=1,2,...$$

Più in generale, in ogni schema successo-insucccesso nel quale la probabilità di successo in una singola prova sia $p\in(0,1)$, si può definire *la v.a. geometrica*

$$T := \textit{num. di prove necessarie ad avere successo.}$$

Diversamente dalla v.a. binomiale, i valori possibili di T sono tutti i numeri naturali:

$$T: \begin{pmatrix} 1 & 2 & ... & k & ... \\ p & pq & ... & pq^{k-1} & ... \end{pmatrix}.$$

La funzione di probabilità è semplicemente

$$f(k)=P(T=k)=p\cdot q^{k-1},\ q=1-p, \quad k=1,2,3,...$$

Sommando questi termini si ottiene la serie geometrica:

$$\sum_{k=1}^{\infty} f(k)=p\sum_{k=1}^{\infty} q^{k-1}=p\sum_{j=0}^{\infty} q^j=\frac{p}{1-q}=\frac{p}{1-(1-p)}=1.$$

La v.a. geometrica si può anche riguardare come un tempo di attesa, misurato in numero di prove:

$$T := \text{tempo di attesa del primo successo.}$$

[Invece qualche autore chiama v.a. geometrica $X=T-1$, cioè "numero di insuccessi precedenti il primo successo", con funzione di probabilità $P(X=j)=pq^j,\ \ j=0,1...$].

Es. Un arciere ha probabilità $\frac{1}{3}$ di far centro in un bersaglio. Trovare la probabilità che gli occorra un numero di prove maggiore di 3.
Sia $T=$"numero di prove necessarie al primo centro nel bersaglio", sapendo che la probabilità di far centro è 1/3. Allora T è geometrica di parametro 1/3. L'evento $T>3$ ha probabilità

$$P(T\geq 4)=1-P(T=1)-P(T=2)-P(T=3)=$$

$$= 1 - \frac{1}{3} - \frac{1}{3} \cdot \frac{2}{3} - \frac{1}{3} \cdot (\frac{2}{3})^2 =$$

$$= 1 - \frac{1}{3}[1 + 0.666 + 0.444] = 1 - (2.111)/3 \simeq 29.6\%$$

LEGGE DI PROBABILITÀ NORMALE O DI GAUSS

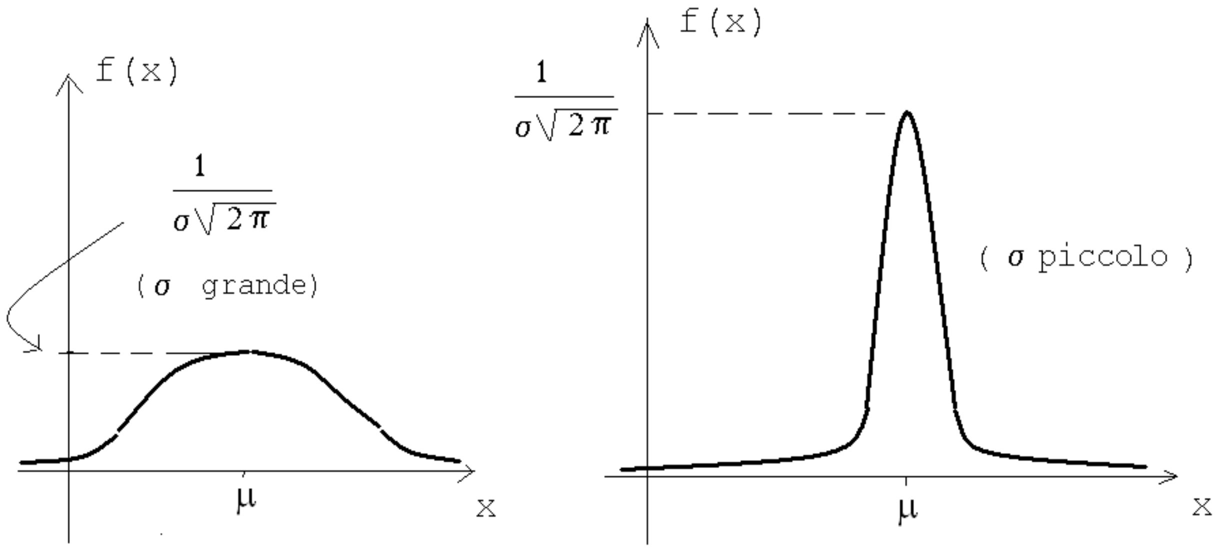
Def. Siano $\mu \in R$, $\sigma > 0$. La v.a. continua X è *normale* con parametri μ, σ, e si scrive $X \simeq N(\mu, \sigma^2)$, sse la densità è:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-(x-\mu)^2/2\sigma^2}.$$

La funzione $f(x)$ è detta funzione di Gauss. È la funzione "a campana" simmetrica rispetto ad $x_\circ = \mu$, che ha un max. per $x_\circ = \mu$ e ivi assume il valore massimo $f(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$. Quest'ultimo ha il significato di fattore di normalizzazione, cioè è quel numero tale che

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

(questa uguaglianza si può dimostrare e dice che f è una densità di probabilità: vuol dire che $P(-\infty < X < +\infty) = 1$).



Si dimostra che la v.a. X ha media μ e varianza σ^2. Si nota che più è piccolo σ, più è alto il picco $f(\mu)$, e dunque è più concentrata la campana intorno alla media μ : ciò concorda perfettamente con il significato di varianza che possiede σ^2. Ecco il calcolo di media e varianza:

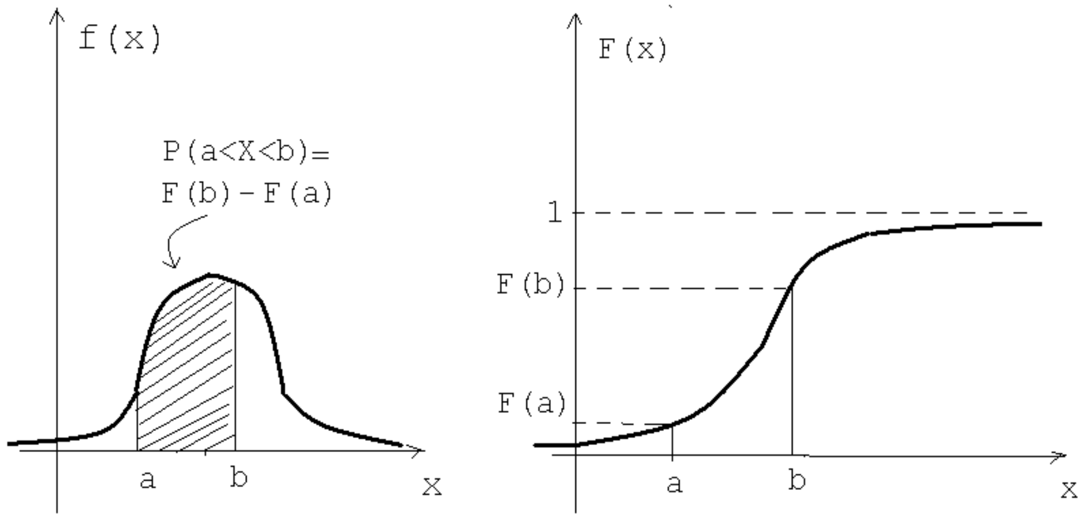
Lemma.

$$E(X) = \mu, \quad Var(X) = \sigma^2.$$

Dim. La media di X è ovviamente μ a causa della simmetria del grafico della densità attorno ad $x_\circ = \mu$.
Per la varianza basta moltiplicare e dividere per $-\sigma^2$ e integrare per parti riconoscendo $-\frac{(x-\mu)}{\sigma^2} \cdot e^{-(x-\mu)^2/2\sigma^2}$ come fattore differenziale :

$$\int_{-\infty}^{+\infty} (x-\mu)^2 e^{-(x-\mu)^2/2\sigma^2} dx = \lim_{R\to\infty} (-\sigma^2)[(x-\mu)\frac{e^{-(x-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}}]_{-R}^{R} -$$

$$\int^{+\infty}$$



La funzione integrale F non si può calcolare coi metodi di integrazione elementari. Tuttavia, detta

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}\, e^{-u^2/2}\, du$$

la funzione distribuzione della **v.a. normale standardizzata** , cioè la v.a. normale con media 0 e varianza 1, si ha:

Proposizione La funzione distribuzione normale di media μ e varianza σ^2 si può rappresentare in termini della funzione distribuzione normale Φ di media 0 e varianza 1 nel seguente modo:

$$F(x) = \Phi(\frac{x-\mu}{\sigma}).$$

In particolare:

$$P(a < X \leq b) \equiv F(b) - F(a) = P(\frac{a-\mu}{\sigma} < \ N(0;1) \ \leq \frac{b-\mu}{\sigma}).$$

Dim. Ponendo $\frac{t-\mu}{\sigma} = u$, si ottiene $\frac{du}{dt} = \frac{1}{\sigma}$, $dt = \sigma du$, e quindi

$$F(x) \equiv \lim_{R \to -\infty} \int_R^x \frac{1}{\sigma\sqrt{2\pi}} \ e^{-(t-\mu)^2/2\sigma^2} \ dt$$

$$= \lim_{R \to -\infty} \int_{\frac{R-\mu}{\sigma}}^{\frac{x-\mu}{\sigma}} \frac{1}{\sigma\sqrt{2\pi}} \ e^{-u^2/2} \sigma du = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \equiv \Phi(\frac{x-\mu}{\sigma}).$$

Conseguenza: se X è normale, anche $(X-\mu)/\sigma$ è normale. Inoltre, come ogni standardizzata, $(X-\mu)/\sigma$ è standard. Quindi $(X-\mu)/\sigma$ è normale standard, cioè $N(0;1)$.

Oss. Di conseguenza si usano le tavole della normale standard $N(0;1)$ per calcolare le probabilità relative a una normale generica $N(\mu, \sigma^2)$. Le tavole di $N(0;1)$ si possono usare in due modi:

1) dato un valore $z \in R$, si cerca la probabilità $P(N(0;1) \leq z) = \Phi(z)$;

2) data una probabilità α (a volte assegnata come percentuale) si cerca il valore $z \in R$ tale che $\alpha = P(N(0;1) \leq z)$. [Tale z è denotato ϕ_α, e chiamato quantile relativo ad α, ovvero percentile $n-$esimo se $\alpha = n/100$].

Ricerche di quantità similari sono riconducibili alla tavola di $N(0,1)$ tramite la proposizione precedente e considerazioni geometriche sulle aree sottese al grafico della densità: ad es. $P[N(0;1) \leq -1.7] = 1 - P[N(0;1) \leq 1.7]$; altro esempio: il quantile $\phi_{0.95}$ è uguale all'opposto del quantile $\phi_{0.05}$.

Es. Sia $X \simeq N(0.8;4)$ ossia X è normale con media 0.8 e varianza 4. Calcoliamo a modo di esempio:

$$P(X \leq -1.16) = P[N(0;1) \leq \frac{-1.16-0.8}{2}) = P[N(0;1) \leq -0.98] =$$

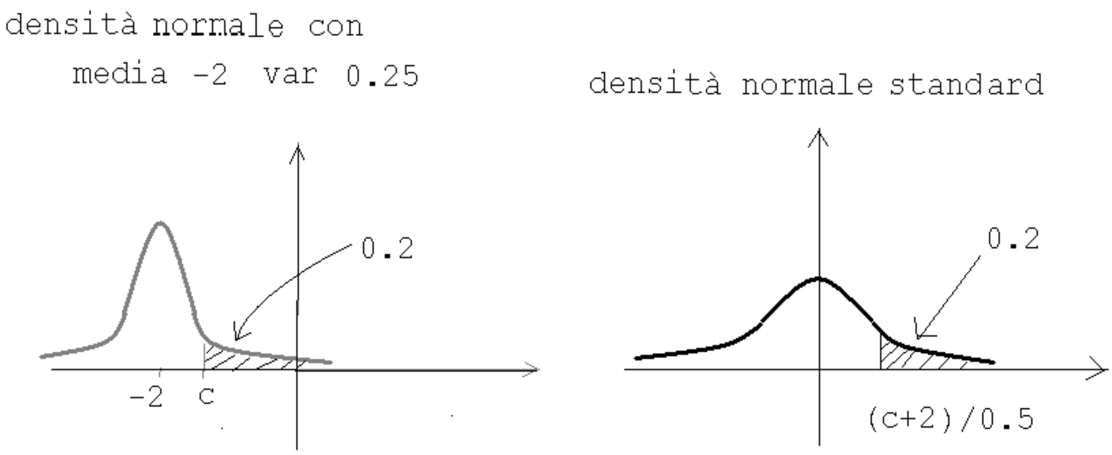
$$= 1 - P[N(0;1) \leq 0.98] = 16.35\%$$

$$P(X \geq 1) = 1 - P[N(0;1) \geq (1-0.8)/2] = 1 - P[N(0;1) \leq 0.1] = 46.02\%$$

$$P(2 \leq X \leq 3) = P[(2-0.8)/2 < N(0;1) \leq (3-0.8)/2] =$$

$$= \Phi(1.1) - \Phi(0.6) = 13.86\%$$

Es. Sia $X \simeq N(-2;\ 0.25)$. Determinare $c \in R$ tale che
(1) $P(X \geq c) = 0.2$; (2) $P(-2- \ c \leq X \leq -2+ \ c) = 0.90$ (3) $P(X \leq c) = 0.43$;



$$(2) \quad 0.9 = P[(-2-c+2)/0.5 \le N(0;1) \le (-2+c+2)/0,5] = \Phi(2c) - \Phi(-2c) =$$

$$= \Phi(2c) - (1-\Phi(2c)) = 2\Phi(2c) - 1;$$

$$\Phi(2c) = (0.9+1)/2 = 0.95; \quad 2c = \phi_{0.95} = 1.64; \quad c = 0.82$$

$$(3) \qquad 0.43 = P(X \le c) = P[N(0;1) \le \frac{c+2}{0.5}]$$

da cui, essendo $43\% < 50\%$, abbiamo (achtung!) che:

$$\frac{c+2}{0.5} = \phi_{0.43} \textbf{ è negativo !!! } \text{ed uguale a} \qquad = -\phi_{0.57} = -0.18$$

e infine $c = -2.09$.

OSS. Quanto spesso [=con quale probabilità] la variabile si discosta dalla media al più per i primi tre multipli della deviazione standard?

$$P(\mu - \sigma \le X \le \mu + \sigma) = P(-1 \le N(0;1) \le 1) \simeq 68\%$$

$$P(\mu - 2\sigma \le X \le \mu + 2\sigma) = P(-2 \le N(0;1) \le 2) \simeq 95,5\%$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = P(-3 < N(0;1) < 3) \simeq 99,7\%.$$

Sarà utile ricordare i quantili $\phi_{0.975}$ e $\phi_{0.995}$ della normale standard:

$$P(\mu - 1,96\sigma < X < \mu + 1,96\sigma) = 95\% \quad \text{perché } 1.96 = \phi_{0.975}$$

$$P(\mu - 2,58\sigma < X < \mu + 2,58\sigma) = 99\% \quad \text{perché } 2.58 = \phi_{0.995}.$$

Enunciamo senza dimostrazione i seguenti importanti teoremi.

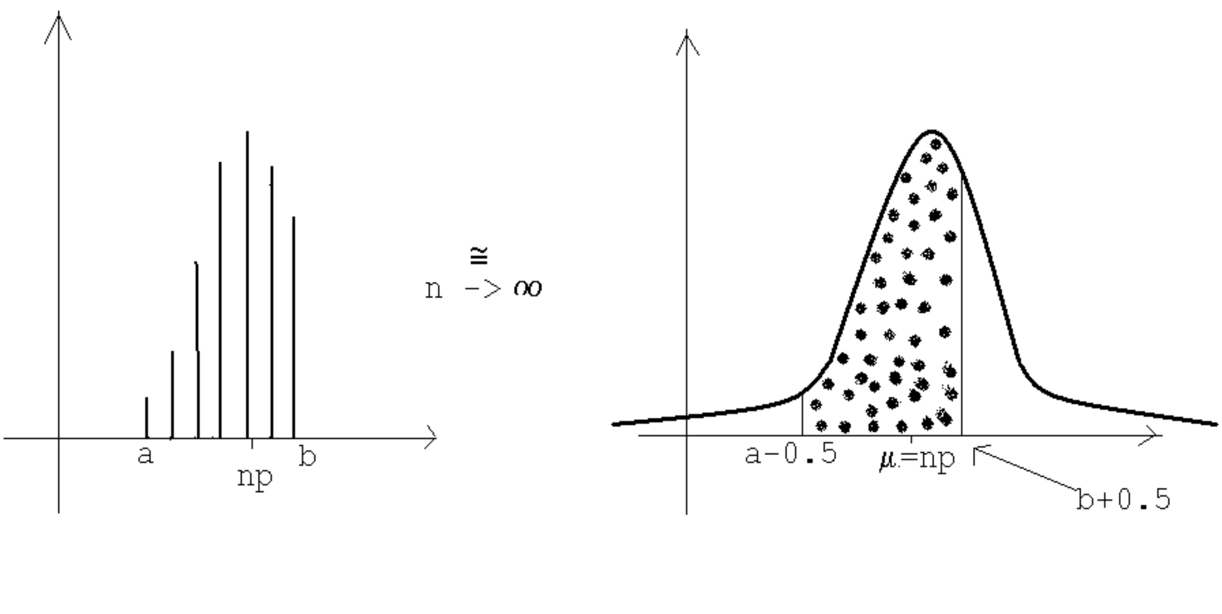
Teorema Se X è normale con media μ e varianza σ^2, allora $X^* = c_1 X + c_2$ $(c_1 > 0)$ è normale con media $\mu^* = c_1\mu + c_2$ e varianza $(\sigma^*)^2 = c_1^2\sigma^2$.

Teorema di approssimazione di De Moivre e Laplace Siano a, b interi qualunque non negativi. Sia X la v.a. binomiale di parametri n, p . Sia Y la v.a. normale avente media np e varianza npq e sia Z la v.a. normale standardizzata, cioè $Z = (Y - np)/\sqrt{npq}$. Vale l'approssimazione:

$$P(a \leq X \leq b) \sim P(a - 0.5 \leq Y \leq b + 0.5) \quad \text{per} \quad n \to \infty$$

$$\text{cioè} \;\; P(a \leq X \leq b) \simeq P(\frac{a - 0.5 - np}{\sqrt{npq}} \leq Z \leq \frac{b + 0.5 - np}{\sqrt{npq}}) \;\; \text{per} \;\; n \to \infty$$

dove " $\sim$ " significa che il quoziente tra le due espressioni tende ad 1 per $n \to \infty$.



OSS. Il termine ± 0.5 è una correzione dovuta al passaggio da una v.a. discreta a una continua. Il caso più evidente: se $a = b$, avremmo "probabilità $= 0$" per il semplice fatto di usare una v.a. continua; si rimedia prendendo $[a-0.5, a+0.5]$, intervallo lungo esattamente 1, che funge da base al rettangolo avente per altezza il valore della densità, che è circa la probabilità binomiale di a; così l'area del rettangolo approssima la probabilità binomiale $P(X = a)$.

Es. *(approssimazione della Binomiale alla Normale).*Il 10% di bulloni prodotti da una certa macchina è difettoso. Trovare la probabilità che, in un campione casuale di 400, al massimo 30 siano difettosi.
$X \sim B(400, \frac{1}{10})$ ha media $\mu = np = 40$, e varianza $\sigma^2 = npq = 36$. Essendo $min\{np, nq\} \geq 5$ ed $n > 50$ è lecita l'approssimazione normale:

$$P(X \leq 30) \simeq P[N(np, npq) \leq \frac{30.5 - np}{\sqrt{npq}}] =$$

$$= P[N(0,1) \leq -1.58] = 1 - P[N(0,1) \leq 1.58] = 1 - 0.9429 = 0.0571$$

Es. *(approssimazione della Binomiale alla Normale).* Determinare la probabilità di ottenere più di 25 "sette" in 100 lanci di una coppia di dadi equi.

La v.a. $X =$ "numero di 'sette' nell'ambito di cento lanci" è binomiale con parametri $n = 100$ e $p = \frac{6}{36} = \frac{1}{6}$. Se Y è normale con la stessa media np e la stessa varianza npq abbiamo:

$$P(X \geq 26) \simeq P[N(0,1) \geq \frac{25.5 - 100/6}{\sqrt{500/36}}) \simeq 1\%$$

OSS. Sussiste anche un'approssimazione della v.a. binomiale alla v.a. di Poisson, sia pure in un diverso regime: la Poissoniana di media μ si può ottenere come caso limite della Binomiale se

$$p \to 0, \quad \text{ed} \quad n \to \infty \quad \text{in modo tale che la media} \quad np \to \mu.$$

In questo regime, in sostanza, n è grande e p è piccolo in modo che $np \simeq npq$ (media $\simeq$ varianza). L'idea che giustifica questa approssimazione è molto semplice: consideriamo una v.a. binomiale $X \sim B(n, \frac{\lambda}{n})$ e studiamo il suo comportamento per $n \to \infty$:

$$P(X = k) = \binom{n}{k} (\frac{\lambda}{n})^k (1 - \frac{\lambda}{n})^{n-k} = \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} (1 - \frac{\lambda}{n})^{n-k} =$$

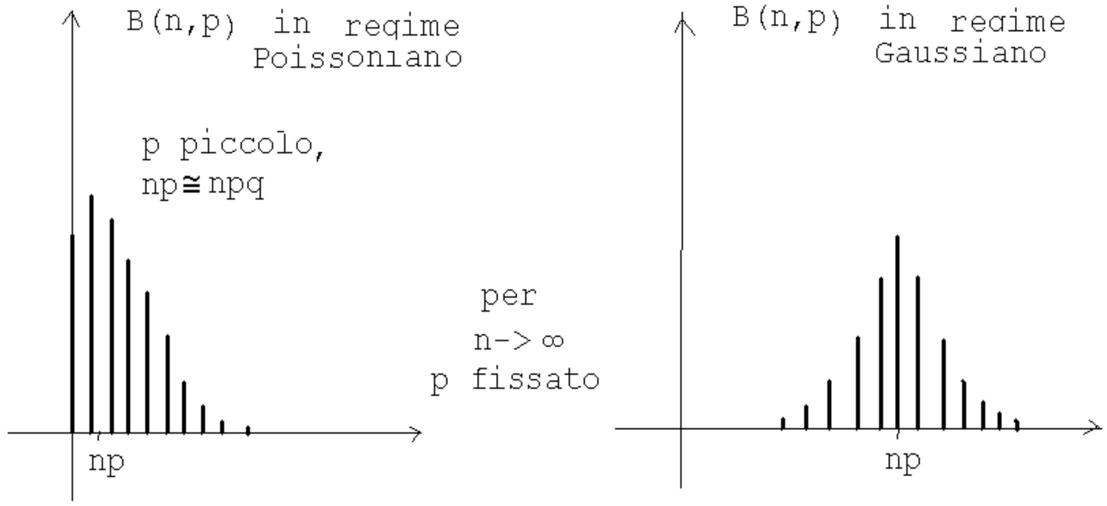
$$= \frac{\lambda^k}{k!} (1 - \frac{\lambda}{n})^n \frac{n(n-1)...(n-k+1)}{n^k} (1 - \frac{\lambda}{n})^{-k} \quad \to \quad \frac{\lambda^k}{k!} e^{-\lambda}$$

dove abbiamo i limiti:

$$\lim_{n\to\infty} (1 - \frac{\lambda}{n})^n \; = \; e^{-\lambda}$$

$$\lim_{n\to\infty} \frac{n(n-1)...(n-k+1)}{n^k} \;=\; 1$$

(il limite per $n \to \infty$ del rapporto di due polinomi aventi lo stesso grado è uguale al quoziente dei coefficienti del termine di grado massimo, qui entrambi uguali ad 1).
Per questo la Poissoniana è detta a volte "la v.a. degli eventi rari": infatti vive in un regime Poissoniano ogni Binomiale con p molto piccolo ed n grande ma non troppo. Tuttavia, se n cresce ulteriormente in rapporto a p, si entra allora nell'altro regime: la funzione di probabilità binomiale diventa sempre più simmetrica e sempre più simile alla funzione densità di Gauss. Ai fini pratici siamo già in regime "Gaussiano" se $n \geq 50$ ed $\min\{np, nq\} \geq 5$.



Es. *(approssimazione della Binomiale alla Poissoniana).*
Un'azienda vende un preparato in partite di 200 confezioni con la garanzia che tutte siano non difettose; se la probabilità che una confezione sia non difettosa è 0.5%, con quale probabilità una partita viola la garanzia ?

La v.a. X = "numero di confezioni difettose nell'ambito di 200" è binomiale con parametri $n = 200$, $p = \frac{5}{1000}$. Essa è bene approssimata dalla Poissoniana Y di media $np = 200\frac{5}{1000} = 1$. Quindi:

$$P(X \geq 1) \simeq P(Y \geq 1) = 1 - P(Y = 0) = 1 - e^{-1} \simeq 63\%$$

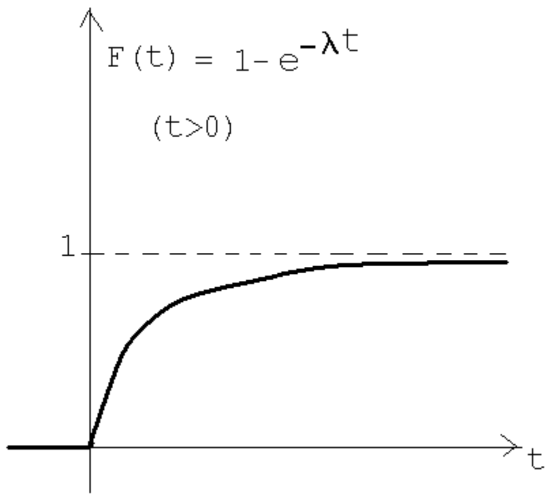
LEGGE DI PROBABILITÀ ESPONENZIALE; LEGGI GAMMA

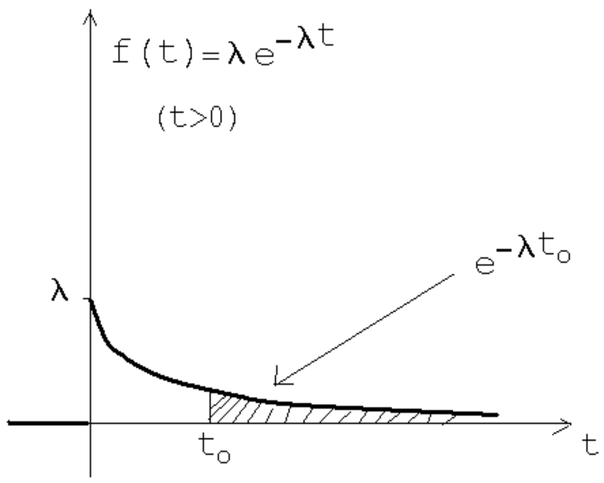
Consideriamo una sequenza di eventi, sapendo che il numero di eventi nell'unità di tempo è una v.a. di Poisson di media λ. Quanto tempo deve aspettare un osservatore della sequenza poissoniana di eventi, per osservare il primo verificarsi dell'evento?

Teorema (v.a. esponenziale). *Se il numero di occorrenze di un evento nell'intervallo di tempo* $[0, t]$ *è v.a. di Poisson con media* $t\lambda$, $\forall t > 0$, *allora l'istante* T *del primo verificarsi dell'evento è una v.a. "esponenziale" con parametro* λ*:*

$$F_T(t) = \begin{cases} 0, & t < 0 \\ 1 - e^{-\lambda t}, & t > 0; \end{cases} \qquad f_T(t) = \begin{cases} 0, & t < 0 \\ \lambda e^{-\lambda t}, & t \geq 0. \end{cases}$$

La v.a. esponenziale ha media $\frac{1}{\lambda}$ *("tempo medio di attesa") e varianza* $\frac{1}{\lambda^2}$.





Dim. Indichiamo con $F_T(t)$ la probabilità $P(T \leq t)$. Allora $1 - F_T(t)$ è la probabilità che l'istante di prima occorrenza sia maggiore di t. Ovvero, $1 - F_T(t)$ è la probabilità che il numero di eventi occorrenti da 0 a t sia zero. Poiché il numero di occorrenze in $[0, t]$ è di Poisson con media λt,

$$1 - F_T(t) = e^{-\lambda t}, \quad \forall t > 0$$

Derivando si ottiene la densità di T: $F_T'(t) = \lambda e^{-\lambda t}, \forall t > 0$. Media e varianza si ottengono integrando per parti:

$$E(T) = \int_0^\infty t f(t) dt = \int_0^\infty t\lambda e^{-\lambda t} dt =$$

$$= [-te^{-\lambda t}]_0^{+\infty} + \int_0^{+\infty} e^{-\lambda t} dt = \frac{1}{\lambda}.$$

$$Var(T) = E(X^2) - \mu^2 = \int_0^\infty t^2 \lambda e^{-\lambda t} \, dt - (1/\lambda)^2$$

$$= [-t^2 e^{-\lambda t}]_0^{+\infty} + \int_0^{+\infty} 2te^{-\lambda t} dt - \frac{1}{\lambda^2} = 2\frac{1}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

□

Teor. (v.a. "Gamma"). *Nelle ipotesi del teorema precedente, il tempo di attesa T_r dell'r-esimo verificarsi dell'evento ($r = 1, 2, ...$) segue una legge "Gamma" con parametri r e λ, denotata $\Gamma(r, \lambda)$:*

$$f_{T_r}(t) = \begin{cases} \lambda^r \frac{t^{r-1}}{(r-1)!} e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

Inoltre le v.a. Γ sono legate tra loro da una proprietà additiva: se T_r, T_s sono indipendenti e di legge $\Gamma(r, \lambda)$ e $\Gamma(s, \lambda)$ rispettivamente,

$$T_r + T_s \sim \Gamma(r + s, \lambda), \quad \forall r, s \geq 0.$$

Infine una legge $\Gamma(r, \lambda)$ ha media r/λ e varianza $r/(\lambda)^2$.
Dim. (cenno) Sia T_2 = "istante del secondo verificarsi dell'evento". Allora la probabilità che "$T_2 > t$" è la probabilità che in $[0, t]$ il numero di occorrenze sia 0 o 1:

$$P(T_2 > t) \equiv 1 - F_{T_2}(t) = e^{-\lambda t} + \frac{\lambda t}{1!} e^{-\lambda t}$$

$$\Longrightarrow \quad F'_{T_2}(t) = \frac{d}{dt}[-(e^{-\lambda t} + \lambda t e^{-\lambda t})] = \lambda e^{-\lambda t} - \lambda e^{-\lambda t} + \lambda^2 t e^{-\lambda t}, \quad t > 0$$

che è proprio una densità $\Gamma(2, \lambda)$...
Media e varianza si potrebbero calcolare con la definizione, oppure usando la funzione generatrice dei momenti (e usando l'identit ben nota:

$$\int_0^\infty t^{r-1} e^{-t} dt = (r-1)! \quad \forall r \in N \ ).$$

$\square$

OSS. Il nome "Gamma" di questa distribuzione viene dalla omonima funzione speciale cosi' definita: $\Gamma : R^+ \to R^+$,

$$\Gamma : R^+ \to R^+, \ \ \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx.$$

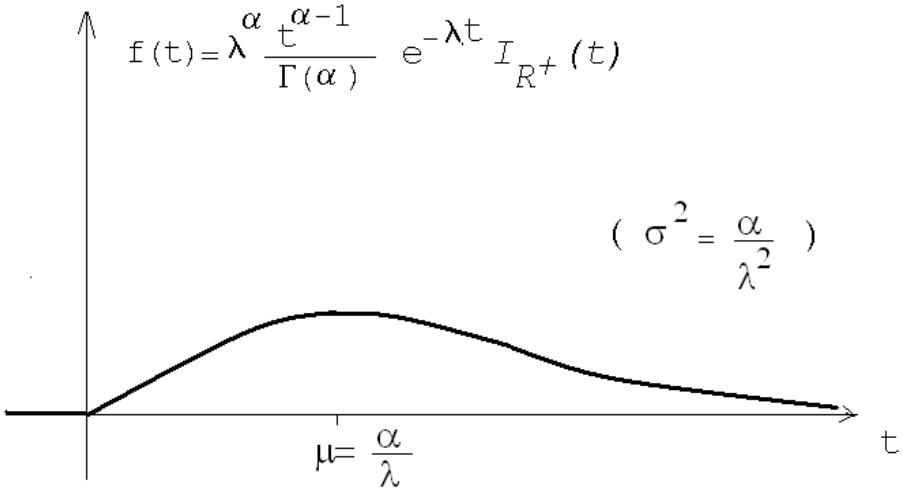
Si osserva subito che $\Gamma(1) = \int_0^{+\infty} e^{-x} dx = 1$ e, per ogni $\alpha > 0$, integrando per parti si ottiene

$$\Gamma(\alpha + 1) = \int_0^{+\infty} x^\alpha e^{-x} dx = [-x^\alpha e^{-x}]_0^\infty + \alpha \int_0^{+\infty} x^{\alpha-1} e^{-x} dx = \alpha \Gamma(\alpha).$$

In particolare $\Gamma(2) = 1$, $\Gamma(3) = 2!$,... ossia:

$$\forall n \in N, \quad \Gamma(n) = (n-1)!$$

Ovviamente si può estendere la definizione di legge Γ da $\alpha \in N$ ad ogni $\alpha > 0$: basta sostiture $(\alpha - 1)!$ con $\Gamma(\alpha)$.



Def. *Sia $m \geq 2$. Le v.a. $X_1, X_2, ..., X_m$ sono indipendenti se, per ogni scelta di insiemi misurabili $A_1, A_2, ..., A_m \subset R$, si ha:*

$$P[X_1 \in A_1, ..., X_m \in A_m] = P[X_1 \in A_1] \cdot ... \cdot P[X_m \in A_m]$$

Diremo che le v.a. $X_1, ..., X_m, ...$ (in numero infinito) sono indipendenti se, per ogni $m \geq 2$, risultano indipendenti tra loro le v.a. $X_1, ..., X_m$.

Se prendiamo ad es. due v.a. X, Y, l'indipendenza significa che $P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)$ per ogni coppia di insiemi misurabili $A, B \subset R$. Si può esprimere la stessa cosa anche in termini di funzioni distribuzione. Infatti, definiamo una funzione "distribuzione congiunta" per la coppia di v.a. X, Y:

$$F_{X,Y}(x, y) := P(X \leq x, \ Y \leq y).$$

Allora le v.a. X, Y sono indipendenti se e solo se $P(X \leq x, \ Y \leq y) = P(X \leq x) \cdot P(Y \leq y)$, cio se e solo se

$$F_{X,Y}(x, y) = F_X(x) F_Y(y), \quad \forall x, y \in R.$$

Quindi due v.a. X, Y sono indipendenti sse la loro funzione distribuzione congiunta può scriversi come prodotto delle singole funzioni distribuzione.

Una propriet importante di una somma di v.a. indipendenti il sommarsi delle varianze:

Teorema. *Se X, Y sono v.a. qualsiasi, la media della somma uguale alla somma delle medie:*

$$E(X + Y) = E(X) + E(Y).$$

Se X, Y sono v.a. indipendenti, la varianza della somma uguale alla somma delle varianze:

$$Var(X \pm Y) = Var(X) + Var(Y).$$

Se poi X, Y, oltre ad essere indipendenti, sono normali, allora normale ogni combinazione lineare:

$$c_1 X + c_2 Y \ \sim \ N(c_1 \mu_X + c_2 \mu_Y; \ c_1^2 \sigma_X^2 + c_2^2 \sigma_Y^2).$$

Es. (a) Siano $X_1, X_2, ..., X_n$ v.a. indipendenti tutte di Poisson con parametro $\lambda = 3$. Si domandano media e deviazione standard di $Z = X_1 + X_2 + ... + X_n$.
Risposta:

$$E(Z) = 3 + 3 + ... + 3 = 3n; \qquad \sigma_Z = \sqrt{3 + 3 + ... + 3} = \sqrt{3n} =$$

Es. (b) Siano $X, Y \sim N(-1, \ 16)$ e indipendenti. Si domanda la probabilità $P(X > \frac{1}{3}Y)$.
Risposta: Si ha

$$P(X > \frac{1}{3}Y) = P(X - \frac{1}{3}Y > 0)$$

Ma per il teorema precedente $X - \frac{1}{3}Y$ normale; inoltre la media è la combinazione lineare delle medie; infine per l'indipendenza la varianza è la somma delle varianze:

$$X - \frac{1}{3}Y \sim N(-1 + \frac{1}{3},\ 16 + (\frac{1}{3})^2 \cdot 16) = \ N(-0.66;\ 17.77)$$

Quindi la probabilit richiesta

$$P[N(0,1) > \frac{0 + 0.66}{\sqrt{17.77}}] = P[N(0,1) > 0.15] = 1 - 0.559 = 0.441$$

Per enunciare l'importante teorema di limite centrale ci serve la suddetta nozione di indipendenza, ma anche un'altra nozione: la "convergenza in distribuzione".

Def. *La successione di v.a. reali $\{X_n\}_n$ converge in distribuzione (o "debolmente", o "in legge") alla v.a X se e solo se, dette F_n ed F le rispettive funzioni distribuzione, si ha*

$$\lim_{n \to \infty} F_n(x) = F(x)$$

per ogni punto $x \in R$ di continuità per F.

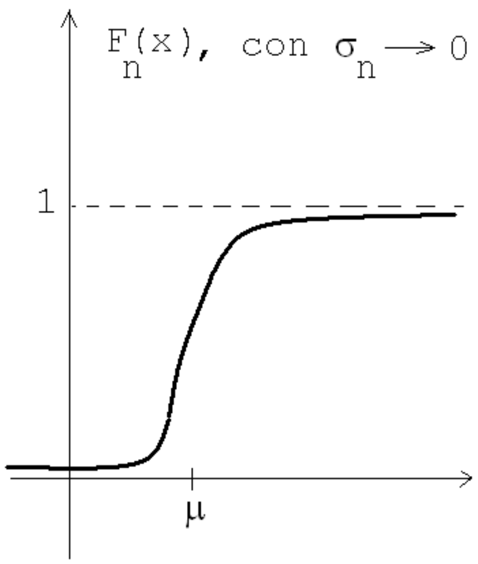
Es. Prendiamo una successione di v.a. normali con media costante ma con varianza tendente a zero:

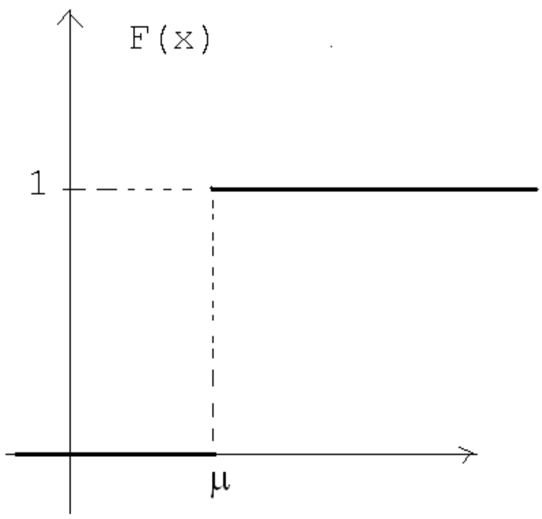
$$X_n \sim N(\mu, \sigma_n^2), \text{ dove } \ \sigma_n \to 0.$$

Allora la funzione distribuzione $F_n(x)$ di X_n converge in tutti i punti tranne in $x = \mu$:

$$\exists \ \lim_{n \to \infty} F_n(x) = F(x) = \begin{cases} 0, & x < \mu \\ 1, & x > \mu \end{cases}$$

Tutti vedono che $F(x)$, in qualunque modo venga definita per $x = \mu$, è la funzione distribuzione della v.a. costante $X = \mu$. Ebbene, la convergenza delle funzioni distribuzione





Teorema di limite centrale *Sia $\{X_n\}_n$ una successione di v.a. indipendenti e identicamente distribuite, di media μ e varianza finita $\sigma^2 > 0$. Allora la loro somma $n-$esima standardizzata*

$$S_n^* = \frac{X_1 + ... + X_n - n\mu}{\sigma\sqrt{n}}$$

converge in distribuzione ad una v.a. $N(0,1)$.

OSS. Il teorema di limite centrale essenzialmente dice che se un fenomeno aleatorio può essere riguardato come sovrapposizione di un gran numero di fenomeni aleatori indipendenti, aventi ciascuno una qualsiasi legge dello stesso tipo, allora tale fenomeno ha una distribuzione che, all'aumentare del numero dei fenomeni, converge alla normale.

N.B. È il teorema di limite centrale che svela il comportamento asintotico di una binomiale $Y \sim B(n,p)$ per p fissato ed n divergente. Infatti si dimostra che la somma di n v.a., di Bernoulli

$$\text{per} \quad i = 1, ..., n, \quad X_i \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

e indipendenti, è distribuito secondo una legge binomiale:

$$Y = X_1 + X_2 + ... + X_n \sim B(n,p).$$

Applicando il teorema di limite centrale, ne consegue il tipo di approssimazione di De Moivre e Laplace nel senso di convergenza delle funzioni distribuzione.
Pur essendo originato dal problema di approssimare la binomiale, l'utilità del teorema di limite centrale va molto oltre: ogni volta che si studia una somma di v.a., ma è sconosciuta la loro densità o ne è proibitivo il calcolo.

Es. Il tempo di sopravvivenza di una lampada è v.a. esponenziale di media $\mu = 10$ giorni. Appena si brucia, essa è sostituita. Trovare la probabilità che 40 lampade siano sufficienti per un anno.
Detta X_i la "durata della $i-$esima lampada", per $i = 1, ..., 40$ le X_i sono indipendenti ed esponenziali con parametro $\lambda = \frac{1}{\mu} = 1/10$. Sappiamo che $E(X_i) = \lambda^{-1} = 10$, $Var(X_i) = \lambda^{-2} = 100$. Allora la loro somma ha media $40 \cdot 10$ e varianza $40 \cdot 100$:

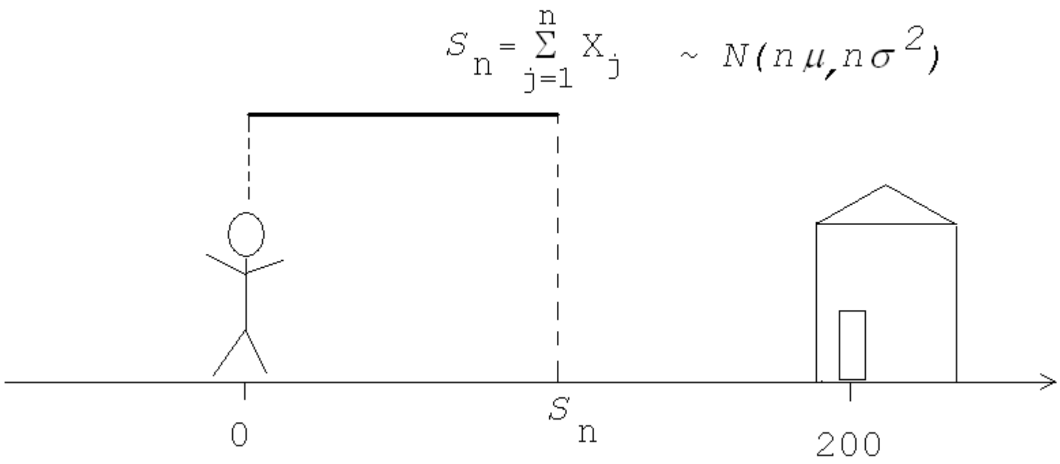
$$P(X_1 + X_2 + ... + X_{40} \geq 365) \simeq P[N(0,1) \geq \frac{365 - 400}{\sqrt{4000}}] = 71\%$$

Es. Supponiamo che un ubriaco disti duecento passi da casa, ma riesca solo a fare, ogni unità di tempo, un passo a destra con probabilità $1/2$ o un passo a sinistra con probabilità $1/2$. Quanti passi deve fare per avere probabilità 20% di arrivare a casa?

Suggerimento: all'istante 0 poniamo l'ubriaco nell'origine, e per ogni $j \in N$ sia $X_j =$ "spostamento al j-esimo istante". Dunque

$$X_j \sim \begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Quanto valgono μ e σ^2 di X_j? Come si comporta, per n grande, la posizione S_n dell'ubriaco



Es. Ogni giorno una ditta guadagna o perde un punto con probabilità 2/4 o 1/4 rispettivamente, mentre resta stazionaria con probabilità 1/4. Se questa tendenza perdura per 100 giorni, con quale probabilità avrà guadagnato 28 punti?

Suggerimento: all'istante 0 poniamo $S_\circ = 0$; per ogni $j \in N$ sia $X_j =$ "incremento (positivo o nullo o negativo) nel j-esimo giorno". Quanto valgono μ e σ^2 di X_j? Come si comporta, per n grande, il guadagno cumulativo S_n fino al giorno n? Quanto vale $P(S_{100} \geq 28)$?

LEGGE CONGIUNTA DI DUE V.A.

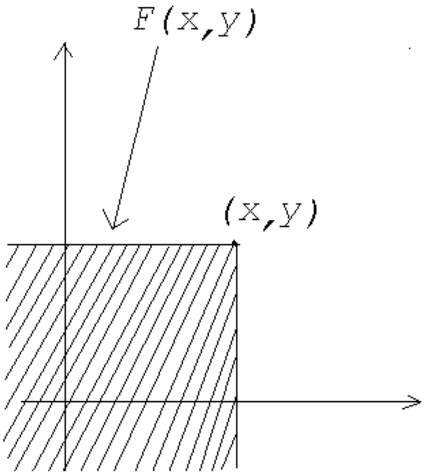
Nel considerare insieme due (o più) v.a., si sottintende sempre che siano definite sullo stesso spazio di probabilità Ω. Ci interessa fare affermazioni di probabilità contemporaneamente su X e su Y, per le quali non sarebbe sufficiente conoscere le due leggi singole di X e di Y. Ad esempio se in un territorio X ed Y sono misure dell'altezza e del peso delle persone, c'è da aspettarsi una certa dipendenza, ossia informazioni sulla concomitanza di variabilità che non sono contenute nelle due singole leggi di probabilità.

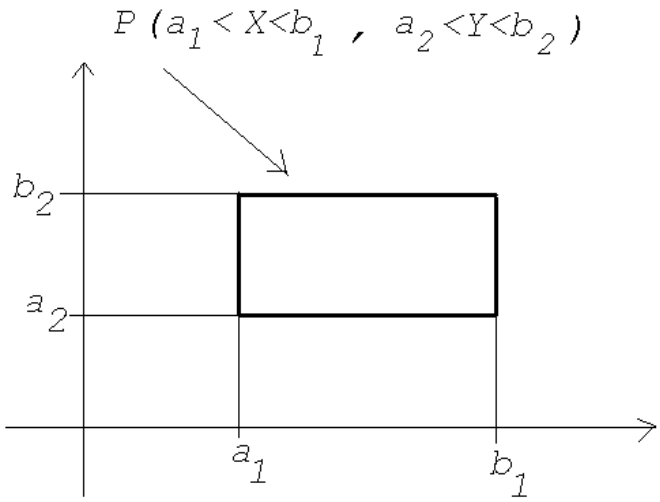
La legge di probabilità congiunta (o distribuzione bidimensionale) di due v.a. è definita su insiemi misurabili B contenuti in R^2:

$$P_{X,Y}(B) = P(\{\omega \in \Omega : \ (X(\omega), Y(\omega)) \in B\}) \equiv P_{X,Y}(B) = \ P[(X,Y) \in B\ ].$$

In particolare se l'insieme B è un prodotto cartesiano $B_1 \times B_2$, con $B_1, B_2 \subset R$, allora $P_{X,Y}(B) = P(X \in B_1, Y \in B_2)$. Si dice *funzione di ripartizone congiunta* $F_{X,Y}(\cdot,\cdot)$ delle v.a. X, Y la funzione di due variabili definita nel piano R^2:

$$F_{X,Y}(x,y) = \ P[X \leq x, Y \leq y] = P_{X,Y}[(-\infty, x) \times (-\infty, y)]$$





Essa contiene tutte le informazioni sulla legge congiunta $P_{X,Y}$, perché tutti gli insiemi misurabili di R^2 si ottengono con unioni numerabili e complementazioni a partire dai rettangoli. Il lettore può verificare la seguente formula mediante considerazioni geometriche in R^2:

$$P[a_1 < X \leq b_1,\ a_2 < Y \leq b_2] =$$

$$= F_{X,Y}(b_1, b_2) + F_{X,Y}(a_1, a_2) - F_{X,Y}(a_1, b_2) - F_{X,Y}(b_1, a_2).$$

Dalla funzione distribuzione congiunta $F_{X,Y}$ si possono ricavare le singole F_X, F_Y, dette *funzioni distribuzione marginali:*

$$F_X(x) = P(X \leq x, Y < +\infty) = \lim_{y \to \infty} F_{X,Y}(x,y)$$

$$F_Y(y) = P(X < +\infty, Y \leq y) = \lim_{x\to\infty} F_{X,Y}(x,y)$$

Ora specializziamo il discorso, fin qui valido in generale, alle v.a. discrete e poi a quelle assolutamente continue.

Definiamo *funzione massa di probabilità congiunta* di due v.a. discrete X, Y la funzione di due variabili

$$f_{X,Y}(x,y) = P(X = x, Y = y)$$

La somma rispetto a tutte le coppie (x, y) su cui tale funzione è positiva fa' 1. La funzione massa di probabilità marginale f_X sarà allora:

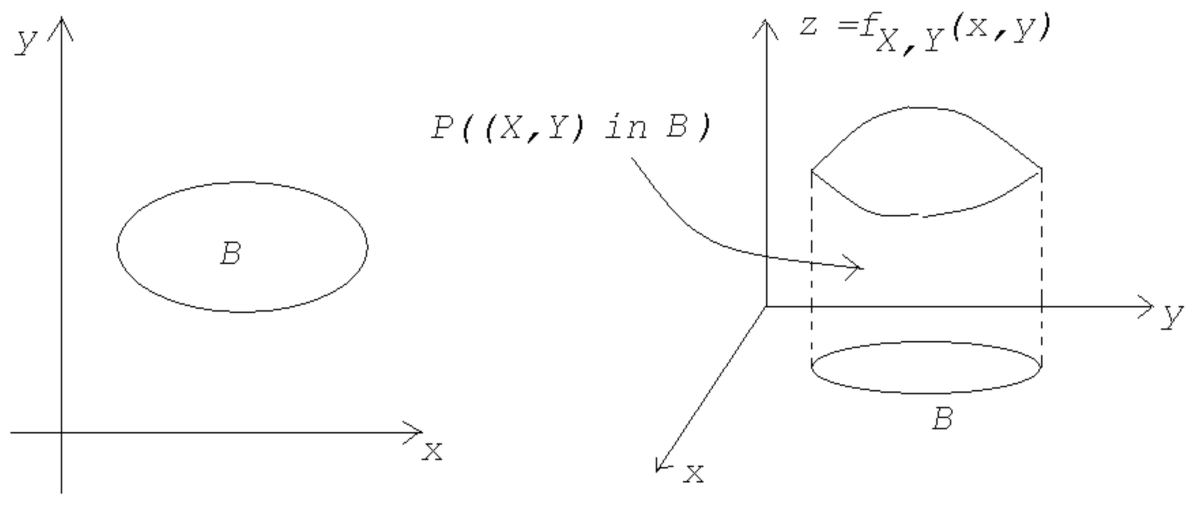
$$f_X(x) \;=\; \sum_{y:\ f_{X,Y}(x,y)>0} f_{X,Y}(x,y)$$

e analoga espressione avrà f_Y.

Def. *Una legge [o distribuzione] congiunta di due v.a. X, Y si dice* **assolutamente continua** *se esiste una funzione densità congiunta: ossia una funzione di due variabili $f_{X,Y}(x,y)$ tale che :*

$$P_{X,Y}(B) = P[(X,Y) \quad in \quad B] = \int_B f_{X,Y}(x,y)dxdy$$

per ogni insieme misurabile B di punti del piano.



Dalla densità si ottiene la funzione distribuzione congiunta:

$$F_{X,Y}(x_\circ, y_\circ) = \int_{-\infty}^{x_\circ} dx \int_{-\infty}^{y_\circ} dy f_{X,Y}(x,y)$$

e viceversa

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

in tutti i punti di continuità di f. La densità marginale e la distribuzione marginale, ad es. della v.a. X, si ottengono in modo naturale:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\ dy$$

$$F_X(x_\circ) = \int_{-\infty}^{x} dx \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\ dy$$

Inoltre serve sapere che due v.a. X, Y con densità e indipendenti producono un vettore aleatorio (X, Y) con densità data da:

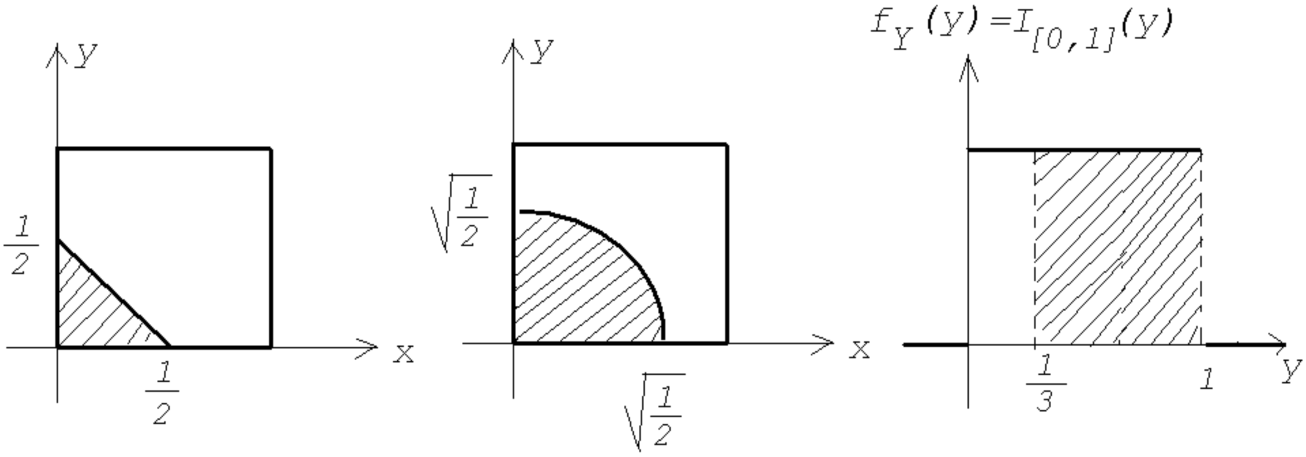
$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y), \quad \forall (x,y) \in R^2.$$

Esempi. Le due v.a. X, Y i.i.d. (=indipendenti identicamente distribuite) siano uniformi in $[0, 1]$. Trovare: *(a)* $P(X + Y < \frac{1}{2})$ *(b)* $P(X^2 + Y^2 < \frac{1}{2})$ *(c)* $P(\cos(\pi Y) < \frac{1}{2})$.
(a) Per indipendenza, la densità congiunta è il prodotto delle densit:

$$f_{X,Y}(x,y) = \begin{cases} 1 & (x,y) \in [0,1]^2 \subset R^2 \\ 0, & \text{altrove} \end{cases}$$

Dunque

$$P(Y < \frac{1}{2} - X) = \int \qquad dx\ dy = \int^{\frac{1}{2}} dx \int^{\frac{1}{2}-x} dy = \frac{1}{8} = 12.5\%$$



Si può anche tracciare, nel supporto $[0,1]^2$ della densità congiunta, il segmento congiungente $(0,\frac{1}{2})$ con $(\frac{1}{2},0)$ e vedere che l'area del triangolino sotteso è $\frac{1}{8}$.
(b) Traccio in $[0,1]^2$ l'arco di cerchio $x^2+y^2=(\frac{1}{\sqrt{2}})^2$, e noto che l'area sottesa in $[0,1]^2$ è $(\frac{1}{\sqrt{2}})^2\cdot\pi\cdot\frac{1}{4}=39.26\%$.
(c) Infine

$$P(\cos(\pi Y)<\frac{1}{2})=P(\pi Y>\frac{\pi}{3})=P(Y>\frac{1}{3})=\frac{2}{3}=66.66\%$$

Es. In due punti di un lago si misura l'intensità del suono causato da rumore di fondo generale (detto "rumore di ambiente"). Siano X,Y le due v.a. intensità del suono. Supponiamo che la loro legge congiunta abbia densit

$$f_{X,Y}(x,y)=\begin{cases} xy\exp[-\frac{1}{2}(x^2+y^2)], & x,y\geq 0\\ 0, \quad \text{altrove}\end{cases}$$

Trovare: *(a)* le densità marginali; *(b)*la distribuzione della intensità massima di rumore, $Z=\max(X,Y)$; (c) la distribuzione dell'intensità minima di rumore $U=\min(X,Y)$.

$$f_X(x)=\int_0^\infty xy\exp[-\frac{1}{2}(x^2+y^2)]dy=x\exp(-\frac{1}{2}x^2),$$
$$f_Y(y)=\int_0^\infty xy\exp[-\frac{1}{2}(x^2+y^2)]dx=y\exp(-\frac{1}{2}y^2).$$

Per un numero positivo z, la probabilità che Z sia minore o uguale a z è data da

$$\begin{aligned}P[\max(X,Y)\leq z]&=P[X\leq z,Y\leq z]\ =\\ &=\int_0^z dx\int_0^z dyxy\exp[-\frac{1}{2}(x^2+y^2)]=\\ &=\int_0^z xe^{-x^2/2}\ dx)^2=(1-e^{-z^2/2})^2\end{aligned}$$

Infine per la v.a. "minimo" conviene considerare la probabilità che essa sia $\geq u$, e poi trovarne il complemento ad 1 :

$$\begin{aligned}P(U\geq u)&=P(\min(X,Y)\geq u)=P(X\geq u,\ Y\geq u)=\\ &=\int_u^{+\infty}dx\int_u^{+\infty}dyxy\exp[-\frac{1}{2}(x^2+y^2)]=\\ &=(\int_u^{+\infty}dxxe^{-x^2/2})^2=e^{-u^2}\end{aligned}$$

Quindi funzione distribuzione e densità sono:

$$F_U(u)=(1-\exp(-u^2))\cdot I_{(0,\infty)}(u),\quad f_U(u)=2u\ e^{-u^2}\cdot I_{(0,\infty)}(u).$$

LEGGE DI PROBABILITA' DI UNA FUNZIONE DI UNA V.A.

Vogliamo ottenere la legge di probabilità di una v.a. $Y = g(X)$, conoscendo la legge della v.a. X e la trasformazione g. L'ipotesi migliore è che g sia un diffeomorfismo $R \to R$, cioè una funzione invertibile e differenziabile insieme con la sua inversa: questa ipotesi, come vedremo, garantisce l'esistenza di una densità f_Y di Y se esiste la densità f_X di X. Ma vorremmo trattare anche situazioni meno fortunate, in cui ci sia solo invertibilit'a a tratti e/o un tipo di regolarità meno forte [un tipico esempio può essere $g(x) = |x|$], ma allora bisognerà esaminare la situazione caso per caso.
Il metodo più comune è trovare prima la funzione distribuzione $F_Y(y)$, da cui si possono ricavare la funzione densità di probabilità [o la funzione di probabilità $f_Y(y)$] nei casi in cui tali funzioni esistono. Avremo:

$$F_Y(y) = P_X(\ g(x) \leq y\ ) \quad \text{se} \quad Y = g(X).$$

Es. Sia g una trasformazione affine: $g(x) = ax + b$, con $a > 0$, $b \in R$. La f.distribuzione di $Y = aX + b$ è:

$$F_{aX+b} = P(aX + b \leq y) = P(X \leq \frac{y-b}{a}) = F_X(\frac{y-b}{a}).$$

Se X è continua con densità $f_X = F'_X$, derivando troviamo

$$f_{aX+b}(y) = \frac{1}{a} f_X(\frac{y-b}{a}).$$

Se X è discreta con funzione massa f_X,

$$f_{aX+b}(y) = f_X(\frac{y-b}{a}).$$

Es. Sia $Y = X^2$. Poiché $g(x) = x^2 \geq 0$, $F_{X^2}(y) = 0$, $\forall y < 0$. Per $y \geq 0$,

$$F_{X^2}(y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) =$$
$$= F_X(\sqrt{y}) - F_X(-\sqrt{y}) + P(X = -\sqrt{y}) \qquad (*)$$

Se X è v.a. continua con funzione densità f_X allora l'ultimo addendo è zero ed $F_{X^2}(y)$ è un integrale: in tal caso, con la regola di derivazione composta, si ottiene la densità di Y:

$$f_{X^2}(y) = \begin{cases} 0, & y < 0 \\ [f_X(\sqrt{y}) + f_X(-\sqrt{y})]\frac{1}{2\sqrt{y}}, & y \geq 0 \end{cases}$$

Se X è discreta anche X^2 è discreta, poiché ciascuno dei due primi termini di $(*)$ è una somma di probabilità di valori puntuali. Perciò:

$$f_{X^2}(y) = \begin{cases} f_X(\sqrt{y}) + f_X(-\sqrt{y}), & y \geq 0 \\ 0, & y < 0 \end{cases}$$

Es. Sia
$$Y = A \sin X, \quad X \sim U[-\pi/2, \pi/2], \quad A > 0.$$
Si tratta di una sinusoide aleatoria, perché la fase X è v.a. uniforme nell'intervallo $[-\pi/2, \pi/2]$. Visto che il codominio di g è $[-A, A]$, interessa $F_Y(\cdot)$ per $|y| < A$:

$$F_Y(y) = \; P(A \sin X \leq y) = P(\sin X \leq \frac{y}{A})$$

$$= P(X \leq \; arcsin \frac{y}{A}) = F_X(arcsin \frac{y}{A})$$

D'altra parte la v.a. angolare X, essendo uniforme, ha funzioni densità e distribuzione:

$$\forall x \in [-\frac{\pi}{2}, \frac{\pi}{2}], \quad f_X(x) = \frac{1}{\pi}, \quad F_X(x) = \frac{1}{\pi}(x + \frac{\pi}{2}).$$

Perciò
$$F_Y(y) = \frac{1}{\pi}(arcsin \frac{y}{A} + \frac{\pi}{2}), \quad |y| \leq A$$

Di conseguenza, la densità di Y è:

$$f_Y(y) = \begin{cases} \frac{1}{\pi A}(1 - (\frac{y}{A})^2)^{-1/2}, & |y| < A, \\ 0, \quad \text{altrove} \end{cases}$$

[*Applicazione:* la gittata di un proiettile di velocità iniziale v lanciato con inclinazione α è $Y = (v^2/g) \sin(2\alpha)$. Se l'angolo di sparo α è aleatorio e in particolare uniforme in un intervallo, la gittata segue una legge del tipo di quella appena ricavata.]
Esercizio: se ad es. l'angolo α è uniformemente distribuito in $(\pi/6, \pi/4)$, come è distribuita la gittata Y in ($(v^2/g) \sin(\pi/3), (v^2/g) \sin(\pi/2)$)?

TEOREMA FONDAMENTALE PER IL CALCOLO DI DENSITA'
Rammentiamo che una v.a. è assolutamente continua se è assolutamente continua la sua funzione di ripartizione $F_X(\cdot)$; o, detto altrimenti, se X ammette una funzione densità di probabilità.
Enumeriamo alcune condizioni sulla trasformazione g sotto le quali $g(X)$ è v.a. assolutamente continua se X è assolutamente continua. Vogliamo inoltre la funzione densità di $Y = g(X)$ in termini della funzione densità di X e delle derivate di $g(\cdot)$.

[Prima di enunciare tale teorema, può essere utile al lettore richiamare la derivazione della funzione inversa: sotto le appropriate ipotesi, se $y = g(x)$ ha inversa $x = g^{-1}(y)$

$$\frac{dx}{dy} = \frac{d}{dy} g^{-1}(y) = \frac{1}{g'(x)}|_{x=g^{-1}(y)}) = \frac{1}{dy/dx}.$$

Ad esempio per $y = sinx$ $(-\pi/2 < x < \pi/2)$ si ha $x = arcsin \; y$ $\;(-1 < y < 1)$:

$$\frac{dx}{dy} = \frac{d}{dy} arcsin \; y = \frac{1}{\frac{d}{dx} sin \; x}|_{x=x(y)} =$$

$$= \frac{1}{cosx}|_{x=x(y)} \;=\; \frac{1}{\sqrt{1-sin^2x}}|_{x=x(y)} = \;\frac{1}{\sqrt{1-y^2}}, \quad -1 < y < 1\;].$$

È utile usare anche la funzione indicatrice di un insieme:

$$I_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A. \end{cases}$$

Teorema fondamentale per il calcolo di densità.
a) *Siano U, V due aperti di R e sia $g: U \to V$ un diffeomorfismo, cioè una funzione invertibile e differenziabile insieme alla sua inversa. Se X è una v.a. assolutamente continua, allora $Y = g(X)$ è una v.a. assolutamente continua con densità di probabilità*

$$f_Y(y) = \; f_X(x) \cdot |g'(x)|^{-1}|_{x=g^{-1}(y)} \cdot I_V(y)$$

$$\textit{o anche} \;\; f_Y(y) = f_X[g^{-1}(y)]|\frac{d}{dy}g^{-1}(y)| \cdot I_V(y).$$

b) *Sia $g: U \to V$ non necessariamente invertibile. Supponiamo che l'insieme dei punti singolari di g cioè*

$$\Sigma_g = \{x \in U: \; \textit{non esiste } g'(x) \;\; o \;\; g'(x) = 0\},$$

sia finito. Per ogni numero $y \in V$ chiamiamo $m(y)$ il numero dei punti non singolari che sono soluzioni dell'equazione $g(x) = y$:

$$m(y) \in N \;\; se \quad \exists x_1(y), ...x_m(y) \;\; :$$

$$per \; k = 1, ..., m(y), \quad g[x_k(y)] = y, \;\; \exists g'[x_k(y)] \neq 0$$

$$m(y) = 0 \;\; se \quad non \; esiste \; x \;\; : \quad g(x) = y \;\; e \;\; g'(x) \neq 0.$$

Ebbene, se X è una v.a. assolutamente continua allora anche $g(X)$ è v.a. assolutamente continua con funzione densità

$$f_Y(y) = \begin{cases} \sum_{k=1}^{m(y)} f_X[x_k(y)]|g'[x_k(y)]|^{-1}, & se \quad m(y) > 0 \\ 0, \quad se \quad m(y) = 0 \end{cases}$$

Dim. Per brevità ci limitiamo alla situazione (a), in cui g è diffeomorfismo, nel qual caso vale la formula del cambio di variabile di integrazione:

$$\int_U h(x)dx = \int_V h[g^{-1}(y)]|\frac{d}{dy}g^{-1}(y)|dy.$$

per ogni funzione integrabile nonnegativa h. Ora, una $f_Y(\cdot) \geq 0$ è la densità di Y sse

$$P(Y \in A) = \int_A f_Y(y)dy, \quad \forall A \text{ aperto } \quad \subset R.$$

Usando l'invertibilità di g e la formula del cambio di variabile di integrazione

$$P(Y \in A) = P[g(X) \in A] = P[X \in g^{-1}(A)] =$$

$$= \int_{g^{-1}(A)} f_X(x)\ dx = \int_A f_X[\ g^{-1}(y)] |\frac{d}{dy} g^{-1}(y)| dy$$

per ogni aperto $A \subset V$. L'uguaglianza si estende ad ogni aperto A di R se definiamo

$$f_Y(y) = \begin{cases} f_X[\ g^{-1}(y)] |\frac{d}{dy} g^{-1}(y)|, & y \in V \\ 0, & \text{altrove.} \end{cases}$$

[Ecco anche un argomento più intuitivo, che però esige di sapere la derivabilità delle funzioni distribuzione F. Per $y \in V$, se g è crescente

$$F_Y(y) = P[g(X) \leq y] = P[X \leq g^{-1}(y)] = F_X[g^{-1}(y)].$$

Se g è decrescente,

$$F_Y(y) = P[g(X) \leq y] = P[X \geq g^{-1}(y)] = 1 - F_X[g^{-1}(y)].$$

Derivando nel primo caso si ottiene:

$$F'_Y(y) = f_X[g^{-1}(y)] \frac{d}{dy} g^{-1}(y), \quad \text{se} \quad y \in V$$

$\frac{d}{dy} g^{-1} > 0$ perché g^{-1} è crescente come lo è g. Derivando nell'altro caso si ottiene

$$F'_Y(y) = f_X[g^{-1}(y)](-1) \frac{d}{dy} g^{-1}(y), \quad y \in V$$

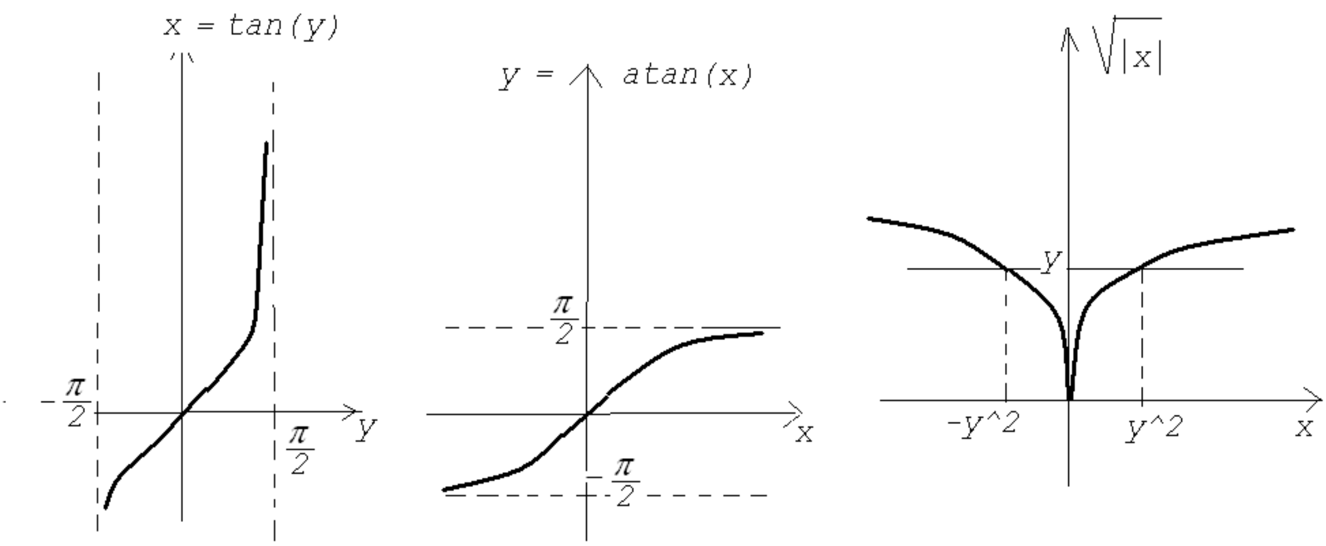
dove si ricompone il modulo di $(g^{-1})'$ perché g^{-1} è decrescente come lo è g.] $\square$

Es. Sapendo che X è v.a. assolutamente continua con densità $f_X(\cdot)$, vogliamo la densità $f_Y(\cdot)$ di $Y = arctan\ X$. Sappiamo che $y = arctan(x)$ è crescente in $(-\infty, +\infty)$ con inversa $x = tan(y)$ crescente in $(-\frac{\pi}{2}, \frac{\pi}{2})$. Per il teorema fondamentale:

$$f_Y(y) = \begin{cases} f_X(tan\ y) \cdot (1 + tan^2 y), & \text{se} \quad |y| < \frac{\pi}{2} \\ 0, & \text{altrove} \end{cases}.$$

Es. Sia $g(x) = \sqrt{|x|}$. Allora, $\forall y < 0$, $g(x) = y$ non ha alcuna soluzione, cioè $m(y) = 0$, cioè ci troviamo nel caso (b). Invece $\forall y > 0$, $g(x) = y$ ha due soluzioni $x_1(y) = y^2$, ed $x_2(y) = -y^2$. Inoltre $x_1'(y) = 2y$, $\quad x_2'(y) = -2y$, ma il modulo è $2y$. Per il teorema fondamentale, se X è assolutamente continua, $Y = \sqrt{|X|}$ ha densità:

$$(\ f_X(y^2) + f_X(-y^2) \ ) \cdot 2y \quad \text{per} \quad y > 0$$



Altro esercizio: mostrare che se X è v.a. assolutamente continua

$$f_{|X|}(y) = \begin{cases} f_X(y) + f_X(-y), & \text{per} \quad y > 0 \\ 0, \quad \text{per} \quad y < 0 \end{cases}$$

Problema. L'ampiezza $X(t)$, ad un tempo t, del segnale emesso da un certo generatore di segnali aleatori, è una v.a. distribuita normalmente con parametri $\mu = 0$, $\sigma^2 > 0$. L'onda $X(t)$ passa attraverso un circuito quadrante; l'uscita $Y(t)$, al tempo t, sia $Y(t) = X^2(t)$. Trovare la densità di $Y(t)$.
Risposta. Sia $g(x) = x^2$. Allora, $\forall y < 0$, $g(x) = y$ non ha alcuna soluzione, cioè $m(y) = 0$. Invece $\forall y > 0$, $g(x) = y$ ha due soluzioni $x_1(y) = \sqrt{y}$, ed $x_2(y) = -\sqrt{y}$. Inoltre $x_1'(y) = \frac{1}{2\sqrt{y}}$, $\quad x_2'(y) = -\frac{1}{2\sqrt{y}}$, e il modulo è in ogni caso $\frac{1}{2\sqrt{y}}$. Per il teorema fondamentale, se X è continua, $Y = X^2$ ha densità:

$$f_Y(y) = \begin{cases} (\ f_X(\sqrt{y}) + f_X(-\sqrt{y}) \ ) \frac{1}{2\sqrt{y}}, & \text{per} \quad y > 0 \\ 0, \quad \text{per} \quad y < 0 \end{cases}$$

Nel caso di $X \sim N(0, \sigma^2)$, essendo $f_X(x) = f_X(-x)$,

$$f_Y(y) = \begin{cases} 2 \cdot f_X(\sqrt{y}) \frac{1}{2\sqrt{y}}, & y > 0 \\ 0, \quad \text{altrove} \end{cases}$$

$$\text{cioè: } f_Y(y) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi y}} e^{-y/2\sigma^2}, & y > 0 \\ 0, \quad \text{altrove.} \end{cases}$$

Nel caso $\sigma = 1$, Y è la famosa v.a. $\chi^2(1)$, detta "chi quadrato con 1 grado di libertà".

Allarghiamo il panorama con funzioni g che non necessariamente mutano una v.a. assolutamente continua in una $g(X)$ continua. Tipicamente Y diventa *v.a. mista* (in parte discreta in parte assolutamente continua) nella seguente situazione:

Proposizione. *Sia X v.a. assolutamente continua. Sia I un intervallo nel quale la funzione $g(\cdot)$ assume un valore costante:*

$$\forall x \in I \subset R, \quad g(x) = \bar{y},$$

senza che esista un sovraintervallo proprio J in cui $g(x) = \bar{y}$. In questo caso, la v.a. Y assume il valore $\bar{y} = g(\bar{x})$ con probabilità

$$P(Y = \bar{y}) \;=\; P(X \in I).$$

e se tale probabilitá è non nulla, la v.a. Y è mista.
Dim. Basta un attimo di riflessione. $\square$

Es. *La parte positiva di una v.a.* Si definisce parte positiva (rispettivamente: parte negativa) di una variabile x reale:

$$x^+ = \begin{cases} x, & \text{se } \; x \geq 0 \\ 0, & \text{se } \; x < 0 \end{cases} \qquad x^- = \begin{cases} 0, & \text{se } \; x \geq 0 \\ -x & \text{se } \; x < 0 \end{cases}$$

Ad es. in tale notazione si ha $x = x^+ - x^-$, e $|x| = x^+ x^-$. Ora, data la v.a. X, la v.a. $Y = X^+$ è nota come *la parte positiva di X*. Cerchiamo per ispezione diretta la funzione distribuzione:

$$y < 0 \implies P(X^+ \leq y) = P(\emptyset) = 0$$

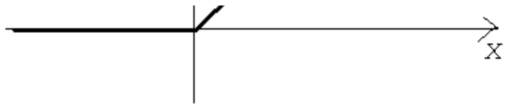
$$\{X^+ \leq 0\} = \{X \leq 0\} \implies P(X^+ \leq 0) = F_X(0)$$

$$y > 0 \implies P(X^+ \leq y) = P(X \leq y) = F_X(y)$$

Perciò

$$F_{X^+}(y) = \begin{cases} 0, \text{ se } \; y < 0 \\ F_X(0), \quad \text{se } \; y = 0 \\ F_X(y), \quad \text{se } \; y > 0 \end{cases}$$

Quindi, ad es., la parte positiva di una v.a. $N(0,1)$ non è né continua né discreta, ma ha una funzione distribuzione di tipo misto: F_Y ha una discontinuità di prima specie in 0, e tale "salto" è la massa di probabilità in 0.



Ad es. vogliamo la funzione distribuzione F_{X^+} se $X \sim U[-1,1]$?

Risposta. Basta partire da

$$F_X(x) = \begin{cases} 0, \ x < -1 \\ \frac{1}{2}(x+1), \quad -1 \leq x \leq 1 \\ 1, \ x > 1 \end{cases}$$

La funzione distribuzione di X^+ concentra in 0 la massa che per X è distribuita in $[-1,0]$:

$$F_{X^+}(x) = \begin{cases} 0, \quad x < 0 \\ \frac{1}{2}, \quad x = 0 \\ \frac{1}{2}(x+1) \quad 0 < x \leq 1 \end{cases}$$

mentre vale 1, $\forall x \geq 1$.

Es. In un istante t l'ampiezza X di un segnale emesso da un generatore di segnali aleatori è una v.a. normale $N(0.3,\ 0.25)$. L'onda X passa attraverso un clipper dando in uscita $Y = g[X]$, dove $g(x) = 1$ o 0 a seconda che $x > 0$ o $x < 0$. Trovare la funzione massa di probabilità di Y per tale istante.
Risposta: si capisce che Y è discreta con valori possibili dati da 1 e da 0.

$$P(Y=0) = P(X<0) = P(N(0,1) < \frac{0-0.3}{0.5}) =$$

$$= 1 - P(N(0,1) \leq 0.6) = 1 - 0.7257 = 27.43\%$$

$$P(Y=1) = P(X>0) = 72.57\%$$

LEGGE DI PROBABILITÀ DI UNA FUNZIONE DI PIU' V.A.

Studiamo la legge di una v.a. $Y = g(X_1, X_2, ..., X)n)$ che nasce come funzione di n v.a. distribuite congiuntamente $X_1, X_2, ..., X_n$.

Supponiamo che sia data la legge di probabilità congiunta $P_{X_1,...,X_n}$ e poniamo $Y = g(X_1,...,X_n)$. Il punto di partenza dei nostri argomenti è la seguente espressione della funzione distribuzione di Y:

$$F_Y(y) \equiv P(Y \leq y) = P_{X_1,...,X_n}[\{(x_1,...,x_n) \in R^n: \ g(x_1,...,x_n) \leq y]$$

Infatti: l'evento $\{Y \leq y\}$ è logicamente l'evento $\{g(X_1,...,X_n) \leq y\}$, e ciò significa: i valori osservati delle variabili $X_1,...,X_n$ sono nella regione di $n-$uple tali che $g(x_1,...,x_n) \leq y$. Ci interessa particolarmente il caso in cui le v.a. siano congiuntamente continue, con densità congiunta assegnata; allora

$$F_Y(y) = \int \dots \int_{g(x_1,\dots,x_n) \square y} f_{X_1,\dots,X_n}(x_1, \dots, x_n) dx_1 \dots dx_n$$

Legge di probabilità della somma. *La funzione densità della somma $Z = X + Y$ di due v.a. continue è*

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)\ dx = \int_{-\infty}^{\infty} f_{X,Y}(z - y, y)\ dy$$

Dim.

$$F_Z(z) = P(X + Y \leq z) = \int_{\{(x,y):x+y \square z\}} f_{X,Y}(x, y) dx dy$$

$$= \int_{-\infty}^{\infty} dx \int_{y=-\infty}^{y=z-x} dy f_{X,Y}(x, y) =$$

Derivando rispetto a z l'ultima uguaglianza otteniamo

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \int_{-\infty}^{+\infty} f_{X,Y}(x, z - x)\ dx.$$

Per simmetria vale anche l'altra espressione. $\square$

OSS. Se le v.a. X, Y sono indipendenti, la densità congiunta è il prodotto delle densità marginali:

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\ dx =$$

$$= \int_{-\infty}^{\infty} f_X(z - y) f_Y(y)\ dy =$$

L'operazione tra le funzioni densità che in questa formula compare è usata in varie branche della matematica: la *convoluzione* fra due funzioni f_1, f_2 è definita:

$$f_3(y) = \int_{-\infty}^{\infty} f_1(x) f_2(y - x) dx$$

e si indica $f_3 = f_1 * f_2$. Possiamo concludere: *la funzione densità della somma $X + Y$ di due v.a. indipendenti e continue è la convoluzione delle funzioni densità f_X ed f_Y delle due v.a.*

Analogo argomento porta a scrivere la funzione massa di probabilità di una *somma di v.a. discrete:*

$$f_{X+Y}(z) = P(X + Y = z) = \sum_{x:P[(X,Y)=(x,z-x)]>0} f_{X,Y}(x, z - x)$$

Si tratta di una somma dove varia un solo parametro, ad es. $x \in R$, e inoltre sappiamo che solo su un insieme discreto di x si ha $f_{X,Y}(x, z-x) > 0$. Conclusione:

$$f_{X+Y}(z) = \sum_{x \in R} f_{X,Y}(x, z-x) = \sum_{y \in R} f_{X,Y}(z-y, y)$$

Con argomenti analoghi si prova:

Teorema. *Se X, Y sono v.a. congiuntamente continue, valgono le seguenti formule per la loro differenza, prodotto e quoziente:*

$$f_{X-Y}(z) = \int_{-\infty}^{\infty} f_{X,Y}(z+y, y)\ dy = \int_{-\infty}^{\infty} f_{X,Y}(x, x-z)\ dx$$

$$f_{X \cdot Y}(z) = \int_{-\infty}^{\infty} dy \frac{1}{|y|} f_{X,Y}(\frac{z}{y}, y) = \int_{-\infty}^{\infty} dx \frac{1}{|x|} f_{X,Y}(x, \frac{z}{x})$$

$$f_{X/Y}(z) = \int_{-\infty}^{\infty} dy |y| f_{X,Y}(zy, y).$$

Dim. Per semplicità $f_{X,Y} = f$. Scrivendo F e derivando si ottiene:

$$F_{X-Y}(z) = \int_{-\infty}^{+\infty} dy \int_{x=-\infty}^{x=y+z} dx f_{X,Y}(x, y)$$

$$\Longrightarrow \qquad f_{X-Y}(z) = \frac{dF}{dz} = \int_{-\infty}^{+\infty} dy f_{X,Y}(y+z, y).$$

$$F_{X \cdot Y}(z) = \int_{\{y>0,\ z/y \geq x\}} f(x,y) dx dy \ + \int_{\{y<0,\ x \geq z/y\}} f(x,y) dx dy =$$

$$= \int_{0}^{+\infty} dy \int_{-\infty}^{z/y} dx\ f(x,y) \ + \int_{-\infty}^{0} dy \int_{z/y}^{+\infty} dx\ f(x,y).$$

Dunque $f_{X \cdot Y}(z)$ non è altro che

$$\int_{0}^{+\infty} dy \frac{1}{y} f(\frac{z}{y}, y) + \int_{-\infty}^{0} dy \frac{(-1)}{y} f(\frac{z}{y}, y)$$

il che ricompone il fattore $\frac{1}{|y|}$ come enunciato.

$$F_{X/Y}(z) = \int_{\{yz \geq x, y>0\} \cup \{x \geq yz, y<0\}} f(x,y) dx dy =$$

$$= \int_{0}^{+\infty} dy \int_{-\infty}^{yz} dx\ f(x,y) \ + \int_{-\infty}^{0} dy \int_{yz}^{+\infty} dx f(x,y) =$$

$$\Longrightarrow \quad f_{X/Y}(z) = \int_0^{+\infty} (dy) y f(yz, y) \; + \; \int_{-\infty}^{0} (dy)(-y) f(yz, y) =$$

$$\int_{-\infty}^{+\infty} (dy)|y| f(yz, y).$$

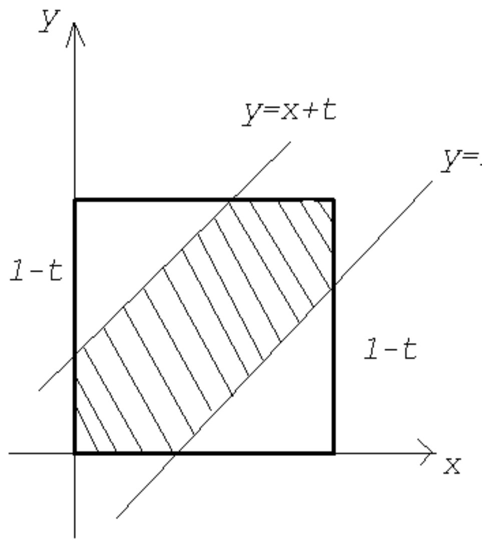
$\square$

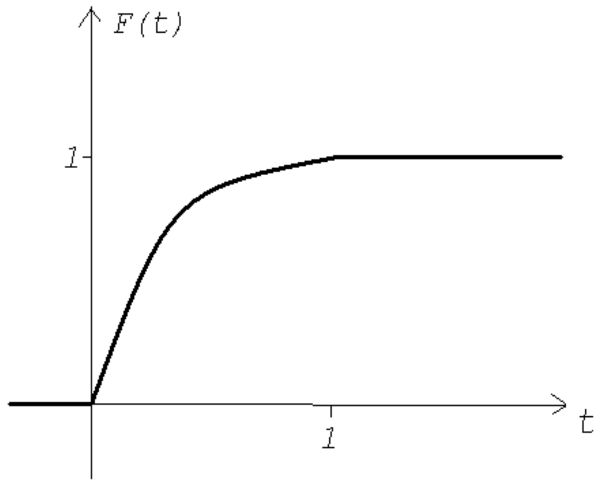
OSS. I teoremi precedenti dicono in tutta generalità la densità di somma, differenza, quoziente... di due v.a. assolutamente continue. In pratica conviene spesso cercare con metodi diretti, adattati a ciascun caso in esame, prima la funzione distribuzione, poi la densità di $X \pm Y$, o di Y/X, ecc.

Es. 1. Se X $N(3;25)$ ed Y $N(-1;4)$ e sono indipendenti, trovare la densità di $4X - 5Y$. Risposta: sappiamo che ogni combinazione lineare $c_1X + c_2Y$ di v.a. normali e indipendenti è normale con media $c_1\mu_1 + c_2\mu_2$ e varianza $c_1^2\sigma_1^2 + c_2^2\sigma_2^2$. Dunque $4X - 5Y$ è normale con media e varianza:

$$\mu = 4 \cdot 3 - 5 \cdot (-1) = 17; \qquad \sigma^2 = 16 \cdot 25 + 5 \cdot 4 = 420$$

di cui è ben nota la densità.





Es. 2. Siano X, Y uniformi in $[0,1]$ e indipendenti; trovare distribuzione e densità di $T = |X - Y|$.

$$F_{|X-Y|}(t) = \int_{\{|y-x|<t\} \cap [0,1]^2} dx\ dy.$$

Il dominio è quindi: $-t \le y - x \le t$, cioè $x - t \le y \le x + t$ per ogni $0 < t < 1$:

$$F_T(t) = \begin{cases} 0 \; t < 0 \\ 1 - (1-t)^2, \; 0 < t < 1 \\ 1, \; t \ge 1 \end{cases}$$

per cui $f_T(t) = +2(1-t) \cdot I_{[0,1]}(t)$.
Applicazione: *il problema dell'incontro.* Con quale probabilità si incontreranno due amici, se concordano di arrivare in piazza in due momenti casuali e indipendenti fra le 5 e le 6, ognuno aspettando l'altro fino a dieci minuti?
Risposta: prendiamo l'ora per unità di misura, riportiamo all'ora "zero" i due istanti indipendenti X e Y di arrivo $X, Y \sim U[0,1]$; cerchiamo la probabilità che i due istanti differiscano in valore assoluto al più per 1/6 di ora:

$$P(|X - Y| \le 1/6) = 1 - (5/6)^2 = 1 - (25/36) = 11/36.$$

Es. Se X_1, X_2 sono v.a. continue con nota densità congiunta, vogliamo la densità della distanza euclidea dall'origine $Y = \sqrt{X_1^2 + X_2^2}$. Per $y < 0$, $F_Y(y) = 0$; per $y \ge 0$ troviamo:

$$F_Y(y) = \int_{\{(x_1,x_2): y^2 \ge x_1^2 + x_2^2\}} f_{X_1,X_2}(x_1, x_2) dx_1 dx_2$$

La simmetria del dominio di integrazione ci suggerisce di esprimere l'integrale doppio in coordinate polari: $x_1 = r\cos\theta$, $x_2 = r\sin\theta$ ed elemento di superfiicie $r dr d\theta$:

$$F_Y(y) = \int_0^{2\pi} d\theta \int_0^y r \; dr \; f_{X_1,X_2}(r\cos\theta, r\sin\theta)$$

La densità di Y si ottiene derivando rispetto ad y; poiché y ivi compare come estremo dell'integrazione rispetto ad r, la derivata di $F_Y(y)$ è l'integranda calcolata in $r = y$:

$$f_{\sqrt{X_1^2+X_2^2}}(y) = y \int_0^{2\pi} d\theta f_{X_1,X_2}(y\cos\theta, y\sin\theta) \cdot I_{(0,+\infty)}(y) \qquad (*)$$

Più in concreto: nel caso che X_1, X_2 siano indipendenti e normali, con $X_1, X_2 \sim N(0, \sigma^2)$, che v.a. è $Y = \sqrt{X_1^2 + X_2^2}$?
Per indipendenza la densità congiunta è il prodotto delle densità:

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{2\pi\sigma^2} \exp[-(x_1^2 + x_2^2)/2\sigma^2]$$

che si può mettere in coordinate polari ($y = \sqrt{x_1^2 + x_2^2}$):

$$f_{X_1, X_2}(y \ \cos\theta, y \ \sin\theta) = \frac{1}{2\pi\sigma^2} e^{-y^2/2\sigma^2}.$$

Si noti che è costante rispetto a θ; dunque l'integrazione fa' la costante 2π; dunque

$$f_{\sqrt{X_1^2+X_2^2}}(y) = \frac{y}{\sigma^2} e^{-y^2/2\sigma^2} \cdot I_{(0,\infty)}(y) \qquad (**)$$

Questa v.a. Y è la ben nota v.a. di Rayleigh con parametro σ^2 . $\square$

CONDIZIONAMENTI, LEGGI CONDIZIONALI

Consideriamo due v.a. X, Y con una nota legge di probabilità congiunta, o misura o distribuzione:

$$A, B \text{ Boreliani } \subset R, \; P_{X,Y}(A \times B) = P(X \in A, Y \in B).$$

Ora, il conoscere che la v.a. X ha assunto un particolare valore x induce *un condizionamento* della v.a. Y; cioè può apportare una modifica della legge di probabilità marginale di Y.
L'idea - *informalmente* - può essere:

$$F_Y(y \,| x \le X \le x+h) = \frac{P(x \le X \le x+h, Y \le y)}{P(x \le X \le x+h)} \cdot \frac{1/h}{1/h} =$$

$$= \frac{\frac{1}{h} \int_{-\infty}^{y} d\eta \int_{x}^{x+h} f_{X,Y}(\xi, \eta) d\xi}{\frac{1}{h} \int_{x}^{x+h} f_X(\xi) d\xi} \quad \rightarrow$$

$$(\text{per } \; h \to 0), \quad \rightarrow \quad \frac{\int_{-\infty}^{y} f_{X,Y}(x, \eta) d\eta}{f_X(x)}$$

supponendo di poter applicare il teorema della media integrale. Ciò motiva la seguente definizione di distribuzione condizionata.

Def. La *funzione distribuzione condizionata* della v.a. Y dato $\{X = x\}$ è per definizione la funzione distribuzione dipendente dal parametro x:

$$F_{Y|X}(y \,|x) \; := \; \frac{\int_{-\infty}^{y} f_{X,Y}(x, \eta) \; d\eta}{f_X(x)}$$

Essa è definita per ogni x tale che $f_X(x) > 0$.
Per le x tali che $f_X(x) > 0$, la *funzione densità di probabilità condizionata* $f_{Y|X}(y|x)$ della v.a. Y dato $\{X = x\}$ è la densità dipendente dal parametro x

$$f_{Y|X}(y|x) \; := \frac{f_{X,Y}(x, y)}{f_X(x)}.$$

Se la funzione distribuzione condizionata è derivabile,

$$f_{Y|X}(y|x) \; = \frac{d \; F_{Y|X}(y|x)}{dy}.$$

OSS. Dunque la densità congiunta è uguale alla densità condizionata di Y dato $X = x$ moltiplicata per la densità marginale di X: :

$$f_{X,Y}(x, y) = \; f_{Y|X}(y|x) \cdot f_X(x).$$

In particolare, se la densità di probabilità *marginale* $f_Y(y)$ della v.a. Y coincide con la densità di probabilità *condizionata* $f_{Y|X}(y|x)$, allora le due v.a. sono *indipendenti; infatti la densità congiunta risulta essere il prodotto delle due marginali:*

$$f_Y(y) = f_{Y|X}(y|x) \quad \Longrightarrow \quad f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y).$$

Inoltre sussiste una specie di *teorema di Bayes per le densità condizionali.* Infatti essendo

$$f_{X,Y}(x,y) = \ f_{X|Y}(x|y) f_Y(y) \ = f_{Y|X}(y|x) f_X(x)$$

si ottiene

$$f_{X|Y}(x|y) \ = \ \frac{f_{Y|X} \cdot f_X(x)}{f_Y(y)} \ = \ \frac{f_{Y|X} \cdot f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x) dx}.$$

Es. Il tempo di decadimento X di una particella in una camera a nebbia sia una v.a. esponenziale con parametro y. Tuttavia il valore y non è uguale per tutte le particelle bensi' è distribuito come la v.a. Y. Assumendo che Y sia uniforme in $[0,1]$, si richiede: 1) la densità marginale di X; 2) la densità condizionata di Y dato $X = x$; 3) la migliore stima di Y dato $X = 1$.

1) L'ipotesi appena descritta sul tempo X non è altro che un'ipotesi sulla legge condizionata di X dato Y:

$$f_{X|Y}(x|y) = y \ e^{-xy} I_{R^+}(x) I_{R^+}(x) I_{(0,1)}(y).$$

Cerchiamo la densità marginale del tempo X, impiegato da una particella a caso per decadere: per $x > 0$

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y) dy =$$

$$= \int_{-\infty}^{+\infty} f_{X|Y}(x,y) \cdot f_Y(y) \ dy = \int_0^1 y \ e^{-xy} dy =$$

$$= [y \frac{e^{-xy}}{-x}]_0^1 - \int_0^1 (\frac{e^{-xy}}{-x}) dy = -\frac{e^{-x}}{x} - \frac{1}{x^2}[e^{-xy}]_0^1 =$$

$$= -\frac{e^{-x}}{x} + \frac{(1 - e^{-x})}{x^2}.$$

2) Per avere la densità condizionata di Y dato $X = x$ uso la formula di Bayes sulle densità:

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y) \cdot f_Y(y)}{f_X(x)} =$$

$$= \frac{y e^{-xy} \cdot I_{[0,1]}(y)}{-x^{-1} e^{-x} + x^{-2}(1 - e^{-x})}.$$

3) La stima richiesta di Y dato $X = x$ è il valor medio di Y condizionato ad $X = x$:

$$\forall x > 0, \quad E[y|X = x] = \int_{-\infty}^{+\infty} y \cdot f_{Y|X}(y|x) dy = \ \frac{1}{f_X(x)} \int_0^1 y^2 e^{-xy} dy.$$

Dunque

$$E[y|X=1] = \frac{1}{f_X(1)} \int_0^1 y^2 e^{-y} dy =$$

$$= \frac{1}{f_X(1)} \{-e^{-1} + \int_0^1 2ye^{-y} dy\} =$$

$$= \frac{1}{f_X(1)} [\ -e^{-1} - 2e^{-1} + 2 \int_0^1 e^{-y} dy\ ] =$$

$$= \frac{1}{f_X(1)} [\ -3e^{-1} + 2[\frac{e^{-y}}{-1}]_0^1\ ]\ =$$

$$= \frac{2 - 5 \cdot e^{-1}}{-e^{-1} + (1 - e^{-1})} =\simeq 0.16/(0.264) \simeq 0.60$$

Es. Ogni anno un tipo di macchina deve essere sottoposto ad alcuni arresti per manutenzione. Questo numero di arresti X è v.a. di Poisson con parametro y. Ma anche y è aleatorio (ad es. può dipendere dalla macchina) e assumiamo che esso segua una legge

$$f_Y(y) = \ ye^{-y} \cdot I_{(0,\infty)}(y).$$

(a) Qual è la probabilità che una singola macchina sia sottoposta a k arresti in un anno?
(b) Sapendo che una macchina ha subito $k = 5$ arresti l'anno scorso, qual è la migliore stima della sua "propensione alla difettosità" Y?

Risposta. Pur non dicendolo esplicitamente, il problema fornisce la funzione di probabilitá condizionata di X dato $Y = y$:

$$P(X = k|Y = y) \equiv \ f_{X|Y}(k|y) = \ \frac{y^k}{k!} e^{-y}.$$

La densità congiunta è il prodotto fra questa e la densità della v.a. condizionante:

$$f_{X,Y}(k,y) \ = \ f_{X|Y}(k|y) \cdot f_Y(y) = \frac{y^k}{k!} e^{-y} \cdot ye^{-y} \cdot I_(0, +\infty)(y)$$

per cui

$$P(X = k) = \int_0^\infty f_{X,Y}(k,y)\ dy = \ \int_0^\infty \frac{1}{k!} y^{k+1} e^{-2y} =$$

$$\text{(posto } \tau = 2y) \quad = \int_0^\infty (\frac{1}{2})^{k+2}) \tau^{k+1} k! e^{-\tau} d\tau = \frac{k+1}{2^{k+2}}$$

(b) Si tratta di calcolare la media condizionata di Y dato $X = 5$, quindi occorre la densità condizionata:

$$f_{Y|X}(y|k) = \frac{f_{X,Y}(k,y)}{f_X(k)} \ =$$

$$= [\ \frac{1}{k!} y^{k+1} e^{-2y}\ ] \cdot \frac{2^{k+2}}{k+1} = \frac{(2y)^{k+1} \cdot 2}{(k+1)!} e^{-2y}$$

Allora

$$E(Y|X=k) = \int_0^\infty y f_{Y|X}(y|k)\ dy\ =$$

$$= \int_0^\infty \frac{2^{k+2}}{(k+1)!} y^{k+2} e^{-2y} dy = \frac{2^{k+2}}{(k+1)!} \int_0^\infty e^{-\tau} (\frac{\tau}{2})^{k+2} \frac{1}{2} d\tau =$$

$$= \frac{(k+2)!}{2 \cdot (k+1)!} = \frac{k+2}{2}, \quad \text{cioé } E(Y|X=5) = \frac{7}{2} = 3.5$$

Essa é la migliore stima di Y in base al dato sperimentale $k = 5$. $\square$

È possibile anche calcolare la probabilità di eventi del tipo $g(X,Y) \leq z$, sotto la condizione, ad es., che $X = x$. *Se X, Y sono indipendenti* allora la condizione $X = x$ può essere usata mediante diretta sostituzione, senza problemi:

$$P[\ g(X,Y) \leq z\ |X = x] =\ P[\ g(x,Y) \leq z\ ].$$

Ecco un esempio, il problema dell'incontro condizionato:

Es. Due giovani decidono di incontrarsi tra le 17 e le 18 con l'accordo che nessuno deve aspettare l'altro per più di 10 minuti. Supponiamo che gli orari X ed Y in cui arrivano siano indipendenti e casuali, variabili fra le 17 e le 18. Trovare la probabilità condizionata che i due giovani si incontrino, dato che lei arriva all'ora x: (a) alle 17:05; (b) alle 17:30; (c) alle ore 17:50.
Risoluzione. Se lei arriva a un orario x e lui a un orario Y, si incontrano solo se $|Y - x| \leq 10$, cioè se $-10 + x \leq Y \leq x + 10$. Indichiamo con A l'evento che i due si incontrino e usiamo come unità di misura i minuti, chiamando minuto *zero* le 17:00.

Allora per $x = 5$ abbiamo

$$P(A|X=5) = P(-10 \leq Y - X \leq 10|X=5) =$$

$$= P(-10+5 \leq Y \leq 5+10) = F_Y(15) - F_Y(-5) = F_Y(15) = 15/60 = 1/4$$

essendo Y uniforme in $[0, 60]$. Similmente per $x = 30$

$$P(A|X=30) = P(-10 \leq Y - X \leq 10|X=30) =$$

$$= P(-10+30 \leq Y \leq 10+30) = F_Y(40) - F_Y(20) = (40-20)/60 = 1/3.$$

e per $x = 50$

$$P(A|X=50) = P(-10 \leq Y - X \leq 10|X=50) = F_Y(60) - F_Y(40) = 1/3.$$

Ricordiamo che, senza condizionamento, la probabilità di incontro era un po' minore di $1/3$: $P(|Y-X| \leq 10) = 11/36$. □

Infine menzioniamo anche un condizionamento di v.a. concettualmente più semplice: cerchiamo la legge condizionata di una v.a X dato un evento B. In altre parole, data la funzione distribuzione di X, $F_X(x) = P(X \leq x)$, cerchiamo la funzione distribuzione di X condizionata al verificarsi di B avente probabilità positiva. A tal fine basta la definizione di probabilità condizionata come rapporto fra la probabilità dell'intersezione e la probabilità condizionante:

$$F_{X|B}(x|B) = \frac{P[(X \leq x) \ \cap \ B]}{P(B)}.$$

Nel modo usuale definiremo allora *funzione densità di probabilità condizionata* $f_{X|B}(x|B)$ come quella funzione, se esiste, da cui si ottiene, per integrazione,la F. Nel caso che $F_{X|B}(x|B)$ sia derivabile, la densità condizionata si ottiene per derivazione:

$$f_{X|B}(x|B) = \frac{dF_{X|B}(x|B)}{dx}.$$

Tali due funzioni godono appieno le proprietà di ogni funzione distribuzione e di ogni densità di probabilità.

Es. La funzione distribuzione di X ="numero uscente nel lancio di un dado equo" ha sei discontinuità in $j = 1, 2, ..., 6$. L'evento condizionante B ="numero pari" modifica la funzione distribuzione; infatti gli unici valori possibili sono $2, 4, 6$ con rispettive probabilità uguali a $1/3$:

$$F_{X|B}(x|B) = \begin{cases} 0, & x < 2 \\ \frac{1}{3}, & 2 \leq x < 4 \\ \frac{2}{3}, & 4 \leq x < 6 \\ 1, & x \geq 6. \end{cases}$$

Es. *Tempo di guasto dopo rodaggio.* Collaudiamo lampadine lasciandole accese fino al guasto (bruciatura del filamento). Un modello per il *tempo di guasto* T di una lampadina è la v.a. esponenziale, avente densità non nulla per $t > 0$:

$$\forall t > 0, \quad f_T(t) = \lambda e^{-\lambda t}; \quad \frac{1}{\lambda} = \text{tempo medio di guasto.}$$

Modifichiamo ora il modo di collaudare: attendiamo un tempo fisso $t_\circ$ fino all'accensione, scartiamo le lampadine che all'istante $t_\circ$ sono già guaste, e collaudiamo solo quelle ancora funzionanti. L'evento condizionante è insomma $B = \{T \geq t_\circ\}$: cerchiamo la distribuzione condizionata $F_{T|B}(t|B)$.

$$P(B) = P(T \geq t_\circ) = 1 - P(T \leq t_\circ) = 1 - F_T(t_\circ).$$

La probabilità dell'intersezione è

$$P[(T \leq t) \cap (T \geq t_\circ)] = \begin{cases} F_T(t) - F_T(t_\circ), & t \geq t_\circ \\ 0 & t < t_\circ. \end{cases}$$

Infatti è vuota l'intersezione fra i due eventi proprio quando $t < t_\circ$. Otteniamo allora:

$$F_{T|B}(t|B) = \frac{P[(T \leq t) \cap (T \geq t_\circ)]}{P(T \geq t_\circ)} = \begin{cases} \frac{F_T(t) - F_T(t_\circ)}{1 - F_T(t_\circ)}, & t \geq t_\circ \\ 0 & t < t_\circ. \end{cases}$$

Quindi la densità della v.a. "tempo di guasto dopo rodaggio" è ottenuta derivando:

$$f_{T|B}(t|B) = \frac{d\ F_{T|B}(t|B)}{dt} = \begin{cases} \frac{f_T(t)}{1 - F_T(t_\circ)}, & t \geq t_\circ \\ 0, & t < t_\circ. \end{cases}$$

Allora, in che modo il rodaggio modifica la densità esponenziale $f_T(t)$ originaria? Primo: la densità del tempo di guasto dopo rodaggio è ovviamente nulla per $t < t_\circ$: il rodaggio è stato superato con certezza, la lampadina si guasta solo dopo $t_\circ$. Inoltre per $t > t_\circ$ la densità modificata ha l'espressione di $f_T(t)$, ma divisa per $1 - F_T(t_\circ)$, al fine di garantire probabilità totale uguale ad 1.

COVARIANZA

Per una coppia di v.a. (X_1, X_2), ecco il parametro più utile a dare indicazioni sul comportamento congiunto di X_1 e di X_2 :

Def. La *covarianza* di una legge congiunta per le v.a. X_1, X_2 con medie μ_1 e μ_2 è per definizione:

$$Cov(X_1, X_2) \ := \ E[(X_1 - \mu_1)(X_2 - \mu_2)]$$

Dunque se la legge congiunta è discreta

$$Cov(X_1, X_2) = \sum_k (x_{1,k} - \mu_1)(x_{2,k} - \mu_2) f_{X_1,X_2}(x_{1,k}, x_{2,k})$$

Se la legge congiunta è continua

$$Cov(X_1, X_2) = \int_{R^2} (x_1 - \mu_1)(x_2 - \mu_2) f_{X_1,X_2}(x_1, x_2) dx_1 dx_2.$$

Prop. *La covarianza di due v.a. è uguale alla differenza "media del prodotto meno prodotto delle medie":*

$$Cov(X_1, X_2) = \ E(X_1 \cdot X_2) \ - \mu_1 \mu_2.$$

Inoltre, mentre $E(X_1 + X_2) = E(X_1) + E(X_2)$*, la varianza di una somma soddisfa:*

$$Var(X_1 + X_2) = Var(X_1) + 2 \cdot Cov(X_1, X_2) + Var(X_2).$$

In particolare, quando X_1, X_2 sono indipendenti

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2) \quad Var(X_1 - X_2) = Var(X_1) + Var(X_2).$$

Dim. Siano, per fissare le idee, X_1, X_2 v.a. discrete che assumono i valori $\{x_{1,i}\}$ ed $\{x_{2,j}\}$, rispettivamente, in modo che $P(X_1 = x_{1,i}, X_2 = x_{2,j}) = f_{X_1,X_2}(x_{1,i}, x_{2,i})$. Poiché le due funzioni di probabilità marginali sono queste somme:

$$\sum_j f_{X_1,X_2}(x_{1,i}, x_{2,i}) = f_{X_1}(x_{1,i}), \quad \sum_i f_{X_1,X_2}(x_{1,i}, x_{2,i}) = f_{X_2}(x_{2,i}),$$

allora:

$$E(X_1 + X_2) = \sum_{i,j}(x_{1,i} + x_{2,i}) f_{X_1,X_2}(x_{1,i}, x_{2,i}) =$$

$$= \sum_i x_{1,i} \sum_j f_{X_1,X_2}(x_{1,i}, x_{2i}) + \sum_j y_j \sum_i f_{X_1,X_2}(x_{1,i}, x_{2i}) =$$

$$= \sum_i x_{1,i} f_{X_1}(x_{1,i}) + \sum_j x_{2,i} f_{X_2}(x_{2,i}) = \mu_1 + \mu_2.$$

Inoltre, usando questa additività della media,

$$Var(X_1 + X_2) = E[(X_1 + X_2 - \mu_1 - \mu_2)^2] =$$

$$= E[(X_1 - \mu_1)^2] + 2 \cdot E[(X_1 - \mu_1)(X_2 - \mu_2)] + E[(X_2 - \mu_2)^2]$$

$\square$

N.B. Covarianza sensibilmente positiva significa che le due variabili tendono a disporsi nella stesso verso rispetto alla loro media, come ad esempio nel caso della densità di autoveicoli e del CO nell'aria. Covarianza accentuatamente negativa rivela che $X_1 - \mu_1$ ed $X_2 - \mu_2$ tendono ad avere segno opposto, cioè una variabile è decrescente al crescere dell'altra, come nel caso di longevità e di stress. Quando la covarianza è nulla, le due v.a. si dicono *incorrelate.*

Def. Si dice *coefficiente di correlazione lineare* fra due v.a. X, Y

$$\rho_{X_1,X_2} := \frac{Cov(X_1, X_2)}{\sigma_1 \ \sigma_2} \ \equiv \ E[\ \frac{X_1 - \mu_1}{\sigma_1} \ \frac{X_2 - \mu_2}{\sigma_2}\ ].$$

N.B. Si ha per definizione $-1 \leq \rho \leq 1$. Due situazioni antitetiche sono il caso $\rho = 0$ (v.a. incorrelate) e il caso $\rho = \pm 1$ (v.a. totalmente correlate). Quando $\rho = \pm 1$, le v.a. X_1, X_2 sono *linearmente dipendenti:*

$$\exists a,\ b \ :\ X_2 \ = a\ X_1 \ + \ b$$

In particolare X_2 cresce ($a > 0$) o decresce ($a < 0$) al crescere di X_1 a seconda che $\rho = 1$ oppure $\rho = -1$.

Oss.

$$X_1, X_2 \text{ indipendenti} \quad \Longrightarrow \quad X_1, X_2 \text{ incorrelate}$$

Infatti, se X_1, X_2 sono indipendenti, la media del loro prodotto è uguale al prodotto delle medie:

$$E(X_1 \cdot X_2) = \int_{R^2} xy f_{X_1,X_2}(x_1, x_2) dx_1 \, dx_2 =$$

$$= \int_R x_1 f_{X_1}(x_1) dx_1 \int_R x_2 f_{X_2} dx_2 = \mu_1 \cdot \mu_2.$$

Tuttavia le due nozioni non sono equivalenti:

$$X_1, X_2 \text{ incorrelate} \quad \text{non implica} \quad X_1, X_2 \text{ indipendenti}$$

Ciò non toglie che, nel caso di due v.a. congiuntamente normali, esse sono indipendenti se e solo se $\rho = 0$. Analoga situazione se si considerano n v.a. congiuntamente normali.

Def. Un vettore aleatorio $X = (X_1, ..., X_n)$ a valori in R^n è detto Gaussiano se la sua densità di probabilità è data da:

$$f_X(x_1, ..., x_n) = \frac{\exp[\ -\frac{1}{2}(x - \mu_X)^T (C_X)^{-1} (x - \mu_X)]}{(2\pi)^{n/2} (\det\ C_X)^{1/2}},$$

per $x \in R^n$, dove $\mu_X \in R^n$ è il (vettore) valor medio di X e C_X è la matrice di covarianza:

$$\mu_X = \begin{pmatrix} \mu_1 \\ ... \\ \mu_n \end{pmatrix}, \qquad C_X = \begin{pmatrix} Cov(X_1, X_1) & ... & Cov(X_1, X_n) \\ ... & ... & ... \\ Cov(X_n, X_1) & ... & Cov(X_n, X_n) \end{pmatrix}$$

Un vettore aleatorio Gaussiano è spesso denotato $X = (X_1, ..., X_n) \sim N(\mu_X;\ C_X)$.

OSS. Poiché $Cov(X_i, X_j) = Cov(X_j, X_i)$ la matrice C_X è simmetrica.
Affinché la definizione abbia significato, si deve ammettere implicitamente che C_X sia matrice non singolare, in particolare che essa sia definita positiva. Perciò gli autovalori $\lambda_1, ...\lambda_n$ di C_X sono reali e positivi; e la matrice inversa C_X^{-1} è anch'essa definita positiva con autovalori $1/\lambda_1, ..., 1/\lambda_n$. Infine gli n autovettori di C_X possono scegliersi ortonormali e coincidono con gli n autovettori di C_X^{-1}.

Il seguente teorema è illuminante sulla struttura di covarianza di un vettore Gaussiano: esso si può "traslare" e "ruotare" opportunamente nello spazio R^n in modo da ottenere n variabili aleatorie Gaussiane di medie nulle e *indipendenti!*

Teorema *Sia X un vettore Gaussiano. Allora X si può esprimere come*

$$X \; = \; W \cdot Y + \; b, \quad b = (b_1, ..., b_n)$$

dove $b = (b_1, ..., b_n) \in R^n$, W una matrice $n \times n$ e $Y_1, ..., Y_n$ sono v.a. normali **indipendenti** *con media* 0. *La matrice W si può scegliere* **ortogonale.**

[[Richiamo di algebra lineare:

1) Una matrice è ortogonale sse l'inversa è uguale alla trasposta, $W^{-1} = W^T$. È in pratica una rotazione degli assi, a meno eventualmente della loro orientazione.

2) Ogni matrice simmetrica ammette n autovalori (distinti o variamente coincidenti) ed n autovettori che risultano ortogonali a due a due. Diagonalizzare una matrice simmetrica C significa trovare autovalori

$$\lambda_1, ..., \lambda_n \quad \text{soluzioni di} \;\; det(C - \lambda) = 0$$

e autovettori $u, ..., v$ di C, soluzioni non nulle di

$$Cu = \lambda_1 u, ..., Cv = \lambda_n v.$$

Una matrice W che diagonalizza C è tale che

$$W^T \cdot C \cdot W = \begin{pmatrix} \lambda_1 & 0 & ... & 0 \\ 0 & \lambda_2 & ... & 0 \\ ... & & & \\ 0 & 0 & ... & \lambda_n \end{pmatrix} \equiv diag(\lambda_1, ..., \lambda_n)$$

Se C è simmetrica, W è ortogonale e si ottiene prendendo come colonne n autovettori di C di norma 1.

3) Nel caso di una densità Gaussiana congiunta $f_X(x)$, supponiamo anzitutto che $(\mu_1, \mu_2, ..., \mu_n) = (0, 0, ..., 0)$: basta effettuare una banale traslazione. La matrice C che diagonalizziamo è la matrice di covarianza $C = (C_X)$, che è definita positiva. Quindi gli autovalori $\lambda_1, ..., \lambda_n$ sono tutti positivi e ammettono reciproci. I reciproci $1/\lambda_1$, ..., $1/\lambda_n$ sono gli autovalori di $(C_X)^{-1}$, per cui

$$x^T (C_X)^{-1} \cdot x \; = \; \sum_{i=1}^{n} \frac{y_i^2}{\lambda_i}, \quad \text{se} \quad X = WY,$$

$$f_X(x_1, ..., x_n) \quad \longrightarrow \quad f_Y(y_1, ..., y_n) = \frac{exp[-y_1^2/(2\lambda_1) - ... - y_n^2/(2\lambda_n)]}{(2\pi)^{n/2}\sqrt{\lambda_1 \cdot ... \cdot \lambda_n}}.$$

Quindi $\sqrt{\lambda_1}, .., \sqrt{\lambda_n}$ sono le deviazioni standard di $Y_1, ..., Y_n$.

4) Ed infine il significato geometrico: L'ellissoide $x^T \cdot (C_X)^{-1} x = 1$, cioè

$$\sum_{i,j} [(C_X)^{-1}]_{i,j} x_i \cdot x_j = 1 \quad \Longleftrightarrow \quad \frac{y_1^2}{\lambda_1} + ... + \frac{y_n^2}{\lambda_n} = 1.$$

ha per semiassi $\sqrt{\lambda_1}, ..., \sqrt{\lambda_n}$. Tale ellissoide, d'altra parte, è superficie di livello della densità congiunta $f_X(x_1, ..., x_n)$; esso unisce i punti che hanno la stessa "frequenza" f_X. [Le curve di livello sono familiari a chi guarda una carta geografica di montagna ("isòcore") o le previsioni del tempo ("isòbare")].
Quando $n = 2$ le superficie di livello sono ellissi: e i due semiassi dell'ellisse sono (proporzionali alle) deviazioni standard delle variabili aleatorie rese indipendenti Y_1, Y_2.]]

Es. Siano Y_1, Y_2 v.a. normali, indipendenti, con medie nulle e varianze $\sigma_1^2 = \frac{1}{2}$, $\sigma_2^2 = \frac{3}{2}$. Trovare la matrice di covarianza delle v.a.

$$X = WY, \quad \text{dove} \quad W = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$$

$$Cov(X_1, X_1) = E[(\cos\alpha Y_1 - \sin\alpha Y_2)^2] =$$
$$\cos^2\alpha E(Y_1^2) + \sin^2\alpha E(Y_2^2) = \frac{1}{2}\cos^2\alpha + \frac{3}{2}\sin^2\alpha.$$
$$Cov(X_1, X_2) = E[(\cos\alpha Y_1 - \sin\alpha Y_2) \cdot (\sin\alpha Y_1 + \cos\alpha Y_2)] =$$
$$= \cos\alpha\sin\alpha E(Y_1^2) - \sin\alpha\cos\alpha E(Y_2^2)$$
$$= \cos\alpha\sin\alpha(\frac{1}{2} - \frac{3}{2}) = -\cos\alpha\sin\alpha.$$
$$Cov(X_2, X_2) = E[(\sin\alpha Y_1 + \cos\alpha Y_2)^2] =$$
$$= \sin^2\alpha E(Y_1^2) + \cos^2\alpha E(Y_2^2) = \frac{1}{2}\sin^2\alpha + \frac{3}{2}\cos^2\alpha.$$

Es. Sia (X_1, X_2) un vettore Gaussiano con medie nulle e con matrice di covarianza

$$C_X = \begin{pmatrix} 5 & 5/2 \\ 5/2 & 5 \end{pmatrix}$$

per qualche $\sigma > 0$. Esprimere $X = WY$ con Y_1, Y_2 indipendenti: (1) trovare le varianze di Y_1, Y_2 e (2) descrivere la curva di livello della densità congiunta di X_1, X_2.

Il problema di esprimere X_1, X_2 come trasformazione lineare di v.a. Y_1, Y_2 *indipendenti* equivale al problema geometrico di *diagonalizzare* C_X. Ciò significa: trovare autovalori e autovettori di C_X e scrivere la matrice W che ha per colonne gli autovettori normalizzati.
(1)

$$0 = det\begin{pmatrix} 5-\lambda & 5/2 \\ 5/2 & 5-\lambda \end{pmatrix} \implies \lambda_1 = \frac{15}{2}, \; \lambda_2 = \frac{5}{2}.$$

Gli autovalori sono le varianze di Y_1, Y_2:

$$Y_1 \sim N(0, \frac{15}{2}), \quad Y_2 \sim N(0, \frac{5}{2})$$

(2) Gli autovettori u e v sono soluzioni non nulle di $C_X u = \lambda_1 u, \;\; C_X v = \lambda_2 v$,

$$5u_1 + \frac{5}{2}u_2 = \frac{15}{2}u_1 \implies u_1 = u_2$$

$$5v_1 + \frac{5}{2}v_2 = \frac{5}{2}v_1 \implies v_1 = -v_2$$

quindi gli autovettori sono

$$u = \begin{pmatrix} t \\ t \end{pmatrix}, \quad v = \begin{pmatrix} -s \\ s \end{pmatrix}, \qquad t, s \neq 0.$$

Li normalizziamo ($t^2 + t^2 = 1, \quad t = \pm 1/\sqrt{2}$), ottenendo gli autovettori ortonormali:

$$u = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \quad v = -\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}.$$

La matrice W che ha per colonne gli autovettori normalizzati la matrice ortogonale cercata. Quindi le v.a. normali X_1, X_2 possono esprimersi

$$\begin{cases} X_1 = \frac{1}{\sqrt{2}}Y_1 - \frac{1}{\sqrt{2}}Y_2 \\ X_2 = \frac{1}{\sqrt{2}}Y_1 + \frac{1}{\sqrt{2}}Y_2 \end{cases}$$

Poiché, in R^2, ogni matrice ortogonale W con determinante 1 è una rotazione, abbiamo

$$X = WY = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$

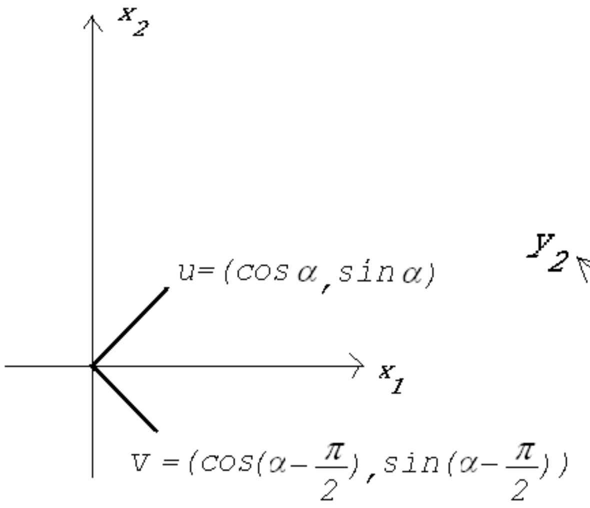
Nel nostro caso riconosciamo che l'angolo di rotazione è $\alpha = \pi/4$; una curva di livello della densità congiunta f_X è un'ellisse con assi propri ruotati di $\pi/4$ rispetto agli assi Ox_1, Ox_2 e con semiassi proporzionali 15/2 e 5/2.

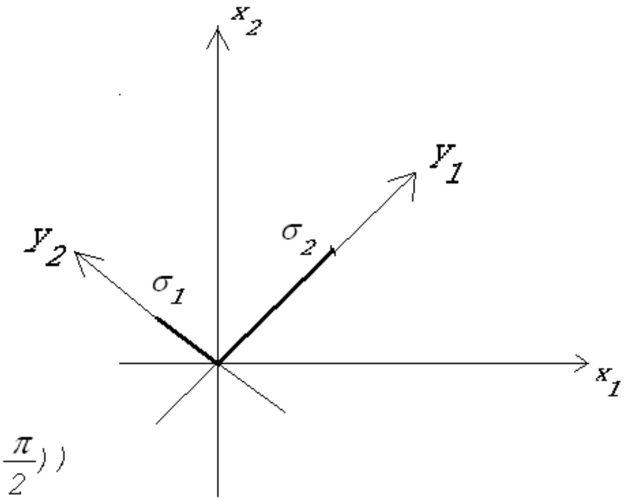
□

Es. Sia (X_1, X_2) un vettore Gaussiano con medie nulle e con matrice di covarianza

$$C_X = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}.$$

Esprimere $X = WY$ con Y_1, Y_2 indipendenti e W ortogonale.





Ciò significa: trovare gli autovalori e gli autovettori di C_X e scrivere la matrice W che ha per colonne gli autovettori normalizzati.

$$0 = det(C_X - \lambda I) = det \begin{pmatrix} 3-\lambda & 1 \\ 1 & 2-\lambda \end{pmatrix} = \lambda^2 - 5\lambda + 5$$

Gli autovalori $\lambda_{1,2} = (5 \pm \sqrt{5})/2$ hanno il significato di varianze di Y_1, Y_2. Gli autovettori u e v sono soluzioni di $C_X u = \lambda_1 u$, $C_X v = \lambda_2 v$:

$$3u_1 + u_2 = u_1(5+\sqrt{5})/2 \implies u_2 = u_1(-1+\sqrt{5})/2$$

$$3v_1 + v_2 = v_1(5+\sqrt{5})/2 \implies v_2 = v_1(-1-\sqrt{5})/2$$

Gli autovettori

$$u = (t, t(-1+\sqrt{5})/2), \quad v = (s, s(-1-\sqrt{5})/2), \ \ t, s \neq 0$$

indicano gli assi ruotati Y_1, Y_2. L'angolo di rotazione di Y_1 rispetto ad X_1 è

$$\alpha = \ arctan(\frac{u_2}{u_1}) = arctan(-1+\sqrt{5})/2) = 0.553$$

[Controprova: l' angolo fra Y_2 ed X_1

$$\beta = arctan(\frac{v_2}{v_1}) = arctan(-1-\sqrt{5})/2) = -1.0163$$

è proprio ottenibile come $\alpha - \pi/2 = 0.553 - 1.57 \simeq -1.163$.]
In conclusione:

$$W = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \quad \text{con} \quad \alpha = 0.553.$$

ESERCIZI PER METODI STATISTICI, Ing. Inf./El./Telecom.

1. Insiemi e probabilità.
2. Variabili aleatorie discrete.
3. Variabili aleatorie continue, approssimazione.
4. Calcolo di leggi, condizionamenti, sistemi di più v.a.

INSIEMI E PROBABILITÀ

1.1 - Due carte sono estratte senza rimpiazzo da un mazzo di 40 ben mescolato. Si calcoli la probabilità che esse siano la prima un asso e la seconda nè asso nè fante.

R.] 8.2%

1.2 - Un gene è composto di due alleli, ciascuno può essere di tipo A oppure a. Nella popolazione vi sono 3 tipi di individui: di tipo AA, Aa, e aa. Ciascun genitore trasmette al figlio uno dei due alleli scelto a caso. Sapendo che inizialmente le proporzioni dei tre tipi sono

$$AA: \frac{1}{3} \quad Aa: \frac{1}{5} \quad aa: \frac{7}{15}$$

quali saranno le proporzioni dei tipi AA, Aa, aa alla generazione successiva?

R.] AA: 18.78%, ...

Soluzione Sia F_A ="il primo dei due genitori trasmette A". Inoltre distinguiamo

$$B-1,\ B_2,\ B_3 = \text{"il primo genitore è } AA, \text{ oppure } Aa, \text{ oppure } aa\text{"}.$$

Infatti sappiamo che nei tre casi cambia la probabilità di trasmettere A:

$$P(F_A|B_1) = 1; \quad P(F_A|B_2) = \frac{1}{2}; \quad P(F_A|B_3) = 0.$$

Allora, per il teorema di probabilità totale,

$$P(F_A) = P(F_A|B_1) \cdot P(B_1) + P(F_A|B_2) \cdot P(B_2) + P(F_A|B_3) \cdot P(B_3) =$$

$$= 1 \cdot \frac{1}{3} + \frac{1}{2} \cdot 15 = \frac{13}{30}$$

Allo stesso modo, l'evento G_A ="il secondo dei due genitori trasmette A" ha una probabilità $P(G_A) = \frac{13}{30}$. In virtù dell' indipendenza,

$$P(\text{"un figlio sia "}AA) = P(F_A \cap G_A) = \frac{13}{30}\frac{13}{30} = \frac{169}{900}$$

$\square$

1.3 - Un compilatore assegna ad ognuna delle variabili che intervengono in un programma una cella di memoria a caso, con indipendenza da una variabile all'altra. In caso di conflitto (cioè se due variabili sono assegnate alla stessa cella), l' operazione di assegnazione deve essere ripetuta. Se vi sono 100 celle di memoria e 4 variabili, qual è la probabilità che si verifichi un conflitto?

R.] 5.89%

1.4 - I componenti prodotti da una certa ditta presentano due tipi di difetti con percentuale del 2% e del 6% rispettivamente e con indipendenza. Qual è la probabilità che un componente presenti il difetto 1, sapendo che è difettoso?

R.] 25.38%

1.5 - Tre malattie A, B, C - e solo queste - causano un certo sintomo con probabilità $f_A = 9/10$, $f_B = 6/10$, $f_C = 4/10$. In Emilia d' estate un individuo è affetto da ciascuna malattia con probabilità $p_A = 0.1\%$, $p_B = 1\%$, $p_C = 5\%$. Sapendo che un paziente emiliano questa estate presenta il sintomo, qual è la probabilità che egli abbia la malattia B?

R.] 22.3%

1.6 - Un dado a 4 facce è lanciato 3 volte. Qual è la probabilità di ottenere "quattro" almeno una volta?

R.] 57.81%

1.7 - Qual è la probabilità che almeno due fra 4 coetanei nati nella stessa stagione festeggino il compleanno nello stesso giorno? (una stagione= 92 giorni).

R.] 6.39%

1.8 - In quanti modi 10 persone possono sedersi su una panchina che ha solo 4 posti?

R.] 5040

1.9 - In uno scaffale ci sono 10 libri, 3 di matematica e 7 di fisica; si trovi la probabilità che i 3 libri di matematica si trovino insieme.

R.] 6.66%

1.10 - L'urna I contiene 3 palline rosse e 5 bianche, mentre l'urna II contiene 4 rosse e 2 bianche. Si sceglie una pallina a caso dall'urna I e la si mette, senza osservare il colore, nell'urna II: si estrae poi una pallina dall'urna II. Qual è la probabilità che la pallina cosi' estratta sia bianca?

R.] 37.5%

1.11 - Una fabbrica produce componenti elettronici, che escono da due linee di produzione, A e B, nelle proporzioni del 35% e del 65% rispettivamente. La Linea A ha una percentuale di pezzi difettosi del 10%, contro il 20% della linea B. Con quale probabilità un chip prodotto da quella fabbrica è difettoso?

R.] 16.50%

1.12 - La popolazione di una regione è affetta da virus Ebola con probabilità 1%. Il miglior test per il virus ha affidabilità 80% tanto sui sani quanto sui malati. Una persona è scelta casualmente e risulta positiva. Qual è la probabilità che sia effettivamente affetta da Ebola?

R.] 3.9%

1.13 - Uno studente è sottoposto a un quiz con 4 risposte possibili. Se ha studiato, egli risponderà certamente in maniera esatta, altrimenti sceglierà una risposta a caso tra le 4 disponibili. Supponiamo che abbia abbia studiato con probabilità 1/2 e che, sottoposto al quiz, abbia scelto la risposta esatta. Sulla base di ciò, qual è la probabilità che abbia studiato davvero?

R.] 80%

VARIABILI ALEATORIE DISCRETE

2.1 - Una compagnia ha un aereo di 19 posti e accetta 21 prenotazioni perché sa che il 10% dei prenotati non si presenta. Con quale probabilità almeno un passeggero resterà a terra?

a] 36.47%

b] 27.46% c] 17.56% d] 21.09%

Risoluzione Sia Z = "il n.o di passeggeri che si presentano fra i 21 che si sono prenotati". Allora Z è binomiale con $n = 21$ e $p = \frac{9}{10}$.

$$P(Z = 20) + P(Z = 21) = \binom{21}{20} (\frac{9}{10})^{20} \frac{1}{10} + \binom{21}{21} (\frac{9}{10})^{21} = 0.3647$$

2.2 - Con quale probabilità esce una o due volte "sette" in 5 lanci di una coppia di dadi?

a] 56.26%

2.3 - Una coppia di dadi è lanciata 3 volte. Con quale probabilità la somma uscente sarà "cinque" nemmeno per una volta?

a] $\sim$ 70%

2.4 - Un calcolatore è collegato a una rete che permette l'accesso ad un massimo di 20 persone. Collegati a questa rete vi sono i terminali di 22 operatori, ognuno dei quali, a un dato istante, richiede con probabilità $p = 0.8$ di essere connesso al calcolatore centrale. Qual è la probabilità che a un dato istante, la rete sia satura (cioè che tutti i 20 accessi siano usati)?

a] 4.7%

2.5 - Una fabbrica produce componenti elettronici, che escono da due linee di produzione, A e B nelle proporzioni del 30% e 70%. La Linea A ha una percentuale di pezzi difettosi del 10%, contro il 17% della linea B. Si considera una confezione di 10 chips di tale fabbrica: con quale probabilità la confezione contiene esattamente un chip difettoso?

a] $\simeq 34\%$

Risoluzione $D =$"un chip è difettoso". Per il teorema di probabilità totale, rispetto alla partizione data $A,\ B$,

$$P(D) = P(D|\ A) \cdot P(A) + P(D|\ B) \cdot P(B) =$$

$$= \frac{1}{10} \cdot \frac{3}{10} + \frac{17}{100} \cdot \frac{7}{10} = 0.03 + 0.119 = 0.149$$

Ora sia $Z =$"n. di chip difettosi nell'ambito di 10". Z è allora binomiale con $n = 10$ e $p = 0.149$.

$$P(Z = 1) = \binom{10}{1} (0.149)(0.851)^9 = 10 \cdot (0.149) \cdot (0.234) = 0.34$$

$\square$

2.6 - Una compagnia di assicurazioni ha 3000 assicurati contro un dato rischio che ha probabilità 0.1% di colpire ogni singolo assicurato in un anno. Sapendo che il numero X di indennizzandi in un anno è di Poisson, che la compagnia indennizza ciascuno con 80000 Euro, che percepisce da ogni assicurato un premio annuale di 100 Euro, quali sono il valor medio e la varianza del beneficio annuale della compagnia?

a] $\mu = 60000, \sigma^2 = 192 \cdot 10^8$

Risoluzione Se ha 3000 assicurati con probabilità 1/1000 di incidente individuale all'anno, allora $X =$"numero di infortunati all'anno" è di Poisson con media $\lambda = 3000 \cdot \frac{1}{1000} = 3$. Se l'indennizzo individuale è $8 \cdot 10^4$ Euro, se il premio annuale è 100 Euro, allora il beneficio annuale è la variabile

$$Y = 100 \cdot 3000 - X \cdot 8 \cdot 10^4$$

Essa ha valor medio $E(Y) = 3 \cdot 10^5 - E(X) \cdot 8 \cdot 10^4 = 300000 - 3 \cdot 80000$ Euro. Essa ha varianza $(8 \cdot 10^4)^2 Var(X) = 64 \cdot 3 \cdot 10^8$. $\square$

2.7 - La memoria secondaria di un calcolatore è composta da 30 unità disco in ognuna delle quali sono archiviati 100 file. Durante l'esecuzione di un programma è necessario accedere a 40 di questi file, tutti diversi. Qual è la probabilità che sia necessario usare l'unità 1? (Cioè qual è la probabilità che tra i 40 file ve ne sia uno contenuto nell'unità 1?)

a] $1 - \frac{2900 \cdot 2899 \cdot \ldots \cdot 2861}{3000 \cdot 2999 \cdot \ldots \cdot 2961}$

Risoluzione Dire che il programma deve accedere al disco "1" equivale a dire che tra le 40 registrazioni ce nè almeno una nel disco "1". È una ipergeometrica la v.a. $Z_1 =$ "numero di files necessari al programma che si trovano nell'unità 1". Infatti l'insieme dei

30 dischi è un'urna contenente $r = 100$ files nel primo disco e $b = 2900$ files negli altri, e facciamo $n = 40$ estrazioni.

$$P(Z_1 = k) = \frac{\binom{r}{k}\binom{b}{n-k}}{\binom{r+b}{n}} \implies \implies P(Z_1 > 0) = 1 - P(Z_1 = 0) =$$

$$= 1 - \frac{\binom{100}{0}\binom{2900}{40}}{\binom{3000}{40}}$$

$$= 1 - \frac{2900!/(40!2860!)}{3000!/(40!2960!)}$$

□

2.8 - Un canale di trasmissione dati può ricevere messaggi binari da due sorgenti diverse A e B con probabilità $\frac{1}{2}$ ciascuna. Ognuna delle due sorgenti produce messaggi in cui i bit successivi sono tra di loro indipendenti. Ma per la sorgente A i bit possono essere 1 o 0 con probabilità $\frac{1}{2}$, mentre per B il valore 1 si verifica con probabilità $\frac{1}{4}$ e 0 con probabilità $\frac{3}{4}$. Un messaggio di lunghezza 10 viene ricevuto e in esso si osservano 4 bit uguali a 1. Qual è la probabilità che si tratti della sorgente A?

a] 58%

Risoluzione Il testo lascia intendere che il messaggio di 10 bits viene o interamente da A o interamente da B. Allora pongo: $X =$"numero di 'uno' in un messaggio di dieci bits".

$$P(A|X = 4) = \frac{P(X = 4|A)P(A)}{P(X = 4|A)P(A) + P(X = 4|B)P(B)}.$$

Ma $P(X = 4|A) = P(X_A = 4)$, dove

$$X_A = \text{"numero di "1" in } n = 10 \text{ bits sapendo che } p = 1/2 \text{ "}.$$

Cos'i $P(X = 4|B) = P(X_B = 4)$, dove

$$X_B = \text{"numero di "1" in } n = 10 \text{ bits sapendo che } p = 1/4 \text{ "}.$$

$$\implies P(A|X = 4) = \frac{P(X_A = 4)P(A)}{P(X_A = 4)P(A) + P(X_B = 4)P(B)} =$$

$$\frac{\binom{10}{4}(1/2)^{10} \cdot \frac{1}{2}}{\binom{10}{4}(1/2)^{10} \cdot \frac{1}{2} + \binom{10}{4}(1/4)^4 \cdot (3/4)^6 \frac{1}{2}}$$

$$= \frac{0.00097}{0.00097 + (0.0039) \cdot (0.1779)} \simeq 0.58$$

□

2.9 - Per depurare un lago artificiale in cui si è rilevato un parassita, si esegue più volte un trattamento. Il trattamento riduce il numero medio di parassiti per litro, λ, portandolo a $\lambda/6$. Se inizialmente il numero di parassiti è una v.a. di Poisson di media 5, quanti interventi occorrono perché al termine ogni litro abbia parassiti con probabilità inferiore ($\leq$) a 0.1% ?
a] 5

Risoluzione Dopo 4 trattamenti Z_4 =”numero parassiti per litro” è una v.a. di Poisson con media $\mu = 5 \cdot (1/6)^4$. Quindi

$$P(Z_4 = 0) = e^{-5\cdot(1/6)^4} < 0.999$$

Dopo il quinto trattamento si ha

$$P(Z_5 = 0) = e^{-5\cdot(1/6)^5} > 0.999$$

□

2.10 - Una moneta è lanciata 3 volte. Se X è il numero di teste che si verificano nei lanci, e se F indica la funzione distribuzione, quanto vale $F(2.9)$?
a] 7/8

2.11 - In un’urna ci sono 5 palline bianche e 3 nere e si estraggono a caso e senza rimpiazzo due palline. Se X è il numero di bianche estratte ed f la sua funzione di probabilità, determinare $f(2)$.
a] 5/14

2.12 - Una v.a. X discreta ha come valori possibili $x = 1, 2, 3, 4,$ con funzione distribuzione

$$\text{per } \; x = 1,2,3,4 \qquad F(x) = \frac{1}{8}, \frac{3}{8}, \frac{3}{4}, 1$$

Detta μ la media, quanto vale la probabilità $P(X \leq \mu)$?
a] 37.5%

Risoluzione Poiché $F(x) = \sum_{i \square x} f(x_i)$,

$$f(4) = F(4) - F(3) = 1/4, \quad f(3) = F(3) - F(2) = 3/8,$$

$$f(2) = F(2) - F(1) = 2/8, \quad f(1) = 1/8$$

$$\Longrightarrow \mu = \sum_i x_i f(x_i) = \frac{1}{8} \cdot 1 + \frac{2}{8} \cdot 2 + \frac{3}{8} \cdot 3 + \frac{1}{4} \cdot 4 = \frac{22}{8} = 2.75$$

$$\Longrightarrow F(2.75) = F(2) = \frac{3}{8} = 0.375$$

2.13 - Una v.a. X discreta ha i tre valori $x = 1, 2, 3$, con funzione distribuzione

$$\text{per } \ x = 1, 2, 3 \qquad F(x) = \frac{1}{8}, \frac{3}{8}, 1$$

Qual è la varianza di X?
a] 1/2

2.14 - Se il numero di annegamenti in un anno è pari a $0,3$ su centomila, si chiede la probabilità che in una città di duecentomila abitanti ci siano 3 o 4 annegamenti.
a] 2.27%

2.15 - Un processo di produzione di viti è controllato ogni ora ispezionando n viti, scelte a caso tra quelle prodotte in quell'ora. Se una o più viti sono difettose, il processo è fermato ed esaminato. Il produttore vuole probabilità 95% di fermare il processo quando l'8% delle viti è difettoso. Quanto deve essere grande n?
a] $n \geq 36$

Risoluzione $\quad X =$"numero di difettose nell'ambito di n viti sapendo che $p = \frac{8}{100}$.". Il produttore vuole che

$$P(X \geq 1) \geq 95\% \ \text{ cioè} 1 - \binom{n}{0}(\frac{8}{100})^0(\frac{92}{100})^n \geq \frac{95}{100}$$

$$\iff \quad (\frac{92}{100})^n \leq \frac{5}{100} \quad \iff \quad n\log(0.92) \leq \log(0.05)$$

cioè $\ -n(0.0833) \leq -2.9957, \ $ cioè $n \geq 35.96$ $\square$

2.16 - Una scatola ha fondo quadrato di lato 1 metro, al centro del quale vi è un foro circolare di diametro 10 *cm*.Nella scatola sono gettate a caso e indipendentemente 10 palline di diametro piccolo (cioè $<<$ 10 *cm*). Con quale probabilità alla fine dei lanci si trovano nella scatola 7 palline?
a] $\sim 5.3 \cdot 10^{-5}$.

2.17 - Un principiante di tiro al piattello lo colpisce con probabilità 2/9. Qual è la probabilità che gli occorrano almeno 5 tiri per colpirlo la prima volta?
a] $\simeq 8\%$

VARIABILI ALEATORIE CONTINUE, APPROSSIMAZIONE

3.1 - Si consideri la costante c tale che

$$f(x) = \begin{cases} cxe^{-2x}, & x \geq 0 \\ 0, & \text{altrove} \end{cases}$$

sia la funzione densità di una v.a. X. (a) Determinare c. (b) Trovare la probabilità che due estrazioni indipendenti di X cadano nell'intervallo $(1, 3)$.

a] $c = 4$ b] 0.15

Risoluzione a] Integrando per parti,

$$1 = c\int_0^\infty xe^{-2x}dx = c\frac{1}{2}[\frac{e^{-2x}}{-2}]_0^\infty$$

che vale $c/4$. b] $F(x) = \int_0^x 4x'e^{-2x'}dx' = 1 - 2xe^{-2x} - e^{-2x}$ quindi $P(X \in (1,3)) = F(3) - F(1) = 1 - 7\cdot e^{-6} - (1 - 3\cdot e^{-2})) = 0.406 - 0.017 = 0.389$. Per indipendenza, la probabilità che entrambi i numeri estratti siano in $(1,3)$ è il prodotto delle probabilità: $(0.389)^2 = 0.151$.

3.2 - Si consideri la costante k tale che

$$F(x) = \begin{cases} 0, & x \leq 0 \\ k(1-e^{-x})^2, & x > 0 \end{cases}$$

sia la funzione distribuzione di una variabile aleatoria X; trovare c tale che $P(X > c) = 90\%$.

a] 0.38

3.3 - Il numero di chilometri (misurato in migliaia) che può percorrere un certo tipo di gomme è una v.a. X con densità $f(x) = 0.05e^{-0.05\ x}$, se $x > 0$, e 0 altrove. Trovare la probabilità: (a) che una gomma duri da 10000 a 30000 Km; (b) che due gomme durino entrambe almeno 20000 Km.

a] 0.38 b] 0.36

3.4 - Il voto ad una prova di ingresso è distribuito normalmente, e il miglior 10% dei candidati verrà assunto. Ad esame finito, il voto medio è stato 72 e la deviazione standard è stata 9. Qual è il voto minimo che un candidato deve ottenere per essere assunto?

a] 84.

Risoluzione Essendo $\mu = 72$, $\sigma = 9$,

$$P(X \geq c) = \frac{10}{100} = P(\frac{X-\mu}{\sigma} \geq \frac{c-\mu}{\sigma}) = P(N(0,1) \geq \frac{c-72}{9})$$

$$\Longrightarrow \quad \frac{c-92}{9} = 1.29 \quad \Longrightarrow \quad c = 72 + 9\cdot(1.29) = 83.61$$

□

3.5 - Supponiamo che k sia la costante che rende densità di probabilità la funzione: $f(x) = kx$, se $0 \leq x \leq 3$, ed $f(x) = 0$ altrove. Trovare c tale che $P(c \leq X \leq 3 - c) = 95\%$.

a] $c = 15/200$.

3.6 - Ogni abitante di un paese in un anno ha probabilità p di vincere una lotteria. Se il paese ha 9000 abitanti, con quale probabilità nel 2007 si conteranno : almeno due vincite, nel caso $p = 1/3000$? (b) da 80 a 100 vincite, nel caso $p = 0.01$?

a] 0.80 b] 0.70

3.7 - La stazione Radio Bruco trasmette il segnale orario allo scoccare di ogni ora. L'ascoltatore tipo sintonizza il proprio radioricevitore sulla stazione Radio Bruco a un istante uniformemente distribuito tra le ore 7 : 10 e le ore 19 : 30 nella giornata. Calcolare la probabilità che l'ascoltatore riceva il segnale orario entro 5 minuti dalla sintonizzazione su Radio Bruco (si adotti il minuto come unità di tempo).
a] 8.1%

Risoluzione. Dalle 7 : 10 alle 19 : 30 ci sono 740 minuti. La v.a. "istante di sintonizzazione" è $X \sim U[0, 740]$, con densità di probabilità uguale a $1/740$ in tale intervallo. Il segnale orario è ai minuti $50 + k \cdot 60$, con $k = 0, 1, 2, ..., 11$. La probabilità richiesta è la somma di 12 aree di piccoli rettangoli di base 5 e altezza $1/740$, cioè $(5/740) \cdot 12 = 0.081$.□

3.8 - Un segnale consiste in una parola di n bit, ciascuno dei quali può assumere i valori 0 oppure 1. Nel corso della trasmissione ogni bit con probabilità $p = 0.01$ può essere distorto. Qual è la probabilità che un segnale di 1000 bit contenga almeno 10 bit distorti?
a] 56.35%

3.9 - Un segnale consiste in una parola di n bit, ciascuno dei quali può assumere i valori 0 oppure 1. Nel corso della trasmissione ogni bit con probabilità $p = 0.01$ può essere distorto. Per ridurre la distorsione si usa il seguente protocollo: ogni bit viene trasmesso 3 volte e ed il vero valore viene deciso a maggioranza: il bit viene posto uguale ad A ($A = 0$ oppure 1) se vi sono almeno due valori A tra quelli ricevuti. Qual è la probabilità che un segnale di 1000 bit contenga bit distorti?
a] 25.77%

Risoluzione Considerando il singolo bit, ha probabilità $p = \frac{1}{100}$ di essere distorto. Ma se viene trasmesso tre volte, si deve considerare la v.a. $X =$ "numero di trasmissioni distorte del bit nell'ambito di tre trasmissioni, sapendo che la probabilità di distorsione in una singola trasmissione è $p = \frac{1}{100}$". Allora $X \sim B(n = 3, p = 0.01)$. In base al protocoillo descritto, il rischio è una probabilità p' che dovrebbe risultare decisamente minore di p: la probabilità di avere due bit distorti su 3 ovvero 3 bit distorti su 3:

$$p' = P(X = 2) + P(X = 3) = \sum_{k=2}^{3} \binom{3}{k} (\frac{1}{100})^2 (\frac{99}{100}) + \binom{3}{3} (\frac{1}{100})^3$$

$$= 0.000297 + 0.000001 = 2.98 \cdot 10^{-4}$$

La domanda chiede di trattare la v.a. binomiale $Y =$"numero di bit distorti nell'ambito di $n = 1000$ bit, sapendo che $p' = 2.98 \cdot 10^{-4}$. Questa binomiale è abbastanza prossima alla normale? No, perché si esigerebbe n tanto grande da rendere np' dell' ordine di alcune unità (qui la condizione pratica $np' \geq 5$ è lontana dall'essere verificata). Dunque la risposta è:

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \binom{n}{0} (p')^0 (1 - p')^n =$$

$$= 1 - \binom{1000}{0} (1 - 2.98 \cdot 10^{-4})^{1000} = 1 - 0.7423 = 0.2577$$

3.10 - Un dado viene lanciato 900 volte e indichiamo con X il numero di volte in cui esce il 6. Sappiamo che esiste una partita di dadi truccati che producono il 6 con probabilità 2/9. Per decidere se il dado è di questi ultimi usiamo questa procedura: lo lanciamo 900 volte e decidiamo che è truccato se il 6 esce almeno ($\geq$) 180 volte. Qual è la probabilità che un dado truccato venga effettivamente individuato?

a] 95%.

Risoluzione Abbiamo due v.a. X, Y binomiali in ballo: entrambe sono del tipo "numero di sei uscenti in 900 lanci"; ma una nell' ipotesi che $p = \frac{1}{6}$ (dado equo) e una nell'ipotesi che si abbia $p' = \frac{2}{9}$ (dado truccato).

$$X \sim B(n = 900,\ p = \frac{1}{6}), \quad \text{con } \mu = 150;$$

$$Y \sim B(n = 900,\ p' = \frac{2}{9}), \quad \text{con } \mu' = 200$$

La procedura descritta sceglie 180 come soglia di decisione statistica. Si domanda qual è la probabilità di accorgerci del trucco nell'ipotesi che il dado sia truccato. Cioè si domanda:

$$P(Y \geq 180) \simeq P(N(np', np'q') \geq 179.5) = \ P(N(0,1) \geq \frac{179.5 - 200}{\sqrt{200 \cdot 7/9}}) =$$

$$= P(N(0,1) \geq \ -1.6437) = \ P(N(0,1) \leq \ 1.6437) = \ 0.949$$

Tale probabilità viene chiamata in statistica $1 - \beta =$ "potenza del test". Invece $P(X \geq 180$ viene chiamata $\alpha =$ "livello di significatività del test".

APPLICAZIONI DEL T. DI LIMITE CENTRALE

4.1 - Il tempo di sopravvivenza di una lampadina è una v.a. di media $\mu = 8$ giorni. Appena brucia, essa è sostituita. Trovare il numero di lampadine occorrenti perché esse siano sufficienti per un anno con probabilità 90%.

4.2 - Un ubriaco asimmetrico dista 80 passi da casa e fa un passo verso casa con probabilità 3/5 e un passo di allontanamento con probabilità 2/5. Quanti passi deve fare per avere probabilità 30% di arrivare a casa?

4.3 - L'audience di un programma ogni giorno aumenta di un punto con probabilità 1/5, diminuisce con probabilità 2/5 e resta invariata con probabilità 2/5. Il proprietario della emittente cancella dal palinsesto i programmi quando la loro audience diminuisce di 8 punti. Con quale probabilità il programma sarà cancellato entro 100 giorni?

4.4 - In un certo istante l'ampiezza X di un segnale emesso da un generatore di segnali aleatori è una v.a. normale $N(0.2,\ 0.36)$. L'onda passa attraverso un selezionatore dando in uscita $Y = X^+$, dove $x^+ = x$ se $x > 0$ ed $x^+ = 0$ se $x < 0$. Trovare: a) $P(Y = 0)$; b) $P(-0.3 \leq Y \leq 0.3)$.

a] $\simeq 0.37$; b) $\simeq 0.56$

LEGGE CONGIUNTA DI DUE V.A.

4.5 - Due numeri X ed Y sono scelti a caso e indipendentemente nell'intervallo $[0, 1]$. Indichiamo con Z la distanza fra loro. Calcolare $P(1/5 < Z < 2/5)$.

a] 28%

Risoluzione Dalle v.a. $X \sim U[0, 1]$, $Y \sim U[0, 1]$, indipendenti, costruiamo geometricamente le proprietà della v.a. $Z = |X - Y|$. Nel piano cartesiano, nel quadrato di vertici $(0, 0), (1, 0), (0, 1), (1, 1)$ tracciamo la diagonale $x - y = 0$, ossia $z = 0$. kk La zona $\{Z \equiv |x - y| > t\}$ sarà delimitata dalle rette $\{|y - x| = t\} \equiv \{y = x \pm t\}$. In particolare la retta $y = x + t$ delimita dal basso il triangolo di vertici $(1, 1)$, $(0, t)$, $(1 - t, 1)$; l'altra delimita dall'alto il triangolo di vertici $(0, 1)$, $(t, 0)$, $(1, 1 - t)$. Ora l' unione di questi due triangoli equivale a un quadrato di lato $(1 - t)$:

$$P(|X - Y| > t) = P(Z > t) = (1 - t)^2.$$

Per calcolare la probabilità richiesta basta fare la differenza fra due aree del tipo $P(Z > t)$:

$$P(\frac{1}{5} < Z < \frac{2}{5}) = P(Z \leq \frac{2}{5}) - P(Z \leq \frac{1}{5}) = 1 - P(Z > \frac{2}{5}) - \{1 - P(Z > \frac{1}{5})\}$$

$$= P(Z > \frac{1}{5}) - P(Z > \frac{2}{5}) = (4/5)^2 - (3/5)^2 = 7/25 = 0.28$$

□

4.6 - In due punti di un lago si misura l'intensità del suono causato da rumore di fondo generale (detto "rumore di ambiente"). Siano X, Y le due v.a. intensità del suono. Supponiamo che la loro legge congiunta sia continua con densità

$$f_{X,Y}(x, y) = \begin{cases} 4xe^{-x}ye^{-2y}, & x, y \geq 0 \\ 0, & \text{altrove} \end{cases}$$

Sia $U = \min(X, Y)$ l'intensità minima di rumore. Calcolare $P(U \geq 0.2)$.

a] $\simeq 92\%$

4.7 - Un punto è scelto a caso nel piano con densità

$$f(x, y) = (2\pi)^{-1} \exp(-\frac{x^2 + y^2}{2}).$$

Indichiamo con Z la distanza del punto dall'origine. Calcolare $P(Z > 1/3)$.

a] 94.59%

4.8 - Sia X una v.a. con densità

$$f_X(x) = \frac{1}{3}|x|e^{-|x|^2/3}, x \in R$$

Se $Y = X^2$, calcolare la densità di Y.

4.9 - Sia X una v.a. con densità $f_X(x) = \frac{1}{3}e^{-x/3}$ per $x > 0$, e nulla altrove. Calcolare la densità di Y: a) se $Y = g(X) = exp(-2X)$; b) se $Y = g(X) = \sin X$.

a] $g : R^+ \to (0,1)$; $f_Y(y) = (1/6)y^{-5/6} \cdot I_{(0,1)}(y)$

b]

$$g : R^+ \to (-1,1); \quad f_Y(y) = (1-y^2)^{-1/2} \sum_{k=1}^{\infty} (1/3)exp(-x_k(y)/3) \cdot I_{(-1,1)}(y), \text{ dove } x_k(y) : \sin[x_k(y)]$$

4.10 - È data una v.a. $X \sim N(0,1)$. Determinare la densità di probabilità $f_Y(y)$ della v.a. lognormale $Y = e^{-X}$.

a] $f_Y(y) = (y\sqrt{2\pi})^{-1} \exp -(\log y)^2/2 \cdot I_{R^+}(y)$

4.11 - (α) Sia X una v.a. uniforme in $[-pi/3, \pi/3]$.
(β) Sia X una v.a. Gamma, $Gamma(2,1)$ [cioè: $f(t) = te^{-t}I_{R^+}(t)$]
(γ) Sia X una v.a. normale, $N(0,1)$
Quali sono le densità di
a) $Y = \tan X$?
b) $Y = \cos X$?
c) $Y = \sqrt{|X|}$?
d) $Y = log(1 + X^2)$?

4.12 - In un certo istante l'ampiezza X di un segnale emesso da un generatore di segnali aleatori è una v.a. normale $N(0.2,\ 0.36)$. L'onda passa attraverso un selezionatore dando in uscita $Y = X^+$, dove $x^+ = x$ se $x > 0$ ed $x^+ = 0$ se $x < 0$. Trovare $P(-0.3 \leq Y \leq 0.3)$.

a] $\simeq 56\%$

UNA FUNZIONE DI PIÚ V.A.

4.13 - Un componente elettronico è formato da tre elementi in serie, (cioè non funziona se uno almeno dei tre non funziona) aventi ciascuno un tempo di vita esponenziale di parametri $\lambda = 1/5$, $\mu = 2/5$, $\gamma = 1/10$ rispettivamente. I tre tempi di vita sono indipendenti. Indichiamo con T la v.a. "tempo di vita" del componente. Quanto vale $E(T)$?

a] 1.43

Risoluzione Poiché i tre elementi sono in serie, T coincide con $\min\{T_1, T_2, T_3\}$, dove T_i è il tempo di vita dell'i-esimo elemento. Per definizione di minimo e $P(T \geq t) = P(T_1 \geq t,\ T_2 \geq t, T_3 \geq t)$, e questo è uguale al prodotto delle tre probabilità in virt'u della indipendenza

$$P(T \geq t) = \int_t^{\infty} \frac{2}{10} e^{-2x/10} dx \int_t^{\infty} \frac{4}{10} e^{-4x/10} dx \int_t^{\infty} \frac{1}{10} e^{-x/10} dx$$

$$= [-e^{-2x/10}]_t^{\infty} [-e^{-4x/10}]_t^{\infty} [-e^{-x/10}]_t^{\infty} = e^{(-7/10)t}$$

$$\Longrightarrow \quad P(T \leq t) = 1 - e^{(-7/10)t} \quad \Longrightarrow \quad f_T(t) = \frac{7}{10} e^{-7t/10}$$

Ma questa è una v.a. esponenziale e sappiamo che la media è l'inverso del parametro. Perci'o: $E(T) = \frac{10}{7} = 1.43$ □

4.14 - Un componente elettronico è formato da due elementi uguali in parallelo (cioè non funziona se ambedue han cessato di funzionare); ciascuno dei due a sua volta è formato da due elementi in serie (cioè non funziona se almeno uno dei due non funziona). Questi due elementi in serie hanno tempo di vita esponenziale di parametri rispettivamente $\lambda = 2/10$, $\mu = 1/10$ e si assume l'indipendenza. Qual è il tempo medio di vita del componente globalmente?

a] 5

4.15 - Se X_1, X_2 sono v.a. normali indipendenti di medie 0 e varianze uguali a 5, si domanda la densità della v.a. $Y = X_1^2 + X_2^2$.

CONDIZIONAMENTI, LEGGI CONDIZIONATE

4.16 - Calcolare $P(XY < \frac{1}{4} | X > 1/2)$, sapendo che X, Y sono v.a. indipendenti e uniformi in $[0, 1]$

a] $\frac{1}{2} \log(2)$.

Risoluzione Nel quadrato $[0, 1] \times [0, 1]$, ci interessa il tratto del ramo di iperbole $xy = \frac{1}{4}$ corrispndente a $\frac{1}{2} \leq x \leq 1$. La probabilità dell' intersezione dei due eventi, $P(XY < \frac{1}{4},\ X > 1/2)$, è l'area della superficie sottesa a tale tratto di grafico:

$$P(XY < \frac{1}{4} | X > 1/2) = \frac{1}{P(X > 1/2)} \cdot P(XY < \frac{1}{4},\ X > 1/2) =$$

$$= \frac{1}{1/2} \int_{1/2}^{1} dx \int_0^{\frac{1}{4x}} dy \ = \ 2 \cdot \frac{1}{4} \cdot [\log x]_{1/2}^{1} = \frac{1}{2}(-\log \frac{1}{2})$$

□

4.17 - Un'apparecchiatura ha un tempo di vita Y che segue una legge esponenziale di parametro x che dipende dalla qualità di uno dei materiali impiegati. Ma nel processo di produzione x non è sotto controllo, e si assume distribuito come una v.a. esponenziale di parametro 2. Qual è la densità congiunta di (X, Y)?

a] $f(x,y) = 2x \cdot e^{-x(2+y)}, \quad x, y > 0$

Risoluzione

$$\bar{f}_{Y|X}(y|\ X = x) = \begin{cases} xe^{-xy}, \ y > 0 \\ 0, \quad y \leq 0 \end{cases} \qquad \Longrightarrow$$

$$f(x,y) = f_X(x) \cdot \bar{f}_{Y|X}(y|\ X = x) = \begin{cases} 2e^{-2x}xe^{-xy}, \ x > 0, y > 0 \\ 0, \quad \text{altrove} \end{cases}$$

da cui segue il risultato. □

4.18 - La densità congiunta di (X, Y) è

$$f(x,y) = \begin{cases} 3x \cdot e^{-x(3+y)}, \quad x, y > 0 \\ 0 \quad \text{altrove} \end{cases}$$

Qual è la densità condizionata di X dato $Y = y$?

a] $f_{X|Y}(x|Y = y) = x(3+y)^2 e^{-x(3+y)}$

4.19 - In una fabbrica ci sono due linee di lavorazione. I pezzi prodotti dalla linea A hanno un tempo di vita esponenziale di parametro $\lambda = 1/5$; i pezzi prodotti dalla linea B hanno il tempo di vita esponenziale di parametro $\mu = 1/3$. Inoltre le linee A e B producono rispettivamente il 70% e il 30% dei pezzi. Un pezzo è scelto a caso e indichiamo con T il suo tempo di vita. Calcolare il valor medio $E(T)$.

a] 22/5

Risoluzione Una v.a. esponenziale di parametro λ ha densità uguale a $\lambda e^{-\lambda x}$ per $x \in (0, \infty)$ e uguale a 0 altrove. Per il teorema di probabilità totale applicato all'evento $(T \leq t)$

$$\begin{aligned} F_T(t) &= P(T \leq t) = P(T \leq t|A) \cdot P(A) + P(T \leq t|B) \cdot P(B) \\ &= (\int_0^\infty \frac{1}{5} e^{-x/5}\ dx) \cdot \frac{7}{10} \ + \ (\int_0^t \frac{1}{3} e^{-x/3}\ dx) \cdot \frac{3}{10} = \\ &= \frac{7}{10}(1 - e^{-t/5}) + \frac{3}{10}(1 - e^{-t/3}). \end{aligned}$$

$$\Longrightarrow f(t) = F'(t) = \begin{cases} \frac{7}{10}\frac{1}{5}e^{-t/5} + \frac{3}{10}\frac{1}{3}e^{-t/3}, \quad t > 0 \\ 0, \quad \text{altrove.} \end{cases}$$

Ricordando che la media di una v.a. esponenziale di parametro λ è $\frac{1}{\lambda}$,

$$E(T) = \frac{7}{10} \int_0^\infty \frac{1}{5} t e^{-t/5} dt + \frac{3}{10} \int_0^\infty \frac{1}{3} t e^{-t/3} dt = \frac{7}{10} \cdot 5 + \frac{3}{10} \cdot 3 = \frac{22}{5}$$

$\square$

4.20 - Ogni anno un tipo di macchina deve essere sottoposto ad alcuni arresti per manutenzione. Questo numero di arresti X è v.a. di Poisson con parametro y. Ma anche y è aleatorio (ad es. può dipendere dalla macchina) e assumiamo che esso segua una legge

$$f_Y(y) = \ 3e^{-3y}, \quad y > 0; \qquad f_Y(y) = 0, \quad y < 0$$

Qual è la probabilità che una singola macchina sia sottoposta a 3 arresti in un anno? Qual è la miglior stima di y sapendo che ci sono stati k arresti?

a] $\simeq 1.17\%$, $E(Y|X = k) = (k + 1)/4$.

4.21 - La densità di probabilità congiunta delle v.a. X ed Y è

$$f_{X,Y}(x, y) = \begin{cases} 8xy, & 0 \le x \le 1, \ 0 \le y \le x \\ 0, & \text{altrove.} \end{cases}$$

Determinare la media condizionata di Y sapendo che $X = \frac{1}{2}$.

a] $\simeq 0.33$

4.21 bis - Le variabili aleatorie reali X, Y hanno densità di probabilità congiunta:

$$f_{X,Y} = \begin{cases} 4 \cdot e^{-2y}, & \text{se } 0 < x < y \\ 0, & \text{altrove.} \end{cases}$$

a) Trovare la densità di X. b) Trovare il valor medio condizionato $E(Y|X = x)$ per $x > 0$.

a] $f_X(x) = 2e^{-2x} \cdot I_{R^+}(x)$

b] $E(Y|X = x) = 1$.

4.22 - Si collaudano lampadine lasciandole accese fino al guasto. Il tempo di guasto T abbia la seguente densità:

$$f_T(t) = \frac{1}{5}e^{-t/5}, \quad t > 0; \quad = 0, \ t < 0.$$

Si decide però di collaudare solo le lampadine sopravvissute dopo un rodaggio di tempo $t_\circ = 1$. Qual è la media condizionata di T dato l'evento $B = \{T > 1\}$?

a] 6

4.23 - Due giovani decidono di incontrarsi tra le 17 e le 18 con l'accordo che nessuno deve aspettare l'altro per più di 5 minuti. Supponiamo che gli orari X ed Y in cui arrivano siano indipendenti e casuali, variabili fra le 17 e le 18. Trovare la probabilità condizionata che i due giovani si incontrino: (a) incondizionata; (b) condizionata, sapendo che lei arriva alle 17 : 57.

4.24 - La legge condizionata di una v.a. X è esponenziale di parametro λ, ma λ è un valore di una v.a. Λ uniformemente distribuita in $[0,1]$. Determinare la densità condizionata

$$f_{\Lambda|X}(\lambda|x) \quad \text{per} \quad \lambda = \frac{1}{2},\ x = 1$$

a] 1.14

COVARIANZA, VETTORI GAUSSIANI

4.26 - Siano Y_1, Y_2 normali indipendenti a media nulla con varianze 2 e 7 rispettivamente. Siano

$$\begin{cases} X_1 = (\sqrt{3}/2)Y_1 - (1/2)Y_2 \\ X_2 = (1/2)Y_1 + (\sqrt{3}/2)Y_2 \end{cases}$$

Si domanda la matrice di covarianza della coppia di v.a. (X_1, X_2). Tracciare una curva di livello della densità congiunta di X_1, X_2.

4.27 - Se X_1, X_2 sono v.a. normali di medie zero e matrice di covarianza

$$C_{X_1,X_2} = \begin{pmatrix} \frac{15}{4} & \frac{3\sqrt{3}}{4} \\ \frac{3\sqrt{3}}{4} & 6 \end{pmatrix}$$

si domandano: (1) le varianze σ_1^2, σ_2^2 delle v.a. normali e indipendenti Y_1, Y_2 tali che

$$\begin{cases} X_1 = \cos\alpha Y_1 - \sin\alpha Y_2 \\ X_2 = \sin\alpha Y_1 + \cos\alpha Y_2 \end{cases}$$

2) l'angolo α, in radianti, che individua tale rotazione degli assi.

R.: 6; 3 $\alpha = \pi/3$