NoiStudiamo

APPUNTI

# Metodi matematici e statistici

Probabilità, variabili aleatorie, distribuzioni, statistica inferenziale, test d'ipotesi, metodi Monte Carlo e catene di Markov

A CURA DI

**Paolo Volpini**

Novembre 2026

INDICE

# Indice degli argomenti

Gli argomenti principali degli appunti e la pagina in cui iniziano.


1. Introduzione ..... 4
2. Eventi condizionati ..... 8
3. Variabili aleatorie ..... 14
4. Distribuzioni note ..... 20
5. Altre distribuzioni note ..... 28
6. Variabili aleatorie continue ..... 32
7. Distribuzioni continue ..... 36
8. Altre distribuzioni continue ..... 40
9. Introduzione alla statistica ..... 44
10. Legge dei grandi numeri, teorema del limite centrale ..... 48
11. Statistica ..... 54
12. Rappresentazione dei dati ..... 58
13. Statistica inferenziale ..... 60
14. Test d’ipotesi ..... 66
15. Generazione di numeri pseudocasuali ..... 74
16. Metodi Monte Carlo ..... 80
17. Catene di Markov ..... 82

# Metodi matematici e statistici

Paolo Volpini

# 00 - Introduzione

## Eventi e probabilità

### Evento

Un evento è una **proposizione** che può essere vera o falsa.

Ad esempio, un evento può essere “lancio un dado ed esce 5”.

Tra gli eventi distinguiamo:

- l’evento certo $\Omega$, sempre vero;
- l’evento impossibile , sempre falso;
- l’evento contrario $E^C$ di $E$, vero quando $E$ è falso e falso quando $E$ è vero;

Due eventi $E_1, E_2$ si dicono **incompatibili** se $E_1$ è vero quando $E_2$ è falso e viceversa. Due eventi sono **uguali** se $E_1$ vero allora $E_2$ vero e se $E_1$ falso allora $E_2$ falso. Il viceversa è altrettanto vero.

### Operazioni tra gli eventi

#### Implicazione

Diciamo che $E_1$ **implica** $E_2$ se $E_1$ vero comporta $E_2$ vero e scriveremo $E_1 \subseteq E_2$.
Se due eventi $E_1, E_2$ sono uguali, allora $E_1 \subseteq E_2$ e $E_2 \subseteq E_1$ e scriveremo $E_1 = E_2$.

#### Somma logica

Siano $E_1, E_2$ eventi. La **somma logica** tra $E_1$ ed $E_2$ è l’evento $E$ tale che $E$ è vero se almeno uno tra $E_1, E_2$ è vero. Scriveremo $E = E_1 \vee E_2$



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

### Prodotto logico o intersezione

Siano $E_1, E_2$ eventi. Il **prodotto logico** tra $E_1$ ed $E_2$ è l'evento $E = E_1 \wedge E_2$ tale che $E$ è vero se e solo se $E_1$ ed $E_2$ sono entrambi veri.

## Probabilità

La probabilità è una **misura di sicurezza** che un certo evento accada. Non vi è tuttavia un'unica definizione di probabilità.

### Definizione classica

La definizione classica di probabilità definisce la probabilità di un evento come il rapporto $P(E) = \frac{k}{n}$, dove $n$ è il numero di eventi equiprobabili totali mentre $k$ è il numero eventi favorevoli al verificarsi di $E$.

### Definizione frequentista

La **definizione frequentista** della probabilità invece consiste nel ripetere un esperimento che permette di verificare $E$, lo si ripete $n$ volte e si osservano le $k$ volte in cui $E$ si verifica. Allora $P(E) = \frac{k}{n}$.

### Definizione soggettiva

Si introduce una quota $p$ e una somma $S$. Si effettua una scommessa di quota $p$ su un evento $E$ se versando un importo pari ad $p \cdot S$ si vince $S$ in caso avvenga $E$. Si introduce allora un guadagno $G$ tale che:

$$G = \begin{cases} S - pS & \text{se si verifica E} \\ -Sp & \text{se non si verifica E} \end{cases}$$

Una scommessa si dice coerente se $G_1 \cdot G_2 < 0$, che implica che $p$ è compreso tra zero ed uno.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

### Definizione assiomatica

Sia $\Omega$ l'insieme degli esiti. Sia $\mathcal{P}(\Omega)$ l'insieme delle parti di $\Omega$. Sia $\mathcal{A}$ la famiglia di sottoinsiemi di $\Omega$, sottoinsieme dell'insieme delle parti $\mathcal{P}(\Omega)$ e **contenente i sottoinsiemi che costituiscono gli eventi**. Sia $P$ una **misura di fiduciosità** dell'evento appartenente alla famiglia $\mathcal{A}$. $P$ allora è una funzione del tipo:

$$P : \mathcal{A} \to [0, 1]$$

Il modo in cui viene assegnata la probabilità dipende caso per caso. La funzione ha le seguenti proprietà (affinché sia coerente):

- $P(\Omega) = 1$;
- dati $A_i \in \mathcal{A}, A_i \wedge A_j = 0 \; \forall i, j \in \{1, ..., n\}$ dove $|\mathcal{A}| = n$, allora $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$.

$(\Omega, \mathcal{A}, P)$ è noto come **spazio di probabilità**. In particolare, $\mathcal{A}$ è un $\sigma-$algebra di $\Omega$, in quanto gli eventi godono di tali proprietà:

- , $\Omega \in \mathcal{A}$. Questo significa che l'evento certo e impossibile sono eventi;
- se $A \in \mathcal{A}$, allora $A^C \in \mathcal{A}$. Questo indica che se esiste un evento, esiste anche il suo contrario;
- dati $n$ eventi $A_1 ... A_n$, allora $\bigcup_{i=1}^{n} A_i \in A$. Questo indica che l'unione di eventi è ancora un evento.

Possiamo allora dedurre tali proprietà:

- $P(A^c) = 1 - P(A)$;
- $P(A) \leq P(B)$ se $A \subseteq B \; \forall A, B \in \mathcal{A}$;
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{i_1, i_2 = 1}^{n} P(E_{i_1} \cap E_{i_2}) + \sum_{i_1, i_2, i_3}^{n} P(E_{i_1} \cap E_{i_2} \cap E_{i_3}) - ...$



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

# 01 - Eventi condizionati

Un evento condizionato è un evento che esiste in condizione di un altro. Più precisamente, un evento $E$ condizionato da un evento $H \neq$ e si scrive con $E|H$ e si legge "$E$ dato $H$" è un ente di tre valori:

- $E|H$ è vero se $E$ è vero ed $H$ è vero;
- $E|H$ è falso se $E$ è falso ed $H$ è vero;
- $E|H$ è **indeterminato** se $H$ è falso.

La probabilità di tale evento è dato dal calcolo:

$$P(E|H) = \frac{P(E \wedge H)}{P(H)}$$

### Eventi indipendenti

Due eventi si dicono **indipendenti** se vale la seguente uguaglianza:

$$P(E \wedge H) = P(E) \cdot P(H)$$

In tal caso possiamo allora affermare:

$$P(E|H) = P(E)$$

### Teorema delle probabilità totali

Sia $(\Omega, \mathcal{A}, P)$ uno spazio di probabilità. Consideriamo $n$ eventi $A_1, ..., A_n \in \mathcal{A}$ disgiunti, ossia tali che $A_i \cap A_j = 0 \ \forall i \neq j$, ed **esaustivi**, ossia tali che $\bigcup_{i=1}^{n} A_i = \Omega$. Allora per un qualunque evento $B \in \mathcal{A}$ si ha che:

$$P(B) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)$$

**Esempio**

> *Una stazione radio riceve segnali da due sorgenti $A$ e $B$. La probabilità che il segnale di $A$ sia distorto è $0.1$. La probabilità che invece il segnale di $B$ sia distorto è $0.2$. Qual è la probabilità che il segnale sia distorto?*

Il problema è banale. Basta applicare il teorema delle probabilità totali:

$$P(S) = P(S|S_A) \cdot P(S_A) + P(S|S_B) \cdot P(S_B) = 0.05 + 0.1 = 0.15$$

**Dimostrazione del teorema**

Il teorema è conseguenza della definizione della probabilità condizionata. Infatti l'evento $P(B)$ può essere visto come:

$$P(B) = P(B \cap \Omega)$$

Usando la definizione di probabilità condizionata:

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

E poiché tutti gli eventi $A_i$ sono esclusivi, allora si ha:

$$P(B) = P(B \cap \Omega) = P(B \cap \bigcup_{i=1}^{n} A_i) = P(\bigcup_{i=1}^{n} B \cap A_i)$$

Dato che $\mathcal{A}$ è una $\sigma-$algebra possiamo applicare la proprietà dell'unione come somma, ma poiché gli eventi sono disgiunti, ogni intersezione è nulla. Per cui si giunge alla tesi.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

### Teorema di Bayes

Siano $A, B \in \mathcal{A} | P(A) > 0, P(B) > 0$. Allora:

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

#### Dimostrazione

Si ha banalmente che:

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

Poiché $P(B \cap A) = P(A \cap B)$, allora possiamo dire che:

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

da cui basta dividere entrambi i membri per $P(A)$.

## Cenni di calcolo combinatorio

Ricordiamo alcune definizioni di calcolo combinatorio.

### Disposizioni

Si definiscono **disposizioni di $n$ elementi di classe $k$** i gruppi ordinati di $k$ elementi a partire da un insieme di $n$ elementi:

$$D_{n,k} = \frac{n!}{(n-k)!}$$

Se ogni elemento è ripetibile, allora parliamo di **disposizioni con ripetizione**:

$$D^r_{n,k} = n^k$$



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

## Combinazioni

Si definiscono **combinazioni di $n$ elementi di classe $k$** i gruppi **non** ordinati di $k$ elementi a partire da un insieme di $n$ elementi, oppure il numero di sottoinsiemi di $k$ elementi a partire da un insieme di $n$ elementi:

$$C_{n,k} = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Si distinguono dalle **combinazioni con ripetizione**:

$$C^r_{n,k} = \binom{n+k-1}{k}$$

### Esempio: paradosso del compleanno

Ci sono $r$ persone. Qual è la probabilità che due persone facciano il compleanno lo stesso giorno?

Il modo più rapido di rispondere è calcolando la probabilità dell'evento contrario, ossia l'evento "nessuno fa il compleanno lo stesso giorno". Allora la probabilità dell'evento è data da:

$$P(A) = 1 - \frac{D_{365,r}}{365^r}$$

Infatti possiamo dire semplicemente che la prima persona può essere scelta per un giorno qualsiasi, la seconda persona per tutti i giorni tranne quello precedente, e così via fino ad arrivare alla $r-$esima persona.

## Paradosso di Monty-Hall

Ci sono tre porte chiuse. Dietro una di queste è presente un premio, mentre dietro le altre due sono presenti le capre. Scelta una porta, il presentatore, **che sa dove sta il premio**, apre una porta e mostra che dietro questa non vi è il premio. Allora **conviene cambiare porta?**

La risposta in questo caso è affermativa. Sia $A_i$ l'evento "il premio è dietro la porta $i$" e sia $C_j$ l'evento "il conduttore apre la porta $j$". Per il teorema di Bayes si ha che:

$$P(A_i|C_j) = \frac{P(C_j|A_i) \cdot P(A_i)}{P(C_j)} = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

Chiaramente la porta $j$ è diversa dalla porta che ha scelto il giocatore. Inoltre, se il conduttore scegliesse a caso una porta **senza però sapere dove sta il premio**, allora c'è la possibilità di aprire la porta col premio dietro. Pertanto si ha che:

$$P(A_i|C_j) = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{3}} = \frac{1}{2}$$

### Osservazioni

La probabilità di eventi condizionati ci permette di capire quanto aumenta la probabilità che un evento accada. Ad esempio, supponiamo di scansionare borse per vedere la presenza di bombe. Se una bomba ha fatto scattare l'allarme, quanto è aumentata la probabilità che ci sia una bomba?

Sia $A$ l'evento "la borsa ha una bomba" e $B$ l'evento "l'allarme è scattato". Allora l'aumento di probabilità è dato dal rapporto:

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

Se tale rapporto è 1, allora gli eventi sono indipendenti.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

# 02 - Variabili aleatorie

Le variabili aleatorie generalizzano il concetto di evento.
Sia $(\Omega, \mathcal{A}, P)$ uno spazio di probabilità. Si dice **variabile aleatoria** un'applicazione:

$$X : \Omega \to \mathbb{R} \mid \forall t \in \mathbb{R}, \{\omega \in \Omega : X(\omega) \leq t\} \in \mathcal{A}$$

L'insieme mostrato è l'insieme di tutti gli esiti tali che $X(\omega) \leq t$ è un evento. Inoltre, poiché $\mathcal{A}$ è una $\sigma-$algebra, si ha che anche l'evento contrario appartiene ad $\mathcal{A}$, ossia:

$$\{\omega \in \Omega : X(\omega) > t\} \in \mathcal{A}$$

Per abbreviare la definizione dell'insieme, scriveremo $\{X \leq t\}$, sebbene sia matematicamente priva di senso. Brevemente, una variabile aleatoria non è altro che una funzione che associa ad un evento un valore reale.

Sia dato ora $X(\Omega)$ immagine della funzione $X$. Se $X(\Omega)$ è finito o numerabile, la variabile aleatoria $X$ si dice **discreta**. Se $X(\Omega)$ è continuo, la variabile aleatoria $X$ si dice **continua**.

## Legge o distribuzione di una variabile aleatoria

Sia $A \subseteq \mathbb{R}$. Il valore $P(\{\omega \in \Omega : X(\omega) \in A\})$ si dice **legge** o **distribuzione** di $X$.

La distribuzione dipende dal fenomeno che osserviamo. Non tutte le distribuzioni sono uguali.

Ad esempio, considerando un lancio di un dado, $X(\Omega) = \{1, 2, 3, 4, 5, 6\}$, $A$ può essere l'insieme $\{1\}$, $\{1, 2\}$. Possiamo considerare allora i due esiti $\omega_1, \omega_2$ rispettivamente "esce il numero 1", "esce il numero 2". Allora $X(1), X(2) \in A$ e avremo che la legge ci esprime la probabilità che $X$ assuma i valori di $A$; in questo caso la probabilità è $\frac{1}{3}$.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

# Variabili aleatorie discrete

## Densità discreta

Sia $X$ una variabile aleatoria discreta. Vogliamo calcolarne la legge.
Consideriamo $A \subseteq \mathbb{R}$. Allora:

$$P(X \in A) = P(X \in \bigcup_{a \in A} \{a\}) = \sum_{a \in A} P(X = a)$$

Da notare che tale somma è ben definita per il fatto che $X$ è una variabile aleatoria discreta, per cui $X(\Omega)$ è numerabile. Definiamo allora **densità discreta** la funzione:

$$p(x) = P(X = x)$$

Essenzialmente tale funzione associa ad ogni valore $x$ assunto dalla variabile discreta $X$ la probabilità che la variabile assuma esattamente quel valore.
Affinché $p$ sia una densità di probabilità, devono valere due proprietà:

- $p(x) \geq 0 \; \forall x \in \mathbb{R}$
- $\sum_{x \in \mathbb{R} \mid p(x) \neq 0} p(x) = 1$

I valori reali $x \in \mathbb{R} \mid p(x) \neq 0$ sono numerabili come visto precedentemente.

## Funzione di ripartizione

Definiamo **funzione di ripartizione** di $X$ la funzione:

$$F_X : \mathbb{R} \to [0, 1]$$

La funzione è tale che:

$$F_X(t) = P(X \leq t) = \sum_{x \leq t} p(x)$$



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

Considerando l'esempio del lancio del dado, la funzione $F_X(t)$ è pari a $0$ quando $t \leq 1$, $\frac{1}{6}$ quando $1 \leq t \leq 2$, $\frac{2}{6}$ quando $1 \leq t \leq 3$ e così via.

### Speranza matematica, varianza

Consideriamo ora $X$. Se esiste la somma finita:

$$\sum_{i=1}^{\infty} |x_i| \cdot p(x_i) < +\infty$$

allora si definisce **speranza matematica** di $X$ e si scrive:

$$E[X] = \sum_{i=1}^{\infty} x_i \cdot p(x_i)$$

Ricordiamo sempre che la somma finita esiste per il fatto che i valori reali $x \in \mathbb{R} \mid p(x) \neq 0$ sono numerabili.
La speranza matematica gode di alcune proprietà:

1. se $c \in \mathbb{R} \implies E[cX] = cE[X]$;
2. se $X$ e $Y$ sono due variabili aleatorie discrete, allora $E[X+Y] = E[X] + E[Y]$;
3. se $X$ e $Y$ sono variabili aleatorie tali che $p(X \leq Y) = 1 \implies E[X] \leq E[Y]$;
4. $|E[X]| \leq E[|X|]$;
5. siano $X$ e $Y$ due variabili aleatorie indipendenti. Allora $E[X \cdot Y] = E[X] \cdot E[Y]$.

Le variabili aleatorie devono essere definite sullo stesso spazio di probabilità.

Supponiamo adesso di avere una generica funzione $f : \mathbb{R} \to \mathbb{R}$ tale che $f(X)$ è una variabile aleatoria (ovvero esiste la legge di $f(X)$). Allora se esiste finita, definiamo **speranza matematica** di $f(X)$ la legge:

$$E[f(X)] = \sum_{i=1}^{\infty} f(x_i) \cdot p(x_i)$$

Ne segue che $\forall k \in \mathbb{N}$ possiamo calcolare $E[X^k]$ sempre se esiste finito. Non è detto che se esiste finita la speranza matematica esiste finita anche la somma per $X^k$. Tuttavia per $k = 1$ esiste la speranza matematica, e generalmente per $k$ si ha:



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

$$E[X^k] = \sum_{i=1}^{\infty} x_i^k \cdot p(x_i)$$

Questa quantità prende il nome di **momento di ordine** $k$.

Il **momento centrato di ordine** $k$ è invece la quantità:

$$E[(X - E[X])^k] = \sum_{i=1}^{\infty} (x_i - E[X])^k \cdot p(x_i)$$

Il valore centrale ricorda la **varianza di** $X$, definita come:

$$Var(X) = E[(X - E[X])^2]$$

La varianza esprime la distanza dal valore medio. La differenza tra il momento di ordine $k$ e il momento centrato di ordine $k$ risiede nel fatto che il momento di ordine $k$ esprime la lontananza dei dati rispetto allo zero.

La varianza è anche nota come $\sigma^2$, dove $\sigma$ è la **deviazione standard**, e gode di alcune proprietà:

1. $\sigma^2 = E[X^2] - (E[X])^2$. Tale proprietà vale **solo** per questo ordine;
2. $a \in \mathbb{R} \implies Var(a \cdot X) = a^2 Var(X)$;
3. $a \in \mathbb{R} \implies Var(a + X) = Var(x)$. Questo perché spostiamo i dati sull'asse $x$, ma visto che la varianza è centrata, allora cambia solo la media;

### Covarianza

Siano $X$ e $Y$ due variabili aleatorie sullo stesso spazio di probabilità. Definiamo la **covarianza** di $X$ e $Y$ come:

$$Cov(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

Se $X$ e $Y$ sono indipendenti, allroa si ha che $E[XY] = E[X]E[Y]$ e la covarianza è nulla.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

**Varianza e Covarianza**

Definite $X$ e $Y$ variabili aleatorie sullo stesso spazio di probabilità, vale la seguente proprietà:

$$Var(X+Y) = Var(X) + Var(Y) + 2 \cdot Cov(X,Y)$$

Se due variabili aleatorie sono indipendenti, la varianza della somma è la somma delle varianze delle singole variabili.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

# 03 - Distribuzioni note

## Distribuzione discreta uniforme

### Densità di probabilità

La densità di probabilità di questa distribuzione è molto semplice. Sia $E$ insieme di $n$ esiti **equiprobabili**. Sia $X : E \to \mathbb{R}$ variabile aleatoria. Sia $X(E) = \{1, 2, ..., n\}$ l'immagine di $E$ attraverso $X$. La densità di probabilità è:

$$p(x) = \begin{cases} \frac{1}{n} & \text{se } x \in X(E) \\ 0 & \text{altrimenti} \end{cases}$$

Ci chiediamo se questa è una densità. Dobbiamo provare due proposizioni:

- $p(x) \geq 0 \quad \forall x \in \mathbb{R}$? Sì, perché $p(x) = \frac{1}{n} > 0$ per definizione, e zero altrimenti;
- $\sum_x p(x) = 1$? Sì, perché sommiamo $n$ volte $\frac{1}{n}$. Formalmente si ha che $\sum_x \frac{1}{n} = \frac{1}{n} \sum_{i=1}^{n} 1 = 1$.

### Funzione di ripartizione

La funzione di ripartizione di questa distribuzione è:

$$F_X(t) = \begin{cases} 0 & \text{se } t < 1 \\ \frac{1}{n} & \text{se } 1 \leq t < 2 \\ ... & \\ 1 & \text{se } t \geq n \end{cases}$$

La funzione di ripartizione aumenta a gradi.

### Speranza matematica

La speranza matematica è:

$$E[X] = \sum_x x \cdot p(x) = \sum_{i=1}^{n} i \cdot p(i) = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}$$

Maggiore è il numero di esiti, maggiore è la speranza matematica.

### Varianza

La varianza è:

$$Var(X) = E[X^2] - (E[X])^2$$

Calcoliamo $E[X^2]$:

$$E[X^2] = \sum_{i=1}^{n} i^2 \cdot p(i) = \frac{1}{n} \sum_{i=1}^{n} i^2 = \frac{1}{n} \cdot \frac{n \cdot (n+1) \cdot (2n+1)}{6} = \frac{(n+1) \cdot (2n+1)}{6}$$

Allora la varianza è:

$$Var(X) = \frac{n^2 - 1}{12}$$

Osserviamo che ciò dipende da $n$ in maniera quadratica. Inoltre, se $n$ tende ad infinito, sia media che varianza divergono.

## Distribuzione binomiale

La distribuzione binomiale è una distribuzione di probabilità che descrive un fenomeno di tipo **successo-insuccesso** in cui si ripete l'esperimento $n$ volte e ogni prova è **indipendente** dalla precedente. Ci sono solamente due esiti. Ad esempio, il lancio di una moneta è uno schema successo-insuccesso (testa = successo, croce = insuccesso). Oppure un caso in cui si ha una certa probabilità che un evento abbia successo o meno.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

Siano $n$ il numero di prove, $k$ il numero di successi su $n$ prove, $p$ la probabilità di successo, $q = 1 - p$ la probabilità di insuccesso. La distribuzione dipende solo da $n$ e $p$, ed è cruciale che gli eventi siano indipendenti.

### Densità di probabilità

La variabile aleatoria $X \sim B(n, p)$ (ossia, "$X$ è una variabile aleatoria che segue una distribuzione binomiale di parametri $n, p$") è la variabile aleatoria che conta i successi avvenuti in $n$ prove. La scrittura $P(X = k)$ indica la probabilità di avere $k$ successi.

Trovare la legge di $X$ binomiale significa calcolare $n + 1$ probabilità, in quanto $k \in \{0, ..., n\}$. Consideriamo pertanto alcuni eventi ausiliari:

- l'evento $A_i$ indica il successo alla $i-$esima prova;
- l'evento $A$ indica $k$ successi solo nelle prime $k$ prove.

Pertanto:

$$P(A) = P(A_1 \cap A_2 \cap ... \cap A_k \cap \overline{A}_{k+1} \cap ... \cap \overline{A}_n)$$

Poiché gli eventi sono indipendenti, possiamo dire:

$$P(A) = P(A_1) \cdot P(A_2) \cdot ... \cdot P(\overline{A}_n) = p^k \cdot q^{n-k}$$

Notiamo che anche altre configurazioni di $k$ successi in $n$ prove hanno la stessa probabilità. Queste configurazioni sono di numero $\binom{n}{k}$.

Pertanto la probabilità di avere $k$ successi in $n$ esperimenti è:

$$P(X = k) = \binom{n}{k} p^k \cdot (1 - p)^{n-k}$$

Possiamo allora derivare la legge di $x$, che sarà:

$$p(x) = \begin{cases} \binom{n}{x} \cdot p^x \cdot (1 - p)^{n-x} & \text{se } x \in \{0, 1, ..., n\} \\ 0 & \text{altrimenti} \end{cases}$$

Da notare che questa conta è **già normalizzata**, ossia si ha che:



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

$$\sum_x p(x) = 1$$

Infatti si ha che banalmente per il teorema binomiale di Newton la somma corrisponde al binomio $(p + 1 - p)^n$

#### Esempi

Supponiamo di avere 3 file da trasccrivere. Durante la trascrizione si sono verificati 5 errori distribuiti in maniera random. Dato uno specifico file, calcolare la probabilità che ci sia un errore.

Possiamo vedere la probabilità che ci sia un errore come un evento successo-insuccesso dove la probabilità è $p = \frac{1}{3}$. Avendo $n = 5$, si ha una distribuzione binomiale. La legge della distribuzione binomiale ci permette di dire che:

$$P(X = 1) = \binom{5}{1} \cdot \frac{1}{3} \cdot (\frac{2}{3})^4$$

Supponiamo adesso di avere un segnale a 5 bit. La probabilità di ricevere un bit distorto è $0.1$. Calcolare la probabilità di ricevere un segnale distorto.

Sia $X \sim B(5, 0.1)$ la variabile aleatoria che descrive l'evento "bit distorto". Dato $S$ l'evento "segnale distorto", avrò:

$$P(S) = 1 - P(X = 0) = 1 - 0.9^5$$

### Speranza matematica

Per calcolare la speranza matematica, introduciamo la legge di Bernoulli.

Definiamo **legge di Bernoulli** una binomiale di parametri $n = 1, p$. Se $X_i \sim B(1, p)$, allora:

$$X_i = \begin{cases} 1 & \text{se ottengo successo} \\ 0 & \text{altrimenti} \end{cases}$$

In questo caso la speranza matematica è molto semplice:



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

$$E[X_i] = 1 \cdot p + 0 \cdot (1-p) = p$$

Sfruttiamo il fatto che:

$$X = \sum_{i=1}^{n} X_i \sim B(n,p)$$

Per cui possiamo ottenere:

$$E[X] = \sum_{i=1}^{n} E[X_i] = n \cdot p$$

### Varianza

Per la varianza, consideriamo la legge di Bernoulli:

$$Var(X_i) = E[X_i^2] - (E[X])^2$$

e notiamo che:

$$Var(X_i) = 1^2 \cdot p = p$$

per cui:

$$Var(X_i) = p - p^2 = p \cdot (1-p)$$

Allora:

$$Var(X) = Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) = n \cdot p \cdot (1-p)$$

Questo risultato ci indica che se $p = 1$ o $p = 0$, allora non si hanno dispersione di eventi, in quanto la media è uguale all'evento stesso.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

## Distribuzione di Poisson

Si tratta di un caso particolare della distribuzione binomiale. Descrive uno schema successo-insuccesso quando abbiamo tante prove ma poca probabilità di successo.

### Densità di probabilità

Sia $X$ la variabile aleatoria che descrive il numero di successi quando $n >> 1$ e $p << 1$. Pertanto $X \sim Pois(\lambda)$, dove $\lambda = np$. Al tendere di $n$ a infinito e $p$ a zero, possiamo dire che $B(n,p)$ tende ad una distribuzione di Poisson $Pois(\lambda)$.
Se $X \sim Pois(\lambda)$, allora la densità di probabilità è:

$$p(x) = \begin{cases} e^{-\lambda} \cdot \frac{\lambda^x}{x!} & \text{se } x = 0, 1, ..., \infty \\ 0 & \text{altrimenti} \end{cases}$$

Ci chiediamo se $p(x)$ è una densità. Notiamo chiaramente che $p(x) \geq 0$. Calcoliamo:

$$\sum_{k \in \mathbb{R} | p(k) > 0} p(x) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1$$

### Speranza matematica

Infine cerchiamo ora di capire quanto vale la speranza matematica. Introduciamo una variabile $Y \sim B(n,p)$ e poniamo:

$$E[X] = \lim_{n \to \infty, p \to 0, n \cdot p \; costante} E[Y] = \lambda$$

Questo perché $E[Y] = n \cdot p$. Questo si può anche dimostrare calcolando:

$$\sum_{x \in \mathbb{R}} x \cdot p(x)$$



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

### Varianza

Adesso calcoliamo la varianza. Abbiamo che:

$$Var(Y) = n \cdot p \cdot (1 - p)$$

Ma sapendo che $p$ tende a 0, $n$ tende a infinito e che il loro prodotto è $\lambda$, si ottiene che anche questo prodotto è $\lambda$. Per cui:

$$Var(X) = \lambda$$

### Proprietà della somma

La distribuzione di Poisson gode di una proprietà importante. Se $X \sim Pois(\lambda)$ e $Y \sim Pois(\mu)$ sono variabili **indipendenti**, allora la somma:

$$X + Y \sim Pois(\lambda + \mu)$$



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

# 04 - Altre distribuzioni note

## Distribuzione multinomiale

La distribuzione multinomiale esprime uno schema successo-insuccesso con più di due esiti. Abbiamo $k$ possibili esiti con probabilità $q_i$.
$Y$ seguirà una distribuzione multinomiale di parametri $n, q_1, q_2, ..., q_k$ e si scriverà $Y \sim B(n, q_1, q_2, ..., q_k)$, con $\sum_{i=1}^{k} q_i = 1$. Più precisamente, $Y$ conta quante volte è uscito ciascun esito in $n$ tentativi. $Y$ è un vettore quindi di $k$ variabili aleatorie $Y_i$ che contano quante volte si è verificato l'esito $i-$esimo. L'evento generico $\omega$ è allora un vettore $w = (w_1, ..., w_k)$ che indica che abbiamo svolto $n$ volte l'esperimento e abbiamo osservato $w_i$ volte l'$i-$esimo risultato. Si ha allora che $\sum_{i=1}^{k} \omega_i = n$. Allora:

$$P(Y = \omega) = \frac{n!}{\omega_1! \cdot \omega_2! \cdot ... \cdot \omega_k!} \cdot q_1^{\omega_1} \cdot q_2^{\omega_2} \cdot ... \cdot q_k^{\omega_k}$$

## Distribuzione geometrica

La distribuzione geometrica è caratterizzata dal solo parametro $p$. $X \sim Geom(p)$ se ha densità di probabilità:

$$p(x) = \begin{cases} p(1-p)^{x-1} & \text{se } x = 1, 2, ..., \infty \\ 0 & \text{altrimenti} \end{cases}$$

Notiamo che $x$ dev'essere pari ad almeno $1$. Verifichiamo che è una densità di probabilità. Chiaramente $p(x) \geq 0 \quad \forall x$. Ora verifichiamo:

$$\sum_x p(x) = \sum_x p(1-p)^{x-1} = p \cdot \sum_{n=1}^{\infty} (1-p)^{n-1} = \ = p \cdot \sum_{m=0}^{\infty} (1-p)^m = p \cdot \frac{1}{1-(1-p)} = p \cdot \frac{1}{p} = 1$$



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

La distribuzione geometrica analizza la probabilità di ottenere il **primo successo** dopo $x$ tentativi. Il numero di lanci sarà aleatorio. Una variante invece considera $x$ come il numero di fallimenti che precedono il primo successo.

La distribuzione geometrica è strettamente legata alla binomiale. Consideriamo $T$ variabile aleatoria che vuole tenere conto quante volte è necessario ripetere un esperimento binomiale prima di ottenere un successo. Dobbiamo allora calcolare $P(T = n)$. Questa può essere calcolata come:

$$P(T = n) = P(T > n - 1) - P(T > n)$$

Valutare questa probabilità è più semplice. Introduciamo una variabile aleatoria $X^{(n-1)} \sim B(n, p)$, che conta il numero di successi durante $n - 1$ lanci. Osserviamo:

$$P(T > n - 1) = P(X^{(n-1)} = 0)$$

Ossia che $X_1 = 0, X_2 = 0, ..., X_{n-1} = 0$. Per calcolare $P(X^{(n-1)} = 0)$, usiamo la legge binomiale:

$$P(X^{(n-1)} = 0) = \binom{n-1}{0} \cdot p^0 \cdot (1 - p)^{n-1}$$

Analogamente possiamo dire che $P(T > n) = \binom{n}{0} \cdot p^0 \cdot (1 - p)^n$. Pertanto otteniamo:

$$P(T = n) = p(1 - p)^{n-1}$$

Pertanto osserviamo che la distribuzione geometrica segue le regole della distribuzione binomiale.

Possiamo anche osservare che **la distribuzione geometrica non ha memoria**. Ciò significa che:

$$P(X = n + m|X > n) = P(X = m) \quad \forall n, m \in \mathbb{N}, m > 0$$

Ossia che la probabilità di avere il successo dopo $n+m$ lanci sapendo che dopo $n$ lanci non ho ottenuto successo è uguale alla probabilità di ottenere successo dopo $m$ lanci (i.e. ogni lancio è indipendente). Banalmente si osserva che:

$$P(X = n + m|X > n) = \frac{P(X = n + m \land X > n)}{P(X > n)} = \frac{P(X = n + m)}{P(X > n)}$$

Per il fatto che $m > 0$, per cui è scontato che se $X = n + m \implies X > n$. Svolgendo i calcoli:



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

$$\frac{P(X=n+m)}{P(X>n)} = \frac{p(1-p)^{n+m-1}}{(1-p)^n} = p(1-p)^{m-1}$$

Questo risultato è ottenuto anche senza la premessa che i lanci siano indipendenti. Ossia, il risultato è dedotto **esclusivamente** dalla densità di probabilità, non dalla premessa che i lanci sono indipendenti.

Se $X \sim Geom(p)$ si ha che $E[X] = \frac{1}{p}$ e $Var(X) = \frac{1-p}{p^2}$. Se $p \to 1$, la varianza converge a zero. Se $p \to 0$, la varianza diverge.

## Distribuzione ipergeometrica

La distribuzione ipergeometrica descrive estrazioni **senza rimpiazzo**. Questo significa che **ogni lancio non è indipendente**. Per capire ciò, consideriamo l'estrazione di $k$ palline con una certa proprietà su $n$ palline. $X$ è la variabile aleatoria che conta quante volte la pallina con la proprietà desiderata è stata estratta dopo $m$ estrazioni. Ad ogni istante possiamo calcolare che:

$$P(X=x) = \frac{\text{casi favorevoli}}{\text{casi possibili}} = \frac{\binom{k}{x} \cdot \binom{n-k}{m-x}}{\binom{n}{m}}$$

La formula può essere ostica, ma in realtà è molto semplice da capire. $\binom{k}{x}$ è il numero di sottoinsiemi di $x$ elementi partendo da $k$ elementi: sono i gruppi di $x$ palline che godono della proprietà cercata. $\binom{n-k}{m-x}$ è il numero di gruppi di $m-x$ palline partendo da $n-k$ palline: sono i gruppi di $m-x$ palline che non godono della proprietà cercata. Stiamo pertanto unendo due insiemi:

- l'insieme di $x$ palline che godono della proprietà cercata;
- l'insieme di $m-x$ palline restanti, che non godono della proprietà cercata.

$\binom{n}{m}$ è il numero di gruppi di $m$ palline, quindi tutte le possibili estrazioni.

Di questa distribuzione possiamo calcolare anche speranza e varianza:

$$E[X] = m \cdot \frac{k}{n} Var(X) = \frac{m \cdot k(n-k)}{n^2} \cdot \frac{m-k}{m-1}$$

La speranza quindi è proporzionale al rapporto tra palline con proprietà e palline totali. La varianza sarà nulla pertanto se tutte o nessuna delle palline godono di quella proprietà.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

# 05 - Variabili aleatorie continue

Sia $(\Omega, \mathcal{A}, P)$ spazio di probabilità e sia $X$ una variabile aleatoria:

$$X : \Omega \to \mathbb{R}$$

Se $X(\Omega)$ è infinito non numerabile, ossia continuo, la variabile aleatoria si dice **continua**.
Per le variabili aleatorie discrete, abbiamo detto che la legge di $X$ è:

$$P(X = x)$$

Il problema nell'insieme continuo riguarda la continuità di $X(\Omega)$; tale insieme infatti ha **misura nulla**. Il punto è zero-dimensionale, per cui non possiamo associare ad un punto una probabilità. In termini più pratici, non si parla di probabilità in un punto, ma in intervalli. La probabilità, inoltre, è uguale sia che si considerino o meno gli estremi dell'intervallo. Pertanto, partiremo dalla probabilità di osservare un intervallo di valori.

Possiamo pertanto dire che $X$ variabile aleatoria è continua se la sua funzione di ripartizione è continua a destra. Ossia, data:

$$F_X : \mathbb{R} \to [0, 1]$$

Abbiamo detto che:

$$F(t) = P(X \leq t)$$

La funzione di ripartizione gode di quattro proprietà:

1. $0 \leq F(t) \leq 1 \quad \forall t \in \mathbb{R}$;
2. $F(t_1) \leq F(t_2) \quad \forall t_1, t_2 \in \mathbb{R} : t_1 \leq t_2$. Ossia la funzione è **monotona crescente**;
3. $\lim_{t \to -\infty} F(t) = 0$;



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

4. $\lim_{\epsilon \to 0^+} F(t+\epsilon) = F(t)$. Questo perché possono esistere funzioni non continue a sinistra, i.e. salti nella funzione. In tal caso si può definire la probabilità nel punto di discontinuità, ma **solo** in questo caso. Di nostro interesse sono tuttavia le funzioni continue.

Dalla continuità di $F_X$ deduciamo l'impossibilità di definire la probabilità nel punto. Infatti si avrebbe che la probabilità nel punto è data da:

$$\lim_{\epsilon \to 0} P(t-\epsilon \leq X \leq t+\epsilon) = \lim_{\epsilon \to 0} F_X(t+\epsilon) - F_X(t-\epsilon) =$$

$$= F_X(t) - F_X(t) = 0$$

Ma questo se assumiamo che $F_X$ è continua. Dalla proprietà $4$ sappiamo solo che $F_X$ è continua a destra, ma non a sinistra. Pertanto:

$$\lim_{\epsilon \to 0} F_X(t-\epsilon) \neq F(t)$$

In questo caso ha senso definire la probabilità $P(X=t)$ e si ha:

$$P(X=t) = F_X(t^+) - F_X(t^-)$$

Dalla funzione di ripartizione è possibile definire la densità di probabilità. Più precisamente possiamo dire che una variabile aleatoria $X$ si dice **assolutamente continua** se ammette densità $f(x)$ tale che:

$$F_X(t) = \int_{-\infty}^{t} f(x)\, dx$$

dove $f(x) \geq 0 \quad \forall x \in \mathbb{R}$ e $\int_{-\infty}^{+\infty} f(x)\, dx = 1$. Quest'ultimo vincolo prende il nome di **vincolo di normalizzazione**.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

## Speranza matematica

Definiamo speranza matematica di $X$ variabile aleatoria continua il valore:

$$E[X] = \int_{-\infty}^{+\infty} f(x) \cdot x \, dx$$

Questa vale se chiaramente esiste finito:

$$\int_{-\infty}^{+\infty} f(x) \cdot |x| \, dx < +\infty$$

Assumeremo sempre che esiste finito tale integrale. Inoltre, se esistono finiti, definiamo **momenti di ordine** $k$ la quantità:

$$E[X^k] = \int_{-\infty}^{+\infty} x^k \cdot f(x) \, dx$$

e **momenti centrati di ordine** $k$:

$$E[(X - E[X])^k] = \int_{-\infty}^{+\infty} (x - E[X])^k \cdot f(x) \, dx$$

La speranza matematica, i momenti di ordine $k$ e i momenti centrati di ordine $k$ godono delle stesse proprietà viste per le variabili aleatorie discrete.

## Quantile di ordine $\alpha$

Definiamo **quantile di ordine** $\alpha$ e scriviamo $q_\alpha$ il valore sull'asse $x$ che divide l'area sottesa al grafico in due parti $\alpha$ e $1 - \alpha$. Pertanto se ad una certa $x$ identifichiamo $q_\alpha$, allora l'integrale:

$$\int_{-\infty}^{q_\alpha} f(x) \, dx = \alpha$$

Pertanto il quantile di ordine $\alpha$ è tale che $P(X \leq q_\alpha) = \alpha$. I quantili aiutano a rispondere alla domanda:"Data una distribuzione di probabilità, qual è la quantità più alta che rappresenta una percentuale $x$ di popolazione?".



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

# 06 - Distribuzioni continue

## Distribuzione uniforme

La distribuzione uniforme ha legge su $[a, b]$:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{se } x \in [a, b] \\ 0 & \text{altrimenti} \end{cases}$$

Una distribuzione uniforme non è ben definita su tutto $\mathbb{R}$. Il grafico è una retta orizzontale $y = \frac{1}{b-a}$ quando $x$ è nel range $[a, b]$. Si tratta di una densità:

- la frazione è positiva in quanto $b \geq a \implies f(x) \geq 0$ sempre;
- $\int_{-\infty}^{+\infty} f(x)\, dx = \int_a^b f(x)\, dx = \frac{1}{b-a} \int_a^b dx = \frac{1}{b-a} \cdot (b-a) = 1.$

La funzione di ripartizione:

$$F_X(t) = \int_{-\infty}^{t} f(x)\, dx$$

è da analizzare in 3 casi:

1. $t < a$. Se $t < a \implies f(x) = 0$ e la funzione di ripartizione è $0$;
2. $t \in [a, b] \implies f(x) = \frac{1}{b-a}$ e si ha l'integrale da $a$ a $t$ e si ha $F_X(t) = \frac{t-a}{b-a}$. $F_X(t)$ è allora lineare con coefficiente angolare $\frac{1}{b-a}$;
3. $t > b \implies f(x) = 0$ e la funzione di ripartizione è uguale a $1$ in quanto abbiamo già incontrato tutti gli altri casi.

Concludendo:

$$F_X(t) = \begin{cases} 0 & \text{se } t < a \\ \frac{t-a}{b-a} & \text{se } t \in [a,b] \\ 1 & \text{se } t > b \end{cases}$$

$F_X$ è continua in $\mathbb{R}$.

### Speranza matematica

La speranza matematica è:

$$E[X] = \int_{-\infty}^{+\infty} x \cdot f(x)\, dx = \int_a^b \frac{x}{b-a} = \frac{1}{b-a} \cdot \frac{b^2-a^2}{2} = \frac{b+a}{2}$$

Ossia il punto medio dell'intervallo.

### Varianza

La varianza è:

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

La varianza è nulla se gli estremi coincidono.

## Distribuzione normale o gaussiana

Questa distribuzione è caratteristica di un sistema con varianza finita.

La densità di probabilità è:

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

Il massimo è in $x = 0$ ed è $\frac{1}{\sqrt{2\pi}}$.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

Una normale standard si indica con $X \sim N(0,1)$, ossia con media pari a $0$ e varianza pari ad $1$ ed è una curva a campana centrata simmetrica rispetto all'asse $y$.

### Speranza matematica, Varianza

La speranza matematica è pari a zero in quanto corrisponde all'integrale:

$$E[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \cdot e^{-\frac{x^2}{2}} = 0$$

Poiché è una funzione dispari integrata su un dominio simmetrico.

Ne segue che la varianza è pari ad $1$. La varianza inoltre indica l'ampiezza della campana.

### Quantili nella gaussiana

Nella gaussiana i quantili sono indicati con $\phi_\alpha$. La normale standard è tale che il $99.9\%$ dei dati giace tra $-3$ a $3$. Le proprietà di simmetria della guassiana sono tali per cui i quantili sono simmetrici. Si ha inoltre che sui quantili $\phi_{0.005}$ e $\phi_{0.995}$ ci sono i punti $-3$ e $3$. Quindi l'integrale tra $-\infty$ e $-3$ di $f(x)$ è pari a $0.005$.

### Variabili aleatorie alternative per ottenere diverse gaussiane

Definiamo $Y$ come:

$$U = \sigma X + \mu$$

dove $X \sim N(0,1)$. Avremo:

$$E[Y] = \sigma \cdot E[X] + \mu = 0 + \mu = \mu$$

La varianza:

$$Var(Y) = Var(\sigma X) + \mu = \sigma^2 \cdot Var(X) = \sigma^2$$



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

Tale variabile segue i parametri $Y \sim N(\mu, \sigma^2)$; ha densità:

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

Cambia il punto di massimo, ossia in $x = \mu$. Avremo allora che la funzione è simmetrica rispetto all'asse $x = \mu$. Maggiore è $\sigma$ più lentamente scende la gaussiana e quindi più è larga la campana. I quantili, inoltre, di ordine $\alpha$ saranno traslati seguendo la trasformazione:

$$\phi_\alpha \to \sigma\phi_\alpha + \mu$$

### Regola dei tre $\sigma$ (o legge $68 - 95 - 99.7$)

Si ha che:

- tra $\mu - \sigma$ e $\mu + \sigma$ l'area è $0.68$;
- tra $\mu - 2\sigma$ e $\mu + 2\sigma$ l'area è $0.95$;
- tra $\mu - 3\sigma$ e $\mu + 3\sigma$ l'area è $0.997$.

### Composizione di gaussiane

Siano $Y_1 \sim N(\mu_1, \sigma_1^2), Y_2 \sim N(\mu_2, \sigma_2^2)$, allora:

$$Y = Y_1 + Y_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

# 07 - Altre distribuzioni continue

## Legge esponenziale

La legge esponenziale descrive la durata di vita di un fenomeno che non invecchia, ossia che **non ha memoria**.

Si ha distribuzione:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x > 0 \\ 0 & \text{altrimenti} \end{cases}$$

Diremo allora che $X \sim \exp(\lambda)$. La distribuzione esponenziale è l'equivalente continuo della distribuzione di Poisson. $\lambda$ esprime quanto rapidamente invecchia il fenomeno.

$f(x)$ è una distribuzione di probabilità, infatti:

- $f(x) \geq 0$ sempre;
- $\int_{-\infty}^{+\infty} f(x) = \int_{0}^{+\infty} \lambda e^{-\lambda x} = -e^{-\lambda x}|_0^{+\infty} = -(-1) = 1.$

La funzione di ripartizione:

$$F_X(t) = \int_0^t f(x)\,dx = 1 - e^{-\lambda t} = P(X \leq t)$$

Solitamente a noi interessa sapere $P(X \geq t)$. Definiamo allora la **funzione di sopravvivenza**:

$$P(X \geq t) = 1 - F_X(t) = e^{-\lambda t}$$

La probabilità di sopravvivenza decade esponenzialmente. Più tempo passa, minore è il tempo di osservazione dell'evento.

Dimostriamo ora che $f(x)$ non ha memoria. Per dimostrare che non ha memoria:

$$P(X \geq t+x|X \geq x) = P(X \geq t)$$

Chiaramente se $x, t > 0$:

$$P(X \geq t+x|X \geq x) = \frac{P(X \geq t+x, X \geq x)}{P(X \geq x)}$$

Ma se $t > 0$, allora è ovvio che se $P \geq t+x$ allora $P \geq x$. Allora:

$$P(X \geq t+x, X \geq x) = P(X \geq t+x) = e^{-\lambda(x+t)}$$

Pertanto:

$$P(X \geq t+x|X \geq x) = \frac{e^{-\lambda(x+t)}}{e^{-\lambda x}} = e^{-\lambda t} = P(X \geq t)$$

### Varianza, speranza matematica

Si ha che:

$$E[X] = \frac{1}{\lambda}$$

Mentre che:

$$Var(X) = \frac{1}{\lambda^2}$$

## Distribuzione del $\chi$ quadro

Si indica con $X \sim \chi^2(n)$. $n$ indica i **gradi di libertà** della distribuzione. Risulta legata alla normale:

$$Y_1, Y_2, ..., Y_n \sim N(0,1) \implies \sum_{i=1}^{n} Y_i^2 \sim \chi^2(n)$$



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

Ossia considerando $n$ variabili aleatorie che seguono distribuzioni normali di media $0$ e varianza $1$, la loro somma al quadrato è una distribuzione del $\chi$ quadro.

La $\chi$ quadro è la distribuzione che descrive la somma dei quadrati di una normale. Vale inoltre:

$$X \sim \chi^2(n), Y \sim \chi^2(m) \implies X + Y \sim \chi^2(n+m)$$

## Distribuzione $T$ di Student

Si indica con $\mathcal{T} \sim t(n)$. Tale distribuzione gode di alcune proprietà:

- è **simmetrica**;
- ha **media nulla**;
- il quantile si indica con $t_\alpha$, e poiché è simmetrica lo sono anche i quantili, quindi $t_\alpha = -t_{1-\alpha}$;
- se $n \to +\infty$, allora $t(n)$ tende ad una gaussiana $N(0,1)$. In buona approssimazione, $n \to +\infty$ significa $n \geq 30$.

Se il campione è piccolo si osservano differenze tra la $\mathcal{T}$ di Student e la gaussiana.

## Cenni di variabili aleatorie multivariate

Siano due variabili aleatorie $X, Y$ che possibilmente seguono distribuzioni differenti. La distribuzione di entrambe prende il nome di **distribuzione** (o **densità**) **congiunta** $f(x,y)$ e darà la probabilità di ottenere la coppia $(x,y)$.

La densità congiunta permette di ricavare le densità di probabilità di $X$ e $Y$ separatamente, ossia $f_X(x)$ e $f_Y(y)$. Queste densità prendono il nome di **densità marginali**:

$$f_X(x) = \int_{\mathbb{R}} f(x,y)\,dy f_Y(y) = \int_{\mathbb{R}} f(x,y)\,dx$$

Scegliere una delle due variabili significa scegliere una direzione dalla quale guardare il grafico. Naturalmente sono densità e sono normalizzate.

Possiamo anche calcolare la **densità condizionale**:



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

$$F_{X|Y}(x,y) = \begin{cases} \frac{f(x,y)}{f(y)} & \text{se } f(y) > 0 \\ 0 & \text{altrimenti} \end{cases}$$

Se le variabile aleatorie sono indipendenti, si ha che la densità congiunta è tale che:

$$f(x,y) = f_X(x) \cdot f_Y(y)$$

### Correlazione

Talvolta è possibile che dati tra di loro siano **correlati**. La correlazione, tuttavia, non implica **causalità**: non diciamo infatti che un evento è la causa di un altro, ma che gli eventi sono legati tra loro.

Se $Y$ e $X$ sono **correlati**, potrebbero essere correlati **linearmente**. La corellazione lineare è rivelata dal **coefficiente di correlazione lineare** $\rho$, ed è il rapporto:

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}} = \begin{cases} 0 & \text{eventi indipendenti} \\ 1 & \text{correlazione lineare} \\ -1 & \text{correlazione antilineare} \end{cases}$$

$\rho$ può assumere tutti i valori tra $-1$ e $1$, e sulla base dei valori assunti da $\rho$ possiamo capire quanto sono correlate le variabili:

- $0 < |\rho| \leq 0.3 \implies$ la correlazione è debole;
- $0.3 < |\rho| \leq 0.7 \implies$ la correlazione è media;
- $0.7 < |\rho| \leq 1 \implies$ la correlazione è forte.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

# 08 - Introduzione alla statistica

Le variabili aleatorie di cui abbiamo parlato devono diventare dati reali. Occorre generalizzare alcuni concetti.

Immaginiamo di avere un **campione** $x_1, ..., x_n$. Di tale campione possiamo dare le seguenti:

- **media**. La media del campione è $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Al momento si riferisce solo al set di dati;
- **varianza empirica**. La varianza empirica è $\overline{\sigma}_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$. La varianza empirica ci dice quanto i dati si allargano dalla media.

Dati due campioni $x_1, ..., x_n$ e $y_1, ..., y_n$, definiamo **covarianza empirica** il valore:

$$\overline{\sigma}_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

Se la covarianza empirica è nulla, i campioni sono **scorrelati**. In caso contrario, i dati possono essere correlati e ha senso chiedersi se sono correlati linearmente.
Possiamo calcolare il coefficiente di **correlazione lineare**:

$$\rho = \frac{\overline{\sigma}_{xy}}{\overline{\sigma}_x \overline{\sigma}_y}$$

## Regressione lineare

Dato un campione di coppie di dati $(x_i, y_i)$, il cui coefficiente di correlazione lineare $\rho \sim 1$, ci chiediamo quale sia la retta che descrive meglio la correlazione. Rappresentiamo allora i dati sugli assi cartesiani, e cerchiamo una metrica da minimizzare per trovare la retta che meglio descrive i dati.

Vogliamo allora:

$$\hat{y} = \beta_0 + \beta_1 x$$

Cerchiamo allora di modulare $\beta_0, \beta_1$ opportunamente. Più precisamente, vogliamo una retta che minimizza gli errori su tutti i punti. Per motivi di segno, minimizziamo l'errore quadratico medio (ossia lo **scarto quadratico medio**). Tale scarto $S^2$ è quindi:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(\beta_0 + \beta_1 x_i - y_i)^2$$

Il problema dipende dalle due incognite $\beta_0, \beta_1$, pertanto possiamo scrivere $S^2(\beta_0, \beta_1)$.

### Metodo dei minimi quadrati

Minimizziamo la funzione $S^2(\beta_0, \beta_1)$ rispetto alle variabili $\beta_0, \beta_1$.
Calcoliamo la derivata rispetto a $\beta_0$:

$$\frac{\delta S^2}{\delta \beta_0} = \frac{\delta}{\delta \beta_0}[\frac{1}{n}\sum_{i=1}^{n}(\beta_0 + \beta_1 x_i - y_i)^2] = \frac{1}{n}\sum_{i=1}^{n} 2(\beta_0 + \beta_1 x_i - y_i)$$

Ci interessa che tale quantità sia uguale a zero:

$$\frac{2}{n}(\sum_{i=1}^{n}\beta_0 + \sum_{i=1}^{n}\beta_1 x_i - \sum_{i=1}^{n} y_i) = 0 \iff$$

$$\iff \frac{2}{n}(n\beta_0 + \beta_1\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i) = 0 \iff$$

$$\iff \frac{2}{n}(n\beta_0 + n\beta_1\overline{x} - n\overline{y}) = 0$$

$$\iff 2(\beta_0 + \beta_1\overline{x} - \overline{y}) = 0 \iff \beta_0 = \overline{y} - \beta_1\overline{x}$$

La derivata invece rispetto a $\beta_1$ è:



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

$$\frac{\delta S^2}{\delta \beta_1} = \frac{2}{n}\sum_{i=1}^{n} x_i(\beta_0 + \beta_1 x_i - y_i) = 0 \iff$$

$$\iff \frac{2}{n}(\sum_{i=1}^{n}\beta_0 x_i + \sum_{i=1}^{n}\beta_1 x_i^2 - \sum_{i=1}^{n} x_i y_i) = 0 \iff$$

$$\iff 2(\beta_0 \overline{x} + \beta_1 \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\sum_{i=1}^{n} x_i y_i) = 0$$

Utilizziamo la relazione precedente per sostituire $\beta_0$:

$$2[(\overline{y} - \beta_1 \overline{x})\overline{x} + \beta_1 \overline{x^2} - \overline{xy}] = 0$$

Cerchiamo il valore di $\beta_1$:

$$2[\overline{y} \cdot \overline{x} - \overline{xy} + \beta_1(\overline{x^2} - \overline{x}^2)] = 0 \iff$$

$$\iff \beta_1(\overline{x^2} - \overline{x}^2) = \overline{xy} - \overline{y} \cdot \overline{x} \iff \beta_1 = \frac{\overline{\sigma}_{xy}}{\overline{\sigma}_x^2}$$

La nostra retta pertanto sarà:

$$\hat{y} = \overline{y} - \frac{\overline{\sigma}_{xy}}{\overline{\sigma}_x^2}\overline{x} + \frac{\overline{\sigma}_{xy}}{\overline{\sigma}_x^2}x$$

Tale modello lineare ci permette di fare previsioni in virtù della correlazione lineare dei dati.

Osserviamo che il punto $(\overline{x}, \overline{y})$ appartiene alla retta. Infatti il modo in cui si descrive la regressione lineare è quella di considerare dei punti che stanno sulla retta ma con un errore gaussiano. Possiamo allora considerare la varianza dei punti rispetto al comportamento della retta.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

### Cenni alla regressione lineare multipla

Chiaramente non è sempre detto che il modello dipenda da un solo parametro, ma possiamo avere un modello più complesso:

$$h_\beta : K^d \to K$$

dove $K$ indica un insieme di dimensione $d$. I parametri inoltre potrebbero anche dipendere non in maniera lineare:

$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2$$

Possiamo allora applicare una trasformazione $\phi(x)$ per passare da uno spazio all'altro:

$$\begin{pmatrix} x \end{pmatrix} \rightsquigarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

in modo tale da poter lavorare linearmente anche con un parametro quadratico.

Per capire se un modello lineare è buono, possiamo calcolare il **coefficiente di determinazione** $R$:

$$R = \rho^2 = \frac{\overline{\sigma}_{xy}^2}{\overline{\sigma}_x^2 \overline{\sigma}_y^2} \in [0, 1]$$

Più $R$ è vicino ad 1, più è buono il modello. In caso contrario, bisogna applicare la trasformazione $\phi(x)$.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

# 09 - Legge dei grandi numeri, teorema del limite centrale

## Premessa per la legge dei grandi numeri

La legge dei grandi numeri ci legittima a calcolare la media teorica utilizzando la media campionaria dei dati.

Immaginiamo di avere una successione di $n$ variabili aleatorie indipendenti $\{X_n\}$ che vengono tutte dalla stessa distribuzione di media $\mu$ e varianza $\sigma^2$. La **media campionaria** è la variabile aleatoria:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

La media è a sua volta una variabile aleatoria in quanto ogni termine $X_i$ assume valori dipendenti dagli esiti. Ovviamente possiamo sommare perché $X_n$ è una successione quindi gli esiti sono finiti. La media campionaria è tale che:

$$E[\overline{X}_n] = \mu$$

Si ha infatti che:

$$E[\overline{X}_n] = E[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\cdot n \cdot \mu$$

Vogliamo anche capire la varianza:

$$Var(\overline{X}_n) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

Al tendere di $n$ ad infinito, non si ha una variabile aleatoria ma un numero, ossia la media della distribuzione da cui provengono le variabili aleatorie.

La successione $(X_n)_{n\in\mathbb{N}}$ **converge in probabilità** a $X$, e si indica con $X_n \to^P X$, se:

$$\lim_{n\to\infty} P(|X_n - X| \geq \eta) = 0 \quad \forall \eta > 0$$

Ossia la distanza tra l'$n-$esimo elemento della successione e una variabile aleatoria ha probabilità nulla di essere diversa da zero al tendere di $n$ ad infinito.

## Legge dei grandi numeri

Sia $X_n$ un insieme di variabili aleatorie indipendenti, tutte della stessa distribuzione di media $\mu$ e varianza $\sigma^2$. Allora:

$$\overline{X}_n \to^P \mu$$

Ossia, $\overline{X}_n$ converge in probabilità a $\mu$. La variabile aleatoria $\overline{X}_n$ è così definita:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Questo, come visto, ci legittima ad utilizzare la media campionaria per il calcolo della media teorica. Tuttavia, nella realtà non è sempre possibile: il campione rappresentativo potrebbe avere una certa **bias**, ossia essere **polarizzato**.

### Dimostrazione

Abbiamo bisogno di fare tre passaggi:

1. dimostrare la **disuguaglianza di Markov**;
2. dimostrare che **disuguaglianza di Chebicev**;
3. dimostrare infine la legge dei grandi numeri.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

**Disuguaglianza di Markov**

Sia $X$ una variabile aleatoria a valori non negativi. Allora:

$$P(X \geq \alpha) \leq \frac{E[X]}{\alpha} \quad \forall \alpha > 0$$

Introduciamo una variabile aleatoria ausiliaria $I$ tale che:

$$I = \begin{cases} 1 & \text{se } X \geq \alpha \\ 0 & \text{altrimenti} \end{cases}$$

Poiché $\alpha > 0$ allora:

$$I = \begin{cases} 1 & \text{se } \frac{X}{\alpha} \geq 1 \\ 0 & \text{altrimenti} \end{cases}$$

Chiaramente nel primo caso si ha che $I \leq \frac{X}{\alpha}$, perché siamo nel caso in cui tale frazione ha valore maggiore o uguale ad $1$. Poiché $I$ in questo caso vale $1$, la proprietà è banalmente verificata. Anche nel secondo caso la proprietà è verificata: $X$ è sempre positiva, così come $\alpha$, per cui la frazione è sempre maggiore o uguale a zero.

A questo punto possiamo concludere che:

$$0 \leq I \leq \frac{X}{\alpha}$$

Calcoliamo il valore atteso di $I$:

$$E[I] = 1 \cdot P(X \geq \alpha) + 0 \cdot P(X < \alpha) = P(X \geq \alpha)$$

Poiché vale la proprietà:

$$X \leq Y \implies E[X] \leq E[Y]$$

Dalla catena deriviamo:



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

$$E[I] \leq E[\frac{X}{\alpha}] \implies E[I] \leq \frac{E[X]}{\alpha} \implies P(X \geq \alpha) \leq \frac{E[X]}{\alpha}$$

**Disuguaglianza di Chebicev**

Sia $X$ una variabile aleatoria che ammette speranza matematica e varianza **finite**. Allora:

$$P(|X - E[X]| \geq \eta) \leq \frac{Var(X)}{\eta^2} \quad \forall \eta > 0$$

Per dimostrarlo, introduciamo la variabile aleatoria $Y$:

$$Y = (X - E[X])^2$$

$Y$ gode di due proprietà:

- $E[Y] = Var(X)$, infatti $E[Y] = E[(X - E[X])^2]$, che per definizione è la varianza;
- $Y \geq 0$, per cui vale la disuguaglianza di Markov.

Si ha pertanto:

$$P(Y \geq \alpha) \leq \frac{Var(X)}{\alpha} \quad \forall \alpha > 0$$

Quindi ciò vale anche per $\eta^2$:

$$P((X - E[X])^2 \geq \eta^2) \leq \frac{Var(X)}{\eta^2}$$

Notiamo che:

$$(X - E[X])^2 \geq \eta^2 \iff |X - E[X]| \geq \eta$$

Basta infatti applicare la radice quadrata ad entrambi i membri. La relazione vale anche senza i quadrati:

$$P(|X - E[X]| \geq \eta) \leq \frac{Var(X)}{\eta^2}$$



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

**Dimostrazione dell'enunciato**

La legge dei grandi numeri ha come tesi:

$$\lim_{n\to+\infty} P(|\overline{X}_n - \mu| \geq \eta) = 0$$

Applichiamo la disuguaglianza di Chebicev alla media campionaria, sotto le ipotesi enunciate precedentemente:

$$P(|\overline{X}_n - E[\overline{X}_n]| \geq \eta) \leq \frac{Var(\overline{X}_n)}{\eta^2}$$

Osserviamo tuttavia che la varianza di $\overline{X}_n$ è pari a $\frac{\sigma^2}{n}$, mentre la speranza è $\mu$. Sostituendo:

$$P(|\overline{X}_n - \mu| \geq \eta) \leq \frac{\sigma^2}{n \cdot \eta^2}$$

Per $n \to +\infty$, si ha che la frazione tende a zero. Pertanto la probabilità dev'essere minore o uguale a zero, e poiché la probabilità non può essere minore di zero, si ha:

$$\lim_{n\to+\infty} P(|\overline{X}_n - \mu| \geq \eta) = 0$$

## Teorema del limite centrale

### Premesse

Una successione di variabili aleatorie indipendenti $(X_n)_{n\in\mathbb{N}}$ **converge in legge a** $X$, e si scrive $(X_n)_{n\in\mathbb{N}} \to^{\mathcal{L}} X$ se:

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

dove $F_n(t)$ è la funzione di ripartizione di $X_n$ e $F(t)$ quella di $X$. Pertanto possiamo utilizzare le proprietà delle singole variabili aleatorie per $n$ molto grandi.

Tale convergenza viene utilizzata nel teorema del limite centrale.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

### Enunciato

Sotto le ipotesi precedenti, definiamo la variabile aleatoria **normale standardizzata**:

$$S_n = \frac{\overline{X}_n - \mu}{\sigma}\sqrt{n}$$

Si ha che:

$$S_n \to^{\mathcal{L}} S \sim N(0,1)$$

per $n \to +\infty$.

Essenzialmente stiamo dicendo che, indipendentemente dall'esperimento, raccogliendo un numero sufficiente di dati, la variabile aleatoria $S_n$ seguira una distribuzione normale.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

# 10 - Statistica

La statistica è il metodo scientifico che analizza i dati. Esistono due tipi di statistica:

- **descrittiva**. Si occupa di descrivere il campione;
- **inferenziale**. Si occupa di inferire le proprietà della popolazione.

Parliamo di fenomeni che hanno un certo grado di casualità. Utilizziamo le densità di probabilità che ci permette di capire quali sono i dati che possiamo osservare. Definiamo **popolazione** l'intero insieme statistico da analizzare. Si dice invece **campione** l'insieme dei dati raccolti e solitamente incluso nella popolazione. Osservando il campione, la statistica inferenziale cerca di estrarre informazioni su tutta la popolazione.

Ci occuperemo per ora di statistica descrittiva.

## Statistica descrittiva

Dato un set di dati grezzi, ossia non organizzati, è utile studiarne uno o più caratteri. Tali caratteri possono essere di due tipi:

- **qualitativi**, ossia non numerici;
- **quantitativi**, ossia numerici, divisi in **discreti** e **continui**.

### Raggruppamento per valore

Dato un carattere quantitativo discreto, possiamo raggrupparlo per singolo valore. Sia $N$ il numero di dati raccolti. I dati possono ripetersi: indichiamo pertanto con $n$ il numero di **dati distinti**. Per ogni valore distinto è possibile determinare:

- **frequenza assoluta** $f_i$, ossia il numero di volte in cui compare l'elemento $i-$esimo. Chiaramente $\sum_{i=1}^{n} f_i = N$;



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

- **variabilità** (o **range**), ossia la differenza tra il valore massimo e il minimo;
- **frequenza relativa** (o **probabilità empirica**) $p_i$, definita come il rapporto $p_i = \frac{f_i}{N}$. Chiaramente $\sum_{i=1}^{n} p_i = 1$. Ricorda la definizione frequentista di probabilità.
- **frequenze relative cumulative** $F_i$, definita come $F_i = \sum_{k=1}^{i} p_k$. Ricorda la funzione di ripartizione. Chiaramente $F_n = 1$.

Operativamente:

- si considera il campione numerico;
- si ordina tale campione in maniera crescente;
- si costruisce una tabella di colonne $x_i, f_i, p_i, F_i$. Tale tabella raccoglie le informazioni principali del campione. La rappresentazione di tale tabella avviene tramite **istogramma**, o meglio un **istogramma normalizzato**. La principale differenza tra i due risiede nell'altezza dei rettangoli, che è scalata rispetto ad $N$, per cui la somma delle aree dei rettangoli (assumendo una base unitaria) è pari ad $1$.

Altri indicatori utili sono:

- **moda**, il valore più frequente. Può non essere unica e in tal caso si parla di **distribuzione multimodale**;
- **media**, ossia la media aritmetica dei dati $\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$. Da non confondere con la media campionaria, che è una **variabile aleatoria**. Più è grande il campione, più la media aritmetica è vicina alla media campionaria. Inoltre, in virtù della definizione di $\overline{x}$, possiamo anche scrivere $\overline{x} = \frac{1}{N} \sum_{i=1}^{n} x_i \cdot f_i = \sum_{i=1}^{n} x_i \frac{f_i}{N} = \sum_{i=1}^{n} x_i p_i$. Questa forma è molto simile alla speranza matematica;
- **mediana** $m_e$, ossia il numero che divide i dati in due insiemi ugualmente numerosi. Richiama il quantile di ordine $\frac{1}{2}$. $m_e$ minimizza la funzione $S(a) = \sum_{i=1}^{N} |x_i - a|$.
- **quartili**, ossia i dati che dividono l'insieme dei dati in quattro sottoinsiemi ugualmente numerosi. La mediana è il secondo quartile. Si hanno quindi il 25% dei dati a sinistra del primo quartile, 50% a sinistra del secondo, 75% a sinistra del terzo. Li indicheremo rispettivamente con $q_1, q_2, q_3$. Definiamo **distanza interquantile** la distanza $q_3 - q_1$, e contiene il 50% dei dati. Inoltre, il primo quartile è la mediana della prima metà dei dati, il terzo quartile è la mediana della seconda metà dei dati;
- **varianza empirica** $\overline{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 p_i$. La semplificazione della formula è analoga a quella della media;



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

- **deviazione standard empirica** $\overline{\sigma} = \sqrt{\overline{\sigma}^2}$;
- **momento empirico di ordine** $k$ $\overline{\mu}_k = \frac{1}{N}\sum_{i=1}^{N} x_i^k = \sum_{i=1}^{n} x_i^k \cdot p_i$;
- **momento empirico centrato di ordine** $k$ $\mu_k = \sum_{i=1}^{n} p_i (x_i - \overline{x})^k$.

### Coefficiente di asimmetria (o skewness)

Definiamo **coefficiente di asimmetria** o **skewness** la quantità:

$$\beta_1 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}}$$

Si confrontano i momenti centrati di ordine $3$ e $2$ per verificare se i dati sono simmetrici o asimmetrici. Se il coefficiente è uguale a zero, la distribuzione dei dati è simmetrica, quindi media e mediana coincidono; se è maggiore di zero, la distribuzione è spostata a destra; infine, se è minore di zero la distribuzione è spostata a sinistra.

Solitamente si può anche trovare la forma:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

perdendo tuttavia l'informazione del segno, e quindi dello "spostamento" della distribuzione a destra o sinistra.

### Coefficiente di curtosi

Definiamo **coefficiente di curtosi** la quantità:

$$\gamma = \beta_2 - 3$$

dove:

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

prende il nome di **curtosi** e misura l'appiattamento della curva rispetto alla distribuzione normale. Pertanto la quantità $3$ è la curtosi della guassiana. Un coefficiente di curtosi maggiore di zero indica



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

una distribuzione più “appuntita” rispetto alla normale, minore di zero più “piatta” e uguale a zero “piatta” quanto la normale.

### Raggruppamento per classi di valore

Invece di raccogliere i dati per valore, possiamo raccoglierli per classi di valore. Si considerano allora sottoinsiemi di dati e si adattano le definizioni analizzate precedentemente in virtù di tale raggruppamento:

- la **frequenza assoluta** $f_i$ è il numero di dati nell'$i-$esima classe;
- la **frequenza relativa** $p_i$ è il rapporto $\frac{f_i}{N}$;
- la **frequenza relativa cumulata** $F_i$ è la quantità $\sum_{k=1}^{i} p_k$. Si assume che cresca linearmente l'estremo sinistro e l'estremo destro dell'intervallo delle classi. Tale approssimazione è utile in quanto la definizione del quartile è identica a quella teorica.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

# 11 - Rappresentazione dei dati

## Istogramma

L'istogramma è ideale per mostrare i dati in possesso e la frequenza con cui occorrono. Ad ogni valore è associato un rettangolo la cui altezza è pari alla frequenza assoluta o relativa. L'istogramma con le frequenze relative prende il nome di **istogramma normalizzato** ed è necessario per confrontare la distribuzione dei dati con quelle teoriche.

## Grafico a torta

Il grafico a torta è ideale per mostrare i dati e le loro frequenze relative. Un cerchio viene diviso in un numero di spicchi pari al numero di valori. L'area di ogni spicchio è proporzionale alla frequenza relativa del valore corrispondente allo spicchio.

## Grafico a scatola con i baffi (box plot)

L'obiettivo è descrivere qualitativamente l'istogramma nella maniera più completa. La mediana divide a metà l'universo dei dati, e i quartili dividono l'insieme dei dati in quarti.

Una box rappresenta il 50% dei dati. L'altezza di questa box è la distanza interquantile. Una linea in questa box individua la mediana, permettendo di capire se la distribuzione è simmetrica o meno. Due linee, noti come "baffi", identifica un range di dati al di fuori del quartile basso e alto. Valori estremi che differiscono dal resto dei dati, noti come **outliers**, vengono rappresentati come singoli punti oltre i baffi del box plot.

Informazioni come media e varianza non possono essere estratte. Il grafico infatti ci permette di dedurre la forma. Media e varianza invece possono essere calcolabili a partire dall'istogramma.

## P-P plot o Q-Q plot

Il plot permette di confrontare i dati con una distribuzione già nota.  Nel caso di un P-P plot, si confrontano le proabilità empiriche con le teoriche; nel caso del Q-Q plot, si confrontano i quantili empirici con quelli teorici.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

# 12 - Statistica inferenziale

La statistica inferenziale è l'analisi del campione per trarre informazioni sulla popolazione.
Sia dato un campione $X_i$ di $n$ variabili aleatorie estratte da una popolazione di media $\mu$ e varianza $\sigma^2$.
Sia $f(x)$ la sua distribuzione.

Ci interessa trovare uno **stimatore** di $\mu$ e $\sigma^2$ per media e varianza. Infatti possiamo ad esempio conoscere la distribuzione $f(x)$, ma non sapere media e varianza. Ci chiediamo allora se è possibile dedurle analizzando un campione.

## Stimatore

Uno **stimatore** è un'opportuna funzione delle $X_i$ per stimare i valori di $\mu$ e $\sigma^2$. Generalmente una funzione di variabili aleatorie prende il nome di **statistica**; uno stimatore è un'opportuna statistica.
Uno **stimatore puntuale** del parametro $\theta$ è una funzione $\hat{\theta}$ che ha valori nell'immagine di $\theta$. Gli stimatori pertanto sono **variabili aleatorie**. Tale stimatore si dice **non distorto** se:

$$E[\hat{\theta}] = \theta$$

La speranza matematica dev'essere chiaramente calcolata utilizzando la densità di probabilità della variabile aleatoria, ma la densità di probabilità è funzione delle variabili aleatorie $X_i$.

### Stimatore della media

Chiamiamo $\hat{\Theta}_\mu$ lo stimatore della media. Tale stimatore è non distorto se:

$$E[\hat{\Theta}_\mu] = \mu$$

Uno stimatore della media è la media campionaria $\overline{X}_n$

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Si osserva infatti che:

$$E[\overline{X}_n] = E[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n} n \cdot \mu = \mu$$

## Stimatore della varianza

### Caso 1: media già nota

Se la media $\mu$ è già nota possiamo calcolare la varianza ricordando la formulazione della varianza empirica:

$$\overline{\sigma}^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 \cdot p_i$$

Possiamo allora costruire lo stimatore:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \mu)^2$$

Svolgendo il quadrato:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\mu X_i + \mu^2$$

Ricordando che:

$$Var(X_i) = E[X_i^2] - E[X_i]^2 \implies E[X_i^2] = E[X_i]^2 + Var(X_i)$$

Si ha:

$$E[S^2] = E[\frac{1}{n}\sum_{i=1}^{n} X_i^2] - 2\mu E[\frac{1}{n}\sum_{i=1}^{n} X_i] + E[\mu^2] =$$



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

$$= \frac{1}{n}\sum_{i=1}^{n} E[X_i^2] - 2\mu\frac{1}{n}\sum_{i=1}^{n} E[X_i] + \mu^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n} Var(X_i) + \frac{1}{n}\sum_{i=1}^{n} E[X_i]^2 - 2\mu\frac{1}{n}\sum_{i=1}^{n} \mu + \mu^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n} \sigma^2 + \frac{1}{n} n \cdot \mu^2 - 2\mu\frac{1}{n} n \cdot \mu + \mu^2 = \sigma^2$$

**Caso 2: media non nota**

Se la media non è nota possiamo comunque utilizzare la media campionaria e costruire lo stimatore sottostante:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Affinché lo stimatore non sia distorto, bisogna dividere tutto per $n-1$. Il cambiamento non viene notato se $n$ è molto grande, ma la differenza è percepita per $n$ piccolo.

Svolgendo il quadrato otteniamo:

$$\frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - 2\overline{X}_n X_i + \overline{X}_n^2$$

Osserviamo che:

$$\frac{1}{n-1}\sum_{i=1}^{n} \overline{X}_n X_i = \frac{1}{n-1}\overline{X}_n \sum_{i=1}^{n} X_i = \frac{n}{n-1}\overline{X}_n \cdot \overline{X}_n$$

Possiamo semplificare ulteriormente i termini:

$$\frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \frac{2}{n-1}\sum_{i=1}^{n} X_i\overline{X}_n + \frac{1}{n-1}\sum_{i=1}^{n} \overline{X}_n^2 =$$

$$= \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \frac{2n}{n-1}\overline{X}_n^2 + \frac{n}{n-1}\overline{X}_n^2 =$$



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

$$= \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \frac{n}{n-1}\overline{X}_n^2$$

Si ottiene che la speranza matematica:

$$E[S^2] = \frac{1}{n-1}\sum_{i=1}^{n} E[X_i^2] - \frac{n}{n-1}E[\overline{X}_n^2] =$$

$$= \frac{1}{n-1}\sum_{i=1}^{n} Var(X_i) + E[X_i]^2 - \frac{n}{n-1}(Var(\overline{X}_n) + E[\overline{X}_n]^2) =$$

$$= \frac{n}{n-1}\sigma^2 + \frac{n}{n-1}\mu^2 - \frac{1}{n-1}\sigma^2 - \frac{n}{n-1}\mu^2 = \sigma^2$$

### Uso degli stimatori

Una volta raccolti i dati del nostro campione, li utilizziamo per calcolare gli stimatori e trarre informazioni sulla popolazione. Per $n \to \infty$ si ottengono i valori di media e varianza della distribuzione, pertanto più grande è il campione più accurati saranno gli stimatori.

## Intervalli di confidenza

Invece di stimare i parametri puntualmente è possibile definire un **intervallo di confidenza** per ammettere che la nostra stima sia corretta con un errore percentuale $\alpha\%$.
Sia $\alpha \in [0,1]$. Sia $X_1, ..., X_n$ un campione. Consideriamo l'intervallo $[I_1, I_2]$. tale intervallo si dice **intervallo con livello di confidenza** $1-\alpha$ per il parametro se la probabilità di trovare il parametro nell'inetrvallo è pari a $1-\alpha$.

Supponiamo di voler trovare la media. Sappiamo che uno stimatore della media è $\overline{X}_n$, infatti $E[\overline{X}_n] = \mu$. Vogliamo pertanto costruire l'intervallo:

$$I_\mu = [\overline{X}_n - \epsilon, \overline{X}_n + \epsilon]$$

Tale intervallo è costruito in modo tale che $\mu \in I_\mu$ con probabilità $1-\alpha$.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

Cerchiamo di stabilire come calcolare $\epsilon$. Assumiamo per ora di conoscere il parametro $\sigma^2$. Possiamo costruire la variabile aleatoria:

$$\mathcal{Z}_n = \sqrt{n} \cdot \frac{\overline{X}_n - \mu}{\sigma} \xrightarrow{\mathcal{L}} N(0,1)$$

Tale risultato deriva dal teorema del limite centrale, per cui se $n \to \infty$ si osserva che la variabile aleatoria standardizzata segue una distribuzione normale di media $0$ e varianza $1$.

Per la definizione di funzione di ripartizione, si osserva che la probabilità che $\mathcal{Z}_n$ sia tra il quantile di ordine $\frac{\alpha}{2}$ e il quantile di ordine $1 - \frac{\alpha}{2}$ è uguale a $1 - \alpha$:

$$P(\phi_{\frac{\alpha}{2}} \leq \mathcal{Z}_n \leq \phi_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

Ricordando che i quantili di una distribuzione normale sono simmetrici, ossia $\phi_\alpha = -\phi_{1-\alpha}$:

$$P(-\phi_{1-\frac{\alpha}{2}} \leq \mathcal{Z}_n \leq \phi_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

Sostituendo a $\mathcal{Z}_n$ la sua espressione:

$$P(-\phi_{1-\frac{\alpha}{2}} \leq \sqrt{n} \cdot \frac{\overline{X}_n - \mu}{\sigma} \leq \phi_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

e isolando $\mu$:

$$P(\overline{X}_n - \frac{\sigma}{\sqrt{n}} \cdot \phi_{1-\frac{\alpha}{2}} \leq \mu \leq \overline{X}_n + \frac{\sigma}{\sqrt{n}} \cdot \phi_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

Abbiamo pertanto trovato l'intervallo di confidenza della media: $I_1$ e $I_2$ sono gli estremi della disuguaglianza trovata. Da notare come questo intervallo sia simmetrico. Infatti se $\epsilon = \frac{\sigma}{\sqrt{n}} \cdot \phi_{1-\frac{\alpha}{2}}$ si ottiene che l'intervallo è:

$$I_\mu = [\overline{X}_n - \epsilon, \overline{X}_n + \epsilon]$$

Ossia è l'intorno di $\overline{X}_n$ di raggio $\epsilon$.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

### Caso in cui la varianza non sia nota

Se la varianza non è nota possiamo comunque stimarla utilizzando lo stimatore $S_n^2$ costruito precedentemente. Si può infatti costruire la variabile aleatoria standardizzata:

$$T_n = \frac{\overline{X}_n - \mu}{S_n}\sqrt{n}$$

dove $S_n$ è la radice quadrata dello stimatore della varianza. Si ha che per $n$ sufficientemente grande o se il campione è gaussiano:

$$T_n \sim \mathcal{T}(n-1)$$

Ossia $T_n$ segue una distribuzione $\mathcal{T}$ di Student a $n-1$ gradi di libertà. Il ragionamento precedente può essere ripetuto per ottenere l'intervallo di confidenza:

$$I_\mu^{S_n} = [\overline{X}_n - \frac{S_n}{\sqrt{n}} t_{1-\frac{\alpha}{2}}, \overline{X}_n + \frac{S_n}{\sqrt{n}} t_{1-\frac{\alpha}{2}}]$$

### Intervallo di confidenza della varianza

Sotto l'ipotesi che il campione sia **strettamente gaussiano**, possiamo costruire la variabile aleatoria sottostante per determinare l'intervallo di confidenza della varianza:

$$\mathcal{W} = \frac{S_n^2}{\sigma^2} \cdot (n-1) \sim \chi^2(n-1)$$

L'intervallo di confidenza sarà:

$$E_\sigma = [\frac{S^2 \cdot (n-1)}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}, \frac{S^2 \cdot (n-1)}{\chi^2_{\frac{\alpha}{2}}(n-1)}]$$

La distribuzione del $\chi^2$ non è simmetrica, pertanto non possiamo dire che $\chi^2_\alpha(n-1) = -\chi^2_{1-\alpha}(n-1)$



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

# 13 - Test d'ipotesi

Data una distribuzione, possiamo fare ipotesi su un parametro $\theta$. Esistono due tipi di ipotesi:

- **ipotesi nulla** $H_0 : \theta \in \Theta_0$, ossia $\theta$ è un determinato valore appartenente all'insieme di valori $\Theta_0$. Tale ipotesi è detta **ipotesi nulla semplice** se $\Theta_0$ è un singoletto $\Theta_0 = \{\theta_0\}$. Analizzeremo principalmente test con ipotesi nulla semplice;
- **ipotesi alternativa** $H_1 : \theta \in \Theta_1$. Nel caso di un'ipotesi nulla semplice, l'ipotesi alternativa è $H_1 : \theta \neq \theta_0$

Il metodo statistico che testa un'ipotesi statistica avanzata sul parametro $\theta$ da studiare prende il nome di **test d'ipotesi**. L'idea è quella di specificare un'ipotesi falsificabile, ritenuta vera fino a prova contraria, e di utilizzare un test per verificare se i dati smentiscono tale ipotesi. L'importante è che $H_0$ e $H_1$ non abbiano un **ruolo simmetrico**. Possiamo trovare una condizione tale che è possibile **rigettare**, ossia rifiutare, l'ipotesi $H_0$; in tal caso allora possiamo affermare che $H_1$ è vera. Ma se $H_0$ non può essere rigettata, allora il test non è concludente e non si può determinare la veridicità di nessuna delle due ipotesi.

Il processo che si segue pertanto può essere diviso nei seguenti passi:

1. si preleva un campione di dati;
2. si calcola uno stimatore $\hat{\theta}$ del parametro su cui si considera l'ipotesi nulla;
3. si definisce una **zona di rigetto** $C_R$;
4. se $\hat{\theta}$ appartiene alla zona di rigetto, si rifiuta l'ipotesi nulla e si accetta quella alternativa. In caso contrario, il test non è concludente.

### Errori nel test d'ipotesi

Il test d'ipotesi può essere soggetto a errori. Definiamo diversi tipi di errori:

- **errore di prima specie** $\alpha$. Si definisce $\alpha$ come la probabilità di **rigettare** $H_0$ a torto (ossia rigettare erroneamente $H_0$ quando $\hat{\theta} = \theta_0$). La probabilità $\alpha$ prende il nome di **livello di significatività** ed è tale che $\alpha = P(\hat{\theta} \in C_R | \theta = \theta_0)$. Questo livello pertanto esprime la **probabilità**



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

**massima di accettare un errore di tipo falso negativo**, in quanto si rifiuta un'ipotesi che in realtà è vera;

- **errore di seconda specie** $\beta$. Si definisce $\beta$ come la probabilità di **non rigettare** $H_0$ se falso. Ossia $\beta = P(\hat{\theta} \notin C_R | \theta \in \Theta_1)$ (o analogamente $\theta \neq \theta_0$ quando l'ipotesi nulla è semplice). Questo è un **falso positivo**, ossia si accetta un'ipotesi nulla che in realtà è falsa.

Si definisce $p-$**value** (o $\pi$-value) il più piccolo $\alpha$ livello di significatività per cui si rigetta l'ipotesi nulla, ossia qualsiasi $p \leq \alpha$ rigetta $H_0$. Ossia, il valore $p$ è **la probabilità, sotto ipotesi nulla, di ottenere risultati compatibili con quelli osservati**.

#### Significato logico del $p-$value

Il valore $p$ è la probabilità, supposto che sia vera una data ipotesi, che un risultato simile o più estremo accada. Questo significa che sotto l'ipotesi $H_0$ la probabilità di osservare i dati che abbiamo a disposizione è pari a $p$. $\alpha$ è definito come la probabilità di rifiutare $H_0$ quando in realtà è vera, quindi è **la soglia che definisce quanto estremo dev'essere un risultato per considerarlo incompatibile con l'ipotesi**. Fissare $\alpha = 0.05$ significa fissare un livello di probabilità per stabilire quanto devono essere rari i dati per dire che sono improbabili rispetto all'ipotesi. Quindi tale valore ci aiuta a capire se la differenza tra il risultato osservato e quello ipotizzato è dovuta ad una significativa differenza statistica o alla casualità del campionamento.

### Zona di rigetto

Determiniamo adesso la zona di rigetto $C_R$. Immaginiamo di avere come ipotesi nulla l'ipotesi nulla semplice:

$$H_0 : \theta = \theta_0$$

Possiamo sottoporre la distribuzione a test a due o una coda. Identifichiamo le **code destra e sinistra** con i quantili $q_{\frac{\alpha}{2}}$ e $q_{1-\frac{\alpha}{2}}$. Le code sottendono un'area $1 - \alpha$, ossia commettiamo un errore $\alpha$. Utilizzando tali code possiamo identificare tre tipi di test:

1. **test bilatero** $H_1 : \theta \neq \theta_0$. Fissato il livello di significatività $\alpha$, rigettiamo l'ipotesi nulla quando, per un opportuno valore $\delta$, si verifica che $|\hat{\theta} - \theta_0| > \delta$, ossia se la distanza tra il valore stimato e il valore dell'ipotesi nulla è superiore a $\delta$. Il valore $\delta$ è scelto il modo tale che la probabilità che ciò accada sia uguale al livello di significatività $\alpha$, ossia se $P(|\hat{\theta} - \theta_0| > \delta) = \alpha$;



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

2. **test unilatero a sinistra**. In questo caso rigettiamo l'ipotesi nulla se $\hat{\theta} < \theta_0 - \delta$, ossia se il valore stimato è a sinistra dell'intorno centrato in $\theta_0$ e di raggio $\delta$. Anche in questo caso $\delta$ è scelto affinchè $P(\hat{\theta} < \theta_0 - \delta) = \alpha$;
3. **test unilatero a destra**. Questo test è identico a quello precedente, tuttavia il valore stimato è a destra dell'intorno centrato in $\theta_0$ e di raggio $\delta$. Si ha allora che $\delta$ è scelto affinchè $P(\hat{\theta} > \theta_0 + \delta) = \alpha$.

Rimane solamente la determinazione del valore $\delta$. Per fare ciò, possiamo ricorrere a variabili aleatorie ausiliarie di cui conosciamo già la distribuzione. Vediamo allora tre possibili test:

- $\mathcal{Z}-$test, basato su variabile aleatoria $\mathcal{Z} \sim N(0,1)$;
- $\mathcal{T}-$test, basato su variabile aleatoria $\mathcal{T} \sim T(n-1)$;
- $\chi^2-$test, basato su variabile aleatoria $\chi^2(n-1)$.

### $\mathcal{Z}-$test

Costruiamo una variabile aleatoria che segue la distribuzione $N(0,1)$.

Consideriamo la variabile aleatoria:

$$\mathcal{Z}_n = \frac{\overline{X}_n - \mu_0}{\sigma}\sqrt{n}$$

e come ipotesi nulla:

$$H_0 : \mu = \mu_0$$

La zona di rigetto da determinare per un test bilatero dev'essere tale da rispettare la sottostante condizione:

$$P(|\overline{X}_n - \mu| > \delta) = \alpha$$

Possiamo moltiplicare per $\frac{\sqrt{n}}{\sigma}$ e ottenere:

$$P(|\mathcal{Z}_n| > \frac{\delta\sqrt{n}}{\sigma}) = \alpha$$

Osserviamo che poiché deve accadere che la probabilità:



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

$$P(|\overline{X}_n - \mu| > \delta) = \alpha$$

Basta considerare i quantili $-\phi_{1-\frac{\alpha}{2}}$ e $\phi_{1-\frac{\alpha}{2}}$. Si ottiene allora che $\phi_{1-\frac{\alpha}{2}}$ è uguale al valore $\frac{\delta\sqrt{n}}{\sigma}$. Le zone di rigetto per il test bilatero sono:

$$\mathcal{Z}_n > \phi_{1-\frac{\alpha}{2}} \quad \mathcal{Z}_n < -\phi_{1-\frac{\alpha}{2}}$$

Infatti desideriamo che la probabilità di commettere l'errore pari ad $\alpha$.

Isolando la media campionaria:

$$\overline{X}_n > \mu + \frac{\sigma}{\sqrt{n}}\phi_{1-\frac{\alpha}{2}} \quad \overline{X}_n < \mu - \frac{\sigma}{\sqrt{n}}\phi_{1-\frac{\alpha}{2}}$$

Per il test unilaterale a sinistra, abbiamo l'ipotesi alternativa $H_1 : \mu < \mu_0$, per cui basta:

$$\mathcal{Z}_n < \phi_\alpha$$

Ossia:

$$\overline{X}_n < \mu_0 + \frac{\sigma}{\sqrt{n}}\phi_\alpha$$

Per il test unilaterale a destra:

$$\mathcal{Z}_n > \phi_{1-\alpha}$$

cioè:

$$\overline{X}_n > \mu_0 + \frac{\sigma}{\sqrt{n}}\phi_{1-\alpha}$$

Per il $p-$value, basta osservare che il minimo valore $\alpha$ è calcolabile tramite la funzione di ripartizione. Si osserva allora:

- $p = 2(1 - F_{\mathcal{Z}}(|\mathcal{Z}_n|))$ nel caso bilatero;
- $p = F_{\mathcal{Z}}(\mathcal{Z}_n)$ nel caso unilatero sinistro;



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

- $p = 1 - F_{\mathcal{Z}}(\mathcal{Z}_n)$ nel caso unilatero destro.

Lo $Z-$test viene anche utilizzato in altri due casi:

1. per stimare la probabilità di una distribuzione di Bernoulli e verificare che sia diversa da $p_0$. In tal caso consideriamo come ipotesi nulla $H_0 : p = p_0$ e si considera la variabile $\mathcal{Z}_n = \frac{\overline{X}_n - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} \sim N(0,1)$;
2. per verificare se le medie di due popolazioni indipendenti con varianze note sono uguali. L'ipotesi nulla è $H_0 : \mu_x = \mu_y$ e la variabile ausiliaria è $\mathcal{Z}_n = \frac{\overline{X}_n - \mu_X - (\overline{Y}_m - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim N(0,1)$.

### $\mathcal{T}-$test

Si costruisce una variabile che segue una distribuzione $\mathcal{T}$ di Student. Questo è utile quando la varianza della distribuzione **non** è nota.

La variabile aleatoria ausiliaria che costruiamo è:

$$T_n = \frac{\overline{X}_n - \mu}{S_n}\sqrt{n} \sim \mathcal{T}(n-1)$$

I risultati per le zone di rigetto e i $p-$value sono identici a quelli precedenti, tuttavia considerando i quantili e la funzione di ripartizione della $\mathcal{T}$ di Student.

Il $\mathcal{T}-$test è anche usato per la statistica test di due popolazioni accoppiate $(X_i, Y_i)$, di medie $\mu_X, \mu_Y$ e varianze $\sigma_X^2, \sigma_Y^2$. Data $D_i = X_i - Y_i$, assumendo che $D_i \sim N(\mu_d, \sigma_D^2)$ tale che $\mu_d = \mu_X - \mu_Y$ e $\sigma_D^2$ non nota, possiamo stimare quest'ultima con la varianza campionaria. L'ipotesi nulla è $H_0 : \mu_D = 0$. La variabile da costruire è:

$$T_n = \frac{\overline{D}_n}{S_n}\sqrt{n}$$

### $\chi-$test

Si costruisce una variabile che segue una distribuzione del $\chi^2$. Questo è utile quando vogliamo testare se la varianza è uguale ad una nota $\sigma_0^2$ (ossia, l'ipotesi nulla è $H_0 : \sigma = \sigma_0$).

Si costruisce la variabile aleatoria ausiliaria:



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

$$\mathcal{W}_n = \frac{S_n^2}{\sigma_0^2} \cdot (n-1) \sim \mathcal{T}(n-1)$$

E si osserva come zona di rigetto:

- $\mathcal{W}_n > \chi^2_{1-\frac{\alpha}{2}}(n-1) \vee \mathcal{W}_n < \chi^2_{\frac{\alpha}{2}}(n-1)$ per il test bilatero;
- $\mathcal{W}_n < \chi^2_{\alpha}(n-1)$ per il test unilatero a sinistra;
- $\mathcal{W}_n > \chi^2_{1-\alpha}$ per il test unilatero a destra.

Il $\chi-$test è utile quando vogliamo testare se un vettore di probabilità di una legge multinomiale è pari ad un vettore noto $p^0 = \{p_1^0, p_2^0, ..., p_m^0\}$ di dimensione $m$, ovvero l'ipotesi nulla è $H_0 : p = p_0$. In tal caso la variabile aleatoria ausiliaria che utilizziamo è:

$$\mathcal{W}_n = n \sum_{k=1}^{m} \frac{(\overline{p}_k - p_k^0)^2}{p_k^0} \sim \chi^2(m-1)$$

dove $\overline{p}_k$ è la probabilità empirica di $k$, ossia:

$$\overline{p}_k = \frac{n_k}{n}$$

Questo test è noto come **test di Pearson**. Da notare come per questo test, i gradi di libertà siano $m-1$, ossia il numero di esiti meno uno.

### Esempio di test: lancio di una moneta

Supponiamo di avere una moneta e di voler verificare che tale moneta sia sbilanciata. Allora eseguiamo 20 lanci e contiamo il numero di volte in cui esce testa.
Idealmente se la moneta è perfettamente bilanciata dovremmo ottenere un numero di teste vicino al 10. L'esperienza tuttavia ci dice che anche se la moneta è bilanciata, possiamo anche ottenere 14,15 o addirittura 20 volte testa. Supponiamo allora di aver osservato 15 esiti "testa".

- Quanto è distante il risultato dal valore medio della variabile aleatoria che tiene conto del fenomeno?
- La distanza dalla media è sufficiente per rigettare l'ipotesi che la moneta sia bilanciata?



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

La variabile aleatoria discreta $X$ che tiene conto di tale fenomeno è tale che si osserva:

$$X \sim B(20, p)$$

Definiamo allora l'ipotesi nulla e l'ipotesi alternativa:

- l'ipotesi nulla è che la moneta è bilanciata. Tale ipotesi nulla è l'ipotesi nulla semplice $H_0 : p = 0.5$;
- l'ipotesi alternativa è che la moneta non è bilanciata, ossia $H_1 : p \neq 0.5$.

Scegliamo il livello di significatività $\alpha = 0.05$, valore tipicamente utilizzato per test non critici. Rigettiamo allora l'ipotesi che la moneta sia bilanciata considerando il $p-$value.

Possiamo ottenere un $p-$value preciso utilizzando la variabile aleatoria $X$ e calcolando la probabilità di ottenere un numero superiore alle 15 teste o inferiore alle 5 teste. Infatti per valutare la distanza tra il valore sperimentale e quello teorico, basta considerare la probabilità di ottenere un valore maggiore o uguale a 15. Consideriamo anche il caso simmetrico, poiché stiamo facendo un test su un valore medio. Questo coincide con il calcolo del $p-$value di un test bilatero:

$$p - \text{value} = P(X \geq 15) + P(X \leq 5) \approx 0.041$$

Il $p-$ value ottenuto esprime la probabilità di ottenere il risultato sperimentale o uno addirittura più estremo **ammesso che la moneta sia bilanciata**. La probabilità è in questo caso pari a $4.1\%$. Questa probabilità è tuttavia bassa rispetto al livello di significatività che abbiamo prefissato. Ossia, **assumere che la moneta sia bilanciata non spiega in maniera adeguata quanto osservato**. Rigettando l'ipotesi nulla, concludiamo dicendo che **la moneta non è bilanciata**.

Da notare che se avessimo scelto un valore $\alpha = 0.01$, il test sarebbe stato **inconcludente**: l'ipotesi potrebbe spiegare i dati, **ma non abbiamo nulla che ci permette di garantire ciò**.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

# 14 - Generazione di numeri pseudo-casuali

Un numero si dice **pseudo-casuale** se è dotato di una certa aleatorietà all'interno di un certo periodo dopo il quale la sequenza si ripete. Il vantaggio nell'utilizzo di numeri pseudo-casuali risiede nella rapidità del calcolo, tuttavia è necessario avere un periodo molto lungo: la generazione del numero si basa su una sequenza di inizio che prende il nome di **seme** (**seed**).

Ci interessa generare numeri **uniformi** in $[0, 1]$. Possiamo utilizzare il metodo delle **congruenze lineari**:

$$X_{n+1} = (a \cdot X_n + c) \mod m$$

La relazione è molto semplice, ma permette di generare numeri anche molto grandi. I parametri del **generatore lineare** sono i seguenti:

- $a$ è il **moltiplicatore**;
- $c$ è l'**incremento**;
- $m$ è il **modulo**;
- $X_0$ è il **seme**.

Si dovrebbero garantire le proprietà:

- $c, m$ sono **primi tra loro**;
- $a - 1$ è **divisibile per ogni fattore primo di** $m$;
- $a - 1$ è **un multiplo di 4 se** $m$ **è multiplo di 4**.

Lavorando con i valori di $a, c, m$ si possono generare numeri pseudo-casuali con periodo molto grande, arrivando fino a $T = 2^{256}$. L'algoritmo Mersenne-Twister arriva ad un periodo $T = 2^{19937} - 1$.

In generale, con il generatore lineare possiamo generare numeri abbastanza random nell'intervallo $[0, 1]$. Vogliamo utilizzare questi numeri random per:

- generare numeri nelle distribuzioni binomiale e multinomiale;

- metodo di inversione della funzione di ripartizione;
- metodo del rigetto.

### Generazione di numeri in distribuzioni binomiale e multinomiale.

Generiamo un numero $\eta$ nel range $[0, 1]$. Se $\eta \leq p$, allora $x = 1$, mentre se $\eta > p$, $x = 0$.

La logica che seguiamo è che dire che il numero random è tra zero e $p$ significa dire che ho un successo. Se $p$ è molto grande, ho chiaramente più esiti $x = 1$ e quindi più successi. La probabilità $p$ identifica quindi la dimensione dell'intervallo di sucesso. Questo ci permette di generare $X \sim B(1, p)$. Per generare $X \sim B(n, p)$, genero $n$ $X_i \sim B(1, p)$ e si sommano. Quindi:

$$X = \sum_{i=1}^{n} X_i$$

Per una multinomiale, poiché abbiamo $n$ esiti è necessario dividere l'intervallo $[0, 1]$ in $n$ intervalli in modo tale che la lunghezza di ogni intervallo sia proporzionale alla probabilità di ciascun esito. Possiamo seguire i seguenti step:

1. introduciamo $F_i = \sum_{k=1}^{i} p_k$, dove $p_k$ è la probabilità del $k-$esimo esito;
2. generiamo $\eta \sim U([0, 1])$. Si ha che:
   - $\eta \in [0, F_1] \implies i = 1$;
   - $\eta \in ]F_1, F_2] \implies i = 2$;
   - …
   - $\eta \in ]F_{n-1}, F_n = 1] \implies i = n$;
3. poniamo $X_j = \begin{cases} 1 & \text{se } j = i \\ 0 & \text{altrimenti} \end{cases}$;
4. poniamo $X = (X_1, ..., X_n) \sim B(1, p_1, ..., p_n)$.

Se vogliamo $Y \sim B(n, p_1, ..., p_n)$, basta generare $n$ volte $X$ e sommarle:

$$Y = \sum_{i=1}^{N} X_i$$



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

## Metodo di inversione della funzione di ripartizione

Il metodo di inversione della funzione di ripartizione si basa sulla proprietà che se la funzione di ripartizione $F$ è invertibile e abbiamo $X \sim U([0,1])$, allora $F^{-1}(X)$ ha funzione di ripartizione $F$.

Sappiamo che:

$$F(t) = \int_{-\infty}^{t} f(x)\,dx$$

Dire che $F^{-1}(X)$ ha funzione di ripartizione $F$ significa dire che la sua distribuzione di probabilità è $f(x)$. Pertanto se calcolo $F^{-1}(X)$ ottengo $Y$ con la stessa distribuzione di probabilità. Questo ci permette di generare dei numeri pseudo-casuali che seguono la distribuzione di probabilità $f(x)$ semplicemente applicando l'inversa della funzione ad $X$. Poiché $X \sim U([0,1])$, allora basta sostituire a $X$ il valore $\eta$ generato in maniera pseudo-casuale.

### Esempio 1: distribuzione uniforme

Data $X \sim U([a,b])$, si ha:

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & \text{altrimenti} \end{cases}$$

$$F_X(t) = \frac{t-a}{b-a} \qquad \forall t \in [a,b]$$

Cerchiamo $F_X^{-1}$, ovvero la funzione tale che:

$$y^{-1}(F_X(t)) = t$$

Invertendo:

$$F_X(t) = \frac{t-a}{b-a} \implies (b-a)F_X(t) = t-a \implies t = a + (b-a)F_X(t)$$

Sostituiamo allora a $F_X(t)$ il valore $\eta \sim U([0,1])$. Otteniamo allora che:



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

$$Y = a + (b-a)\eta \sim U([a,b])$$

**Esempio 2: distribuzione esponenziale**

Consideriamo $X \sim \exp(\lambda)$, che ha funzione di ripartizione:

$$F_X(t) = 1 - e^{-\lambda t}$$

Si ha:

$$e^{-\lambda t} = 1 - F_X(T) \implies -\lambda t = \ln(1 - F_X(t)) \implies t = -\frac{\ln(1-F_X(t))}{\lambda}$$

Si sostituisce allora il valore $1 - F_X(T)$ con il valore randomico $\eta$. Infatti poiché $1 - F_X(t) \in [0,1]$, allora possiamo direttamente prendere $\eta$.

**Esempio 3: distribuzione normale**

Siano $X, Y \sim N(0,1)$. Si ha che $X^2 + Y^2 \sim \chi^2(2)$. Tale distribuzione è equivalente alla distribuzione $\exp(\frac{1}{2})$.

Notiamo che $X^2 + Y^2 = Z^2$ è l'equazione di una circonferenza di raggio $Z$. Possiamo allora generare un raggio (poiché abbiamo visto sopra la generazione di un valore nella distribuzione esponenziale), generare l'angolo e considerare $X$ e $Y$ ottenuti goniometricamente:

$$X = Z\cos\theta \qquad Y = Z\sin\theta$$

Allora $X, Y \sim N(0,1)$ e possiamo generarli con il seguente algoritmo:

1. generiamo $\eta_1, \eta_2 \in [0,1]$;
2. generiamo il raggio, ossia $Z^2 \sim \exp(\frac{1}{2})$. Invertendo si ha che $Z^2 = -2\ln\eta_1 \implies Z = \sqrt{-2\ln\eta_1}$. L'argomento della radice è positivo;
3. generiamo l'angolo, ossia $\theta \sim U([0, 2\pi])$ con $a = 0, b = 2\pi$. Consideriamo $\theta = 2\pi \cdot \eta_2$;
4. generiamo $X, Y$ tali che:



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

$$X = \sqrt{-2\ln\eta_1}\cdot\cos(2\pi\eta_2) \qquad Y = \sqrt{-2\ln\eta_1}\cdot\sin(\eta_2\cdot 2\pi)$$

Se si vuole generare $W \sim N(\mu,\sigma^2)$, si genera $X \sim N(0,1)$ e si pone $W = \sigma X + \mu$.

### Metodo del rigetto

Conoscendo la densità di probabilità $f(x) : [a,b] \to \mathbb{R}$. Chiamiamo $M$ il massimo della funzione in $[a,b]$ (che esiste per Weierstrass). Vogliamo allora generare $X$ tale che $X$ abbia distribuzione di probabilità $f(x)$. Utilizziamo il **metodo del rigetto**.

Il metodo del rigetto consiste nei seguenti passi:

1. generiamo $\eta \sim U([a,b])$ e $\xi \sim U([0,M])$;
2. se $\xi \leq f(\eta)$, si accetta la coppia $(\eta,\xi)$ e si pone $X = \eta$.

Si può dimostrare che $X$ ha densità $f(x)$.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

# 15 - Metodi Monte Carlo

I **metodi Monte Carlo** sono tecniche che utilizzano variabili aleatorie per risolvere problemi deterministici difficili da risolvere analiticamente. Ad esempio, il calcolo di un integrale potrebbe essere difficile, ma è possibile utilizzare il **metodo hit or miss**.

Solitamente i metodi montecarlo vengono impiegati nel momento in cui bisogna scoprire lo stato finale di un sistema. Se il problema è computazionalmente pesante, si può scrivere una sequenza di variabili stocastiche che ci permettono di approssimare lo stato finale se il numero è sufficientemente grande. Essenzialmente l'evento stocastico viene ripetuto più volte e si analizza la media.

## Metodo hit or miss

Il metodo hit or miss è un metodo Monte Carlo per il calcolo di integrali definiti. Le ipotesi del metodo sono le seguenti:

1. $f(x) \geq 0$;
2. $f(x)$ continua $\forall x \in [a, b] \subset \mathbb{R}$.

L'obiettivo è calcolare l'integrale:

$$\int_a^b f(x)\, dx$$

Chiaramente osserviamo che l'integrale è l'area sottesa alla curva $f(x)$. Preso un punto nel rettangolo di base $b-a$ e altezza $M = \max_{x \in [a,b]} f(x)$, si ha che la probabilità che tale punto appartenga al trapezoide è pari al rapporto:

$$p = \frac{\displaystyle\int_a^b f(x)\, dx}{M \cdot (b-a)}$$

Otteniamo allora che l'integrale può essere calcolato moltiplicando la stima $p$ per l'area del rettangolo:

$$p \cdot M \cdot (b-a) = \int_a^b f(x)\,dx$$

Per cui possiamo generare $N$ punti nel rettangolo $[a,b] \times [0,M]$. Di questi $N$ punti, solo $N_s$ saranno nel trapezoide. Approssimiamo allora $p$ al rapporto:

$$p_N = \frac{N_s}{N}$$

Otteniamo che:

$$p_N = p \quad N \to \infty$$

Quindi si ha:

$$I = \frac{N_s}{N} M \cdot (b-a)$$

Il numero $N$ dev'essere sufficientemente grande affinché $p_N = p$. Inoltre, se il trapezoide ricopre maggiormente l'area del rettangolo $[a,b] \times [0,M]$, allora la convergenza sarà più rapida.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

# 16 - Catene di Markov

## Processi stocastici

I processi stocastici sono una famiglia di variabili aleatorie. Ad ogni variabile associamo un pedice temporale $t$. Indichiamo tale famiglia con $(X_t)_{t\in T}$.

Se $T$ è discreto, il processo stocastico è **a tempi discreti**, altrimenti è **a tempi continui**.

I valori assunti da $X_t$ appartengono all'insieme $E$. Tale insieme è detto **insieme degli stati**. Ad ogni istante di tempo possiamo valutare la probabilità di trovarci in un certo stato, sapendo la storia precedente del processo:

$$P(X_{t+1} = j|X_t = i_t, X_{t-1} = i_{t-1}, ..., X_1 = i_1) \qquad i, ..., i_{t-1}, j \in E$$

## Catene di Markov

Si definisce **catena di Markov** un processo stocastico **a stati discreti** che gode della proprietà di assenza di memoria, nota anche come proprietà di Markov:

$$P(X_{t+1} = j|X_t = i, X_{t-1} = i_{t-1}..., X_1 = i_1) = P(X_{t+1} = j|X_t = i) = p_{ij}(t)$$

$p_{ij}$ è la probabilità che da $i$ ci si sposti in $j$. Se $p_{ij}(t)$ non cambia col tempo, la catena di Markov si dice **omogenea**. In questo caso:

$$p_{ij}(t) = p_{ij}$$

Lavoreremo sempre con catene di Markov omogenee e con $E$ finito.

Se la catena di Markov è omogenea possiamo costruire la **matrice di transizione**:

$$P = \begin{pmatrix} P_{00} & P_{01} & ... & P_{0N} \\ ... & ... & ... & ... \\ P_{N0} & P_{N1} & ... & P_{NN} \end{pmatrix}$$

L'elemento $P_{ij}$ esprime la probabilità di spostarsi dallo stato $i$ allo stato $j$. Ogni elemento della matrice gode delle proprietà:

- $P_{ij} \geq 0 \quad \forall i, j \in E$;
- $\sum_{j \in E} P_{ij} = 1 \quad \forall i \in E$.

Può accadere che anche $\sum_{i \in E} P_{ij} = 1 \quad \forall j \in E$. In tal caso si dice che $P$ è **bistocastica**.

Data $P$, si può conoscere la probabilità di andare da $i$ a $j$ in $m$ passi. Indichiamo tale probabilità con $p_{ij}^{(m)}$ e conserviamo l'informazione nella matrice $P^{(m)}$. Da notare che questa probabilità **non dipende dai passi precedenti**, ma possiamo prevedere in qualche modo cosa succede tra $m$ passi.

Definiamo $p_{ij}^{(m)}$ come la probabilità:

$$p_{ij}^{(m)} = P(X_{t+m} = j | X_t = i)$$

Si può dimostrare:

$$P^{(m)} = P^m = P \cdot P \cdot ... \cdot P$$

Questo corrisponde alla computazione dei percorsi di lunghezza $m$ avendo la rappresentazione di un grafo tramite matrice di adiacenza.

### Matrice e grafo

Avendo la matrice $P$, possiamo costruire un grafo in maniera molto semplice. La matrice $P$ sarà di dimensione $n \times n$, pertanto il nostro grafo avrà $n$ nodi. Se $P_{ij} \neq 0$, allora esiste l'arco $(i, j)$. Dal grafo, di conseguenza, possiamo costruire la matrice seguendo la medesima regola.

Diremo che due stati sono **comunicanti** se vale la condizione:

$$\exists n \in \mathbb{N} | P_{ij}^{(n)} > 0$$



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

Ciò si indica con $i \to j$; se non sono comunicanti, si usa la notazione $i \nrightarrow j$. Notiamo alcuni aspetti:

1. se $i \to j$ non è detto che $j \to i$, cioè **la proprietà non è simmetrica**;
2. se $i \to j, j \to h$, allroa $i \to h$, cioè **la proprietà è transitiva**;

Se tutti gli stati sono comunicanti, la catena si dice **irriducibile**. Diremo invece che $C$ è una **classe chiusa** se gli stati appartenenti a $C$ non comunicano con gli stati in $E \setminus C$. Una classe si dice irriducibile se tutti i suoi stati comunicano.

Per individuare facilmente le classi chiuse, basta notare i cicli nella rappresentazione a grafo tali che i nodi del ciclo non hanno archi uscenti al di fuori dei nodi del ciclo.

Quando la classe chiusa ha un elemento solo, tale elemento si dice **stato assorbente**. Tale stato indica il termine di una certa evoluzione.
Uno stato, invece, si dice **ricorrente** se si ha probabilità $1$ di ritornarci; altrimenti, si dirà **transitorio**.

### Legge della catena

La **legge** o **distribuzione** della catena di Markov permette di determinare la probabilità ad ogni istante di tempo di trovarsi in un dato stato $i \in E$. Se $X_0$ è lo stato all'istante iniziale $t = 0$ e $v$ è la sua legge:

$$v = (v_1, v_2, ..., v_n) \quad n = |E|$$

tale che:

- $v_k = P(X_0 = k)$;
- $\sum_{k \in E} v_k = 1$

Indichiamo con $w$ la legge al passo $n$, ossia $w = (w_1, ..., w_n)$ e $w_k = P(X_n = k)$. Si può dimostrare:

$$w = v \cdot P^n$$

Pertanto nota la matrice di transizione è molto semplice costruire la legge $w$.
Se calcolassimo il vettore $w \cdot P$, troveremo comunque $w$. Capiamo allora che il processo può diventare **stazionario**. Definiamo allora la **legge stazionaria** o **distribuzione stazionaria** come una legge che non evolve più:

$$v = v \cdot P$$



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 

Se ciò accade:

$$v \cdot P^n = v \cdot P = v$$

### Teorema di Markov-Katukani

Si osserva che se $E$ è finito, allora esiste **almeno** una distribuzione stazionaria.

### Teorema di Markov

Se $P$ è **regolare**, ossia:

$$\exists n \in \mathbb{N} | \, P_{ij}^n > 0 \quad \forall i, j \in E$$

ed $E$ è finito, allora esiste **una ed una sola** distribuzione stazionaria $\pi$ e vale:

$$\pi_j = \lim_{n \to +\infty} P_{ij}^{(n)}$$

Da questa distribuzione **non possiamo più tornare indietro**, pertanto non è possibile esaminare gli step precedenti. In maniera algoritmica è possibile moltiplicare $P$ molte volte per ottenere $\pi$. Alternativamente, si può utilizzare un processo analitico: si trova la distribuzione stazionaria come autovettore sinistro di $P$ con autovalore $1$. Basta allora diagonalizzare la trasposta di $P$ e verificare se esistono autovalori pari ad $1$. Nel caso in cui ci fosse più di un autovalore pari ad 1, ci sono più stati stazionari: sulla base della condizione iniziale cambia lo stato di arrivo. Se lo stato stazionario è unico, anche partendo da un'altra distribuzione di probabilità giungo lo stesso alla legge stazionaria $\pi$.

### Reversibilità

Una distribuzione $\pi$ si dice **reversibile** se vale la seguente proprietà:

$$\pi_i \cdot P_{ij} = \pi_j \cdot P_{ji} \quad \forall i, j \in E$$

Si ha che se una distribuzione è reversibile, allora **è stazionaria**.



Materiale fornito da noistudiamo.it | Contatti: info@noistudiamo.it 