

# Appunti di Metodi Matematici e Statistici

<http://rosarioterranova.blogspot.it/>

v0.71

## STATISTICA DESCRITTIVA

Branca della statistica che raccoglie metodi e strumenti matematici atti ad organizzare in modo sintetico e comprensibili informazioni e dati. Dato l'insieme di dati  $E = \{x_1, x_2, \dots, x_n\}$ , l'unità elementare su cui vengono osservati i dati è l'**unità statistica**, ovvero  $x_n \in E$ . Il numero totale di dati che si hanno a disposizione, ovvero la **numerosità** (o popolazione statistica) è indicato con  $n$ . Ciascuna unità statistica presenta un **carattere**, cioè cosa descrive il dato (Es: voto, colore occhi, ...); mentre la **modalità del carattere** è l'insieme dei valori  $S = \{s_1, s_2, \dots, s_n\}$  che un carattere può assumere (Es: modalità del carattere voto = ottimo, buono, ...; modalità del carattere colore occhi = azzurro, marrone, ...).

## Frequenza

Rappresenta il numero delle volte in cui una determinata modalità del carattere si presenta.

- **Frequenza assoluta:** indica quante volte si ripete una determinata modalità del carattere

$$f_j = n^\circ x_n \in E \text{ con valore } S_j$$

- **Frequenza relativa:** indica quante volte si ripete una determinata modalità del carattere e la divido per il totale delle modalità

$$p_j = \frac{f_j}{n}$$

Le frequenze relative sono comprese in  $[0,1]$ , moltiplicandole per 100 trovo le **frequenze percentuali**.

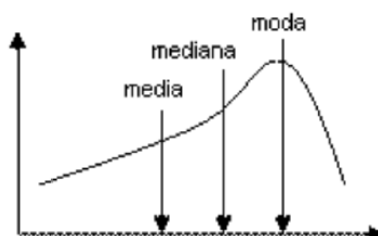
- **Frequenza cumulata assoluta:** è la somma della frequenza assoluta della modalità osservata più le modalità precedenti
- **Frequenza cumulata relativa:** è la somma della frequenza relativa della modalità osservata più le modalità precedenti

Es: sondaggio di  $n=10$  studenti per il gradimento di un professore

Voto	Frequenza assoluta	Frequenza relativa	Frequenza percentuale	Frequenza cumulata assoluta	Frequenza cumulata relativa	Frequenza cumulata percentuale
Cattivo	3	$3/10=0,3$	30	3	0,3	30
Medio	2	$2/10 = 0,2$	20	$3+2=5$	$0,3+0,2=0,5$	$30+20=50$
Buono	5	$5/10 = 0,5$	50	$5+5=10$	$0,5+0,5=1$	$50+50=100$

## Indici di tendenza centrale

Consentono di sintetizzare un insieme di misure tramite un unico valore rappresentativo.



- **Media:** data dalla somma delle misure osservate diviso la numerosità

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Formula excel =MEDIA(valori)

- **Media pesata (o ponderata):** si applica quando i valori  $x_n \in E$  non hanno tutti la stessa importanza; essa è data dalla somma delle misure osservate diviso la numerosità, in cui ogni valore viene moltiplicato per il suo peso

$$\bar{x}_p = \frac{1}{n} \sum_{i=1}^n s_i x_i$$

Formula excel =MATR.SOMMA.PRODOTTO(valori;pesi)/SOMMA(valori)

- **Mediana:** valore che occupa la posizione centrale di una sequenza ordinata. Se la sequenza ha n pari, allora si considera la media dei due elementi centrali.  
Formula excel =MEDIANA(valori)
- **Moda:** valore più frequente nella serie di dati osservata. Se essa è unica si dice unimodale, se invece è di due valori si dice bimodale.  
Formula excel =MODA(valori)
- **Quantile:** il quantile di ordine  $\alpha$ , con  $\alpha \in \mathbb{R}$  e  $\alpha \in [0,1]$ , è un valore  $Q_\alpha$  che divide la popolazione in due parti proporzionali ad  $\alpha$ , ovvero  $\text{valori} < Q_\alpha < \text{valori}$ . Se  $\alpha = 1/2$  allora stiamo parlando della mediana. Il quantile viene chiamato **percentile** quando  $\alpha$  è espressa sotto forma di percentuale.
- **Quartile:** è un percentile in cui  $\alpha$  assume valori della frazione di quattro, ovvero:
  - $Q(1)$  valore tale che almeno il 25% (1/4) dei dati è minore o uguale a questo
  - $Q(2)$  valore tale che almeno il 50% (2/4) dei dati è minore o uguale a questo
  - $Q(3)$  valore tale che almeno il 75% (3/4) dei dati è minore o uguale a questo
 Formula excel =QUARTILE(valori;quartile) permette di ottenere il quartile desiderato

#### Procedura di calcolo dei quartili

1. Fissiamo  $\alpha = 0.25, 0.50$  e  $0.75$  e calcoliamo  $\alpha(n+1)$
2. Se  $\alpha(n+1) = m \in \mathbb{N}$ , allora  $Q_\alpha = x_m$
3. Se  $\alpha(n+1) = m \notin \mathbb{N}$ , allora si usa uno dei seguenti metodi
  - a. METODO MEDIA: si prende parte intera di  $\alpha(n+1)$   $m \in \mathbb{N}$  tale che  $m < \alpha(n+1) < m+1$  e quindi si fa la media aritmetica  $Q_\alpha = \frac{x_m + x_{m+1}}{2}$
  - b. METODO INTERPOLAZIONE LINEARE: detta  $m$  la parte intera di  $\alpha(n+1)$  e  $\beta$  la sua parte frazionaria avremo  $Q_\alpha = x_m + \beta(x_{m+1} - x_m)$

Es: Siano assegnati i seguenti dati, che rappresentano le età di un campione di persone

$E = \{16, 18, 18, 19, 20, 20, 20, 20, 21, 21, 21, 22, 23, 25, 28, 30, 31, 37\}$   $n = 18$ , media = 18, mediana = 21, moda = 20

Calcolo  $\alpha = 0.25$ :

1. Calcolo  $\alpha(n+1) = 0.25(18+1) = 4.75$
2. 4.75 non è  $\in \mathbb{N}$ , quindi  $m = 4$ ,  $\beta = 0.75$
- 3a.  $Q_{0.25} = (x_4 + x_5)/2 = (19+20)/2 = 19.5$
- 3b.  $Q_{0.25} = x_4 + 0.75(x_5 - x_4) = 19 + 0.75(20-19) = 19.75$

Il calcolo  $\alpha = 0.50$  e  $0.75$  avviene come sopra. Da questa analisi possiamo concludere che il 25 % delle persone del campione hanno un'età minore o uguale a 19,5 anni

#### Indici di variabilità

Valutano in modo sintetico la variabilità della distribuzione di dati

- **Scarto:** valore dato dalla differenza tra il valore i-esimo e la media della distribuzione  
 $x_i - \bar{x}$

- **Varianza:** somma dei quadrati degli scarti fratto la numerosità

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

Minore è la varianza, maggiore è la concentrazione dei dati intorno al valore medio; maggiore è la varianza, maggiore è la dispersione dei dati intorno al valor medio

Formula excel =VAR.POP(valori)

- **Scarto quadratico medio (o deviazione standard):** è la varianza sotto radice, esprimere la dispersione dei dati intorno ad un indice di tendenza

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

Formula excel =DEV.ST(valori)

- **Scarto medio assoluto:** è la somma degli scarti fratto la numerosità

$$s.a. = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

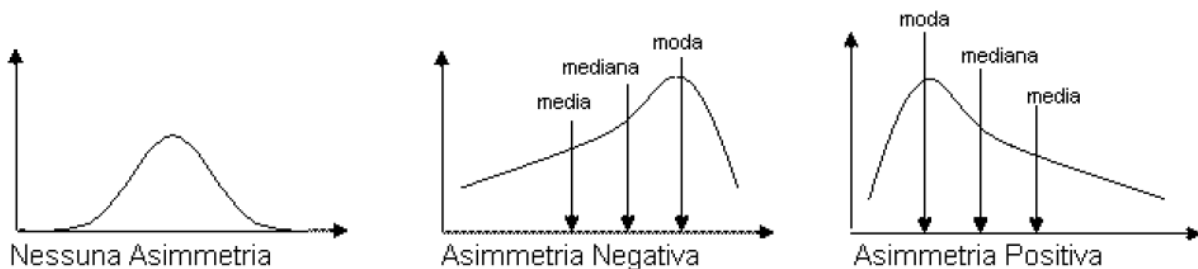
- **Coefficiente di variazione:** è il rapporto tra deviazione standard e media; un grande coefficienti di variazione indica una maggiore variabilità dei dati

$$c.v. = \frac{s}{\bar{x}}$$

### Indici di forma

Valutano la forma di una distribuzione di una serie di dati

- **Indice di simmetria:** una distribuzione è simmetrica rispetto alla mediana se le modalità che sono equidistanti dalla mediana hanno la stessa frequenza. Una distribuzione è asimmetrica se invece non è possibile individuare nell'istogramma un asse verticale che taglia la distribuzione in due parti specularmente uguali.



L'indice più usato per valutare l'asimmetria di una distribuzione è **Skewness**

$$sk = \frac{m_3}{\sqrt{m_2^3}} \quad m_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k \quad k = 2, 3, 4$$

- Se  $sk > 0$  la distribuzione ha coda più lunga a destra
- Se  $sk < 0$  la distribuzione ha coda più lunga a sinistra
- Se  $sk = 0$  la distribuzione è simmetrica rispetto alla media

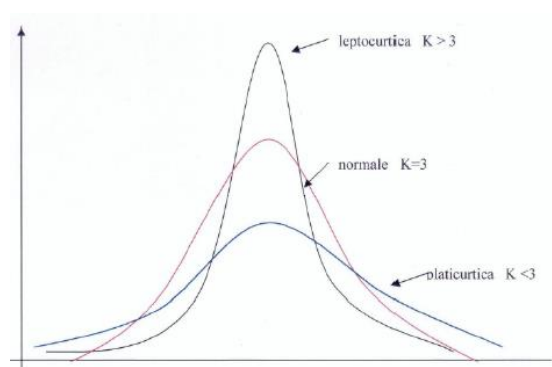
Formula excel =ASIMMETRIA(valori)

- **Curtosi:** ci dice se una distribuzione è piatta o appuntita

$$k = \frac{m_4}{m_2^2}$$

- Se  $k > 3$  la distribuzione è leptocurtica
- Se  $k < 3$  la distribuzione è platicurtica
- Se  $k = 3$  ha la stessa altezza di una normale

Formula excel =CURTOSI(valore1;valore2;...)



## Correlazione

Indica la tendenza che hanno due variabile a variare insieme, ovvero a covariare. Nella correlazione tra due serie di dati vengono confrontate le variazioni delle i-esime coppie di dati rispetto ai rispettivi valori medi mediante il prodotto

$$(x_i - \bar{x})(y_i - \bar{y})$$

Tanto più i prodotti hanno concordanza di segno, tanto più i dati hanno una forte dipendenza.

Formula excel =CORRELAZIONE(matriceA;matriceB) se il valore restituito è prossimo a 1 vuol dire che i dati sono ben correlati.

**Covarianza:** indica se due serie di dati sono correlate positivamente, negativamente o non lo sono affatto. Essa è la somma dei prodotti degli scarti

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

- Se  $C_{xy} > 0$  le due serie di dati sono correlate positivamente
- Se  $C_{xy} < 0$  le due serie di dati sono correlate negativamente
- Se  $C_{xy} = 0$  le due serie di dati sono incorrelate, e quindi **statisticamente indipendenti**

Formula excel =COVARIANZA(matriceA;matriceB)

## Analisi di regressione

Tecnica che permette di ricavare informazioni sui valori futuri di una variabile partendo da un'altra variabile o insiemi di variabili note. La **regressione lineare** rappresenta un metodo di stima del valore atteso partendo da due variabili y e x legati da una retta. Tale retta è del tipo

$$y = mx + q$$

Dove q è l'**intercetta**, ovvero il punto dove la retta incontra l'asse delle ordinate, mentre m è il **coefficiente di regressione** che indica quando varia la y al variare dell'unità x. La retta di regressione si ottiene applicando il metodo dei minimi quadrati.

## Metodo dei minimi quadrati

Tecnica che permette di trovare una funzione che minimizza lo scarto quadratico tra un generico punto y e la retta di regressione. La formula è

$$g(f) = \sum_{i=1}^n |f(x_i) - y_i|^2$$

Per trovare i coefficienti della retta di regressione, partiamo da questa formula e applichiamo le variabili m e q

$$g(m, q) = \sum_{i=1}^n [mx_i + q - y_i]^2$$

Risolvendo il sistema e applicando la formula della varianza e covarianza, troveremo che

$$m = \frac{C_{xy}}{S_x^2} \quad q = \bar{y} - \frac{C_{xy}}{S_x^2} \bar{x}$$

Formula excel =REGR.LIN(colonnaY;colonnaX;vero;vero) selezionando prima due caselle e poi premendo CTRL+SHIFT+INVIO

## Coefficiente di Pearson (o coefficiente di correlazione lineare)

Il metodo dei minimi quadrati fornisce la retta che meglio approssima i dati, ma non il grado di approssimazione. Per questo motivo si usa il coefficiente di correlazione lineare (o di Pearson), che trova il grado di approssimazione della retta di regressione attraverso il processo di regressione lineare. Tale coefficiente ci fa capire se i dati seguono l'andamento di una retta o meno

$$r_{xy} = \frac{C_{xy}}{S_x S_y}$$

Ovvero l'indice di covarianza diviso la deviazione standard di x moltiplicata la deviazione standard di y.

- $r_{xy} \in [-1; +1]$
- Se  $r_{xy} = \pm 1$  i dati sono perfettamente allineati con la retta di regressione
- Se  $r_{xy} > 0$  la retta è ascendente
- Se  $r_{xy} < 0$  la retta è discendente

Formula excel =PEARSON(valori\_indipendenti;valori\_dipendenti)

## PROBABILITÀ

Disciplina che si occupa dello studio dei fenomeni aleatori (o stocastici o non deterministici), ovvero quei fenomeni che ripetendosi sotto le medesime condizioni iniziali producono risultati diversi (Es: lancio dado); in caso contrario sono detti deterministici. Ad ogni evento associa un numero  $P$  compreso tra  $[0,1]$  detto probabilità.

- **Probabilità classica:** 
$$P(A) = \frac{\text{numero casi favorevoli}}{\text{numero casi possibili}}$$
 supponendo che tutti i casi siano egualmente possibili. Es: nel lancio di un dado c'è una probabilità  $\frac{1}{6}$  che esca la faccia con il numero 2 al primo tentativo

- **Probabilità frequentista:** è il limite del rapporto tra gli esperimenti favorevoli e il totale di quelli effettuati per il numero di esperimenti totali

$$P(A) = \lim_{n \rightarrow \infty} \frac{\text{numero casi favorevoli}}{\text{numero casi possibili}}$$

Es: lancio di una moneta, probabilità che esca testa per tanti lanci consecutivi è  $\lim_{n \rightarrow \infty} 1/2$

- **Probabilità soggettivista:** la probabilità degli eventi viene attribuita soggettivamente a seconda del grado di fiducia del verificarsi dell'evento, il quale dipende dalla persona che la valuta e dalle informazioni disponibili

$$P(E) = \frac{\text{prezzo da pagare}}{\text{somma vinta al verificarsi di } E}$$

Essa risponde al **criterio di coerenza**, dove il soggetto che attribuisce la probabilità dell'evento deve conoscere ogni aspetto dell'esperimento. Es: il soggetto dovrà essere sia scommettitore che generatore di eventi.

### Spazio di probabilità

È la terna  $(\Omega, \mathbb{P}, P)$  dove

- $\Omega$  è lo spazio campionario, ovvero l'insieme dei possibili risultati di un evento
- $\mathbb{P}$  è una famiglia di sottoinsiemi di  $\Omega$  per i quali si può calcolare una probabilità
- $P$  è la misura della probabilità tale per cui  $P(\Omega) = 1$ .

Es: lancio di un dado

- $\Omega = \{1,2,3,4,5,6\}$
- $\mathbb{P} = \text{numero pari} = \{2,4,6\}$
- $P(E) = \frac{\text{numero casi favorevoli a } E}{\text{numero casi possibili di } \Omega} = \frac{3}{6}$

### Definizione assiomatica di probabilità

Due eventi sono incompatibili quando la loro intersezione forma l'insieme vuoto. Dato  $P(A)$  l'insieme delle parti, ovvero l'insieme dei suoi sottoinsiemi, avremo che:

- $A \cap B$  i due eventi  $A$  e  $B$  si verificano entrambi
- $A \cup B$  almeno uno dei due eventi si verifica
- $A^C$  l'evento associato ad  $A$  non si verifica

La definizione assiomatica di Kolmogorov dice che la probabilità è una funzione che associa a ciascun sottoinsieme di  $\Omega$  un numero reale non negativo

$$P: \mathbb{P}(\Omega) \rightarrow \mathbb{R}_+$$

1.  $\forall A \subseteq \Omega \quad P(A) \geq 0$
2.  $P(\Omega) = 1$

### Probabilità della somma logica di eventi

Dati due eventi  $A$  e  $B$

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Se i due eventi sono incompatibili allora  $P(A \cup B) = P(A) + P(B)$

Es: Un'urna contiene 15 palline numerate da 1 a 15. Calcolare la probabilità che, estraendo una pallina,

- a) esca un numero dispari o maggiore di 10;
- b) esca un numero minore di 6 o maggiore di 10.

Definiamo gli eventi: A = esce un numero dispari, B = esce un numero > 10, C = esce un numero < 6.

Dalla definizione di probabilità classica  $P(A)=8/15$ ,  $P(B)=5/15$ ,  $P(C)=5/15$ ,  $P(A \cap B)=3/15$  dal teorema segue :

a)  $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 2/3$

b) essendo gli eventi B e C incompatibili  $P(B \cup C) = P(B) + P(C) = 2/3$

### Probabilità condizionata

Probabilità che l'evento A si verifichi supponendo che l'evento B si sia già verificato (A subordinato o condizionato da B)

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

Es: Si lanci una coppia di dadi. Se la somma è 6 (evento B) si determini la probabilità che uno dei due dadi abbia come esito 2 (evento A).

In questo caso si avrà:

$B = \{\text{somma dei dadi } 6\} = \{(1, 5)(2, 4)(3, 3)(4, 2)(5, 1)\} \rightarrow |B| = 5$

$A = \{\text{un 2 si presenta su almeno un dado}\}$

Ne segue che  $A \cap B = \{(2, 4), (4, 2)\} \rightarrow |A \cap B| = 2$  quindi  $P(A|B) = (P(A \cap B)) / (P(B)) = |A \cap B| / |B| = 2/5$

- **Teorema della probabilità composta:** la probabilità che due eventi si verifichino contemporaneamente è pari alla probabilità di uno dei due eventi moltiplicando con la probabilità dell'altro il quale è subordinato al primo

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

- **Eventi indipendenti:** Due eventi A e B sono stocasticamente indipendenti se il presentarsi dell'evento B non influenza la probabilità dell'evento A (e viceversa).

$$P(A|B) = P(A) \text{ o } P(B|A) = P(B) \Rightarrow P(A \cap B) = P(A)P(B)$$

- **Teorema della probabilità assoluta (o totale):** è uguale alla somma delle probabilità degli eventi parziali incompatibili

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

Dove  $A_i$  è l'insieme di probabilità che condizionano B

### Teorema di Bayes

Viene impiegato per calcolare la probabilità di una causa che ha scatenato l'evento verificato. La conoscenza della verifica di un particolare evento (causa) può modificare la probabilità di un altro evento (effetto) aumentando o riducendo la probabilità iniziale. Esso calcola, quindi, la probabilità di una causa che ha scatenato l'evento verificato.

Es: si può calcolare la probabilità che una certa persona soffra della malattia per cui ha eseguito il test diagnostico o viceversa non sia affetta da tale malattia conoscendo la frequenza con cui si presenta la malattia e la percentuale di efficacia del test diagnostico.

Considerando un insieme di alternative  $A_i$  e  $A_j$  che partizionano  $\Omega$ , la formula di Bayes è:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^n P(B|A_j)P(A_j)}$$

Dove

- $P(A)$  è la probabilità a priori di A. "A priori" significa che non tiene conto di nessuna informazione riguardo B.
- $P(A|B)$  è la probabilità condizionata di A, noto B. Viene anche chiamata probabilità a posteriori, visto che è derivata o dipende dallo specifico valore di B.
- $P(B|A)$  è la probabilità condizionata di B, noto A.
- $P(B)$  è la probabilità a priori di B.

Es: Si consideri una scuola che ha il 60% di studenti maschi e il 40% di studentesse femmine. Le studentesse indossano in egual numero gonne o pantaloni; gli studenti indossano tutti quanti i pantaloni. Un osservatore, da lontano, nota un generico studente coi pantaloni. Qual è la probabilità che quello studente sia una femmina?

Poniamo

- A = studente osservato femmina
- B = studente osservato indossa i pantaloni
- P(A) = probabilità che lo studente sia femmina senza nessun'altra informazione. Dato che l'osservatore vede uno studente a caso, ciò significa che tutti gli studenti hanno la stessa probabilità di essere osservati. Essendo le studentesse il 40% del totale, la probabilità risulterà  $4/10=2/5$ .
- P(B|A) = probabilità che uno studente indossi i pantaloni, noto che lo studente è femmina. Poiché indossano gonne e pantaloni in egual numero, la probabilità sarà di  $1/2$ .
- P(B) = probabilità che uno studente qualsiasi (maschio o femmina) indossi i pantaloni. Poiché il numero di coloro che indossa i pantaloni è di 80 (su 100 studenti, 60 i maschi che indossano tutti i pantaloni, 20 le donne che indossano anche i pantaloni dato che sono in egual numero di quelle che indossano le gonne), la probabilità P(B) è  $80/100=4/5$ .

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{\frac{1}{2} * \frac{2}{5}}{\frac{4}{5}} = \frac{1}{4}$$

Abbiamo pertanto  $1/4$  cioè il 25% di probabilità che lo studente osservato sia femmina.

### Variabile aleatorie (o casuali o stocastiche)

Assegnato uno spazio di probabilità  $(\Omega, \mathbb{P}, P)$  una variabile aleatoria X è una funzione che ad ogni elemento di  $\Omega$  associa un numero reale  $X: \Omega \rightarrow \mathbb{R}$ . Una variabile aleatoria è una grandezza che nel corso di un esperimento può assumere diversi valori imprevedibili a priori. Es: lancio di un dado, non può conoscere a priori il valore della faccia che si presenterà, tale valore è aleatorio.

- **Variabili aleatorie discrete:** una variabile aleatoria si dice discreta se l'insieme dei valori S da essa assumibili (detto **supporto** della variabile aleatoria) è un insieme finito.

$$X: \Omega \rightarrow S \quad \text{con} \quad S = \{s_1, s_2, \dots, s_n\}$$

- **Distribuzione discreta di probabilità:** è rappresentata dall'elenco delle modalità che la variabile assume

$$P_X(s) = \begin{cases} P(X = s) & \forall s \in S \\ 0 & \text{altrimenti} \end{cases}$$

- **Funzione di ripartizione:** poiché non si conosce a priori il valore assunto da una variabile aleatoria, si può fare una valutazione probabilistica sui valori che assumerà usando la funzione di ripartizione. Essa è un numero reale compreso tra  $[0,1]$  e vale:

$$F_X(t) = \sum_{s \in S, s \leq t} P(s) \quad \forall t \in \mathbb{R}$$

- **Variabili aleatorie continue:** una variabile aleatoria è continua se la corrispondente funzione di ripartizione è continua. È una variabile che può assumere qualsiasi valore (entro un certo intervallo). Alcuni esempi di variabili aleatorie continue sono: velocità di un mezzo, statura di una persona, temperatura di un forno

- **Variabili aleatorie assolutamente continue:** quando esiste il seguente integrale

$$F_X(t) = \int_{-\infty}^t f_X(u) du \quad \forall t \in \mathbb{R}$$

La derivata della funzione di ripartizione, se esiste, è detta **funzione densità di probabilità**, e deve essere uguale a 1, che indica il 100% delle probabilità  $\int_{-\infty}^{+\infty} f_X(u) = 1$

### Indici di tendenza centrale per le variabili aleatorie

Sintetizzano con un solo valore aleatorio le caratteristiche della distribuzione.

- **Valore atteso (o media o speranza matematica):** rappresenta il valore previsto che si potrà verificare in un gran numero di prove; è definito come

- La sommatoria dei vari volumi s pesati con le singole probabilità di s

$$\mu = E[X] = \sum_{s \in S} sp(s) \quad \text{con} \quad S = \text{supporto di } X$$

- L'integrale definito della funzione densità di probabilità per ogni singolo valore

$$\mu = E[X] = \int_{-\infty}^{+\infty} uf(u)du$$

- **Varianza:** misura la dispersione dei valori assunti da X rispetto al suo valore atteso; è definita come
  - Caso discreto  $\sigma_x^2 = V[X] = \sum_{s \in S} (s - \mu)^2 p(s)$
  - Caso continuo  $\sigma_x^2 = V[X] = \int_{-\infty}^{+\infty} [u - \mu]^2 f(u)du$

Dove  $\mu$  è il valore atteso. Tanto più grande è la varianza, tanto più i valori X saranno lontani l'uno dall'altro; tanto più la varianza è piccola, tanto più i valori di X saranno vicini tra loro.

- **Deviazione standard:**  $\sigma_x = \sqrt{\sigma_x^2}$

## DISTRIBUZIONI NOTEVOLI

La distribuzione è una rappresentazione del modo in cui le diverse modalità di un carattere si distribuiscono nelle unità statistiche che compongono il collettivo oggetto di studio.

- **Distribuzione di Bernoulli:** distribuzione di una variabile aleatoria X che segue l'andamento di una curva bernoulliana. In questa distribuzione di probabilità si fanno n prove ripetute e indipendenti e i risultati possibili di ciascuna prova possono essere soltanto successo o insuccesso. La funzione di ripartizione è

$$F_x(t) = \sum_{s \leq t} P_x(s)$$

- **Distribuzione binomiale:** siano  $x_1, x_2, \dots, x_n$  variabili bernoulliane, di uguale parametro p e stocasticamente indipendenti tra loro (l'indipendenza stocastica di due eventi A e B si ha quando il verificarsi di uno non modifica la probabilità di verificarsi dell'altro, ovvero quando la probabilità condizionata  $P(A|B)$  oppure  $P(B|A)$  è pari rispettivamente a  $P(A)$  e  $P(B)$ ). Sara X la somma delle variabili  $x_1, x_2, \dots, x_n$  distribuita seconda una binomiale di parametri n e p

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

- $P(X=k)$  è la probabilità che X assuma valore k
- n è il numero di lanci o di variabili aleatorie X
- k è il numero di volte che si può presentare ogni singolo valore
- $\binom{n}{k} = \frac{n!}{(n-k)!k!}$

Formula excel **=DISTRIB.BINOM(num\_successi;prove;probabilità\_s;cumulativo)** se cumulativo = VERO viene distribuita la funzione di ripartizione, se FALSO viene restituita la distribuzione di probabilità

- **Distribuzione di Poisson:** si ottiene quando si effettuano molte prove, ciascuna con bassa probabilità di successo. Il grafico della distribuzione è suddiviso in un numero molto grande di sottointervalli
  - La probabilità che si verifichi un evento in ogni sottointervallo è molto piccola
  - La probabilità che si verifichi un evento è costante per tutti i sottointervalli
  - L'evento non si può verificare più di una volta in ciascuno dei sottointervalli
  - Gli eventi che si verificano in intervalli disgiunti sono indipendenti

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$\lambda$  è parametro caratteristico sia del valore medio che della varianza. La funzione di ripartizione è

$$F_X(t) = \sum_{k \in \mathbb{N} \ k \leq t} \frac{\lambda^k}{k!} e^{-\lambda} \quad \forall t \in \mathbb{R}$$

Es: Il titolare di una locanda rinomata per i suoi antipasti a base di pesce sta tenendo sotto controllo il ciclo di produzione per verificare quanti gamberetti vengono inseriti su ogni antipasto. Supponendo che tale numero segua una distribuzione Poissoniana con  $\lambda = 6$  qual è la probabilità che in un particolare antipasto si trovino esattamente 5 gamberetti?

$$P(X = 5) = \frac{\lambda^k}{k!} e^{-\lambda} = \frac{6^5}{5!} e^{-3} = \frac{7776 * 0.0024}{120} = 0.155$$

Formula excel **=POISSON(x;media;cumulativo)** se cumulativo = VERO viene distribuita la funzione di ripartizione, se FALSO viene restituita la distribuzione di probabilità



- **Distribuzione esponenziale:** distribuzione importante nello studio di quelle variabili che descrivono i tempi di attesa al verificarsi di un evento (Es: il tempo di attesa per arrivo trasporto). Essa è l'unica distribuzione che verifica la **proprietà di non memoria** (tiene conto solo dell'evento precedente a quello considerato). Questa distribuzione descrive la durata di vita di un fenomeno che non invecchia o privo di memoria.

- **Densità di probabilità**  $f_x(t) = \begin{cases} 0 & \text{se } t < 0 \\ \lambda \exp(-\lambda t) & \text{se } t \geq 0 \end{cases}$
- **Funzione di ripartizione**  $F_x(t) = \begin{cases} 0 & \text{se } t < 0 \\ 1 - \exp(-\lambda t) & \text{se } t \geq 0 \end{cases}$
- **Proprietà di non memoria**  $P(X > s + t | X > s) = P(X > s) \forall t, s \in \mathbb{R}_+$   
 $P(X > s + t | X > s)$  è una probabilità condizionata, infatti l'evento che accade nel tempo s+t dipende da quello che accade al tempo s.

Es: Consideriamo il gioco del lotto ed indico con X la variabile aleatoria {numero minimo di estrazioni su una data ruota affinché esca il numero 34}. In questo caso s, t saranno numeri naturali.

- $P(X > s)$  è la probabilità che nelle prime s prove non esca il numero 34;
- $P(X > s + t)$  è la probabilità che nelle prime s + t prove non esca il numero 34;
- $P(X > s + t | X > s)$  è la probabilità che, supposte che nelle prime s prove non sia uscito il numero 34, nelle successive t non esca in numero 34.

La proprietà di non-memoria afferma che la probabilità che non si verifichi alcun successo (ovvero estrazione del numero 34) fino alla prova s + t (supposto che non si sia verificato nelle prime s prove) non dipende da s ossia da quanto si è atteso (ovvero il ritardo), ma solo dal numero t di prove ancora da effettuarsi. È come se, ad ogni estrazione, il numero 34 non avesse memoria di quello che è accaduto nel passato: infatti ogni numero ha sempre la stessa probabilità di essere estratto.

- **Distribuzione di Weibull:** descrive la durata di vita per un fenomeno la cui probabilità di morire può variare nel tempo, in funzione di un parametro  $\beta$ . Quando il parametro  $\beta$  è uguale a 1, la distribuzione di Weibull equivale alla distribuzione esponenziale.

- **Densità di probabilità**  $f_x(t) = \begin{cases} 0 & \text{se } t < 0 \\ \alpha \beta t^{\beta-1} \exp(-\alpha t^\beta) & \text{se } t \geq 0 \end{cases}$
- **Funzione di ripartizione**  $F_x(t) = \begin{cases} 0 & \text{se } t < 0 \\ 1 - \exp(-\alpha t^\beta) & \text{se } t \geq 0 \end{cases}$
- **Funzione di affidabilità**  $R_x(t) = P(X > t) = 1 - P(X \leq t) = 1 - F_x(t)$   
 Questa è la probabilità che un componente si ancora funzionante al tempo t.
- **Funzione di intensità di rottura**  $\lambda(t) = \frac{n^\circ \text{ oggetti guasti al tempo } t}{n^\circ \text{ oggetti funzionanti al tempo } t} = \frac{f_x(t)}{R_x(t)} = \frac{f_x(t)}{1 - F_x(t)}$   
 Questa funzione è il tasso di guasto; si può provare che  $\lambda(t) = \alpha \beta t^{\beta-1}$  da cui
  - per  $\beta < 1$  il tasso di guasto diminuisce nel tempo (alta mortalità infantile)
  - per  $\beta = 1$  il tasso di guasto è invariante nel tempo (memoryless)
  - per  $\beta > 1$  il tasso di guasto aumenta con il tempo (invecchiamento)

Formula excel =WEIBULL(x;alfa;beta;cumulativo)

- **Distribuzione normale (o di Gauss):** Descrive variabili casuali che tendono a concentrarsi attorno a un valor medio. Si ha che  $X \simeq N(\mu, \sigma)$ , ovvero che la variabile aleatoria X è distribuita secondo una normale di parametri  $\mu$  (valore atteso o media,  $E[X]$ ) e  $\sigma$  (scarto quadratico medio,  $V[X]$ ) se ha densità di probabilità uguale a

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

- **Funzione di ripartizione**  $F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt = P(X \leq t)$
- X è una variabile casuale che prende valori in tutto l'asse dei numeri reali.
- La curva di Gauss ha una forma campanulare ed è simmetrica rispetto al valore atteso
- La curva ha un massimo in  $\left(\mu, \frac{1}{\sigma\sqrt{2\pi}}\right)$
- Gli indici di tendenza centrale coincidono
- La curva è crescente in  $x < \mu$
- Ha un asintoto orizzontale in  $y=0$
- Il suo range interquartile copre un intervallo compreso tra  $\mu - \frac{2}{3\sigma}$  e  $\mu + \frac{2}{3\sigma}$
- L'area sottesa alla curva di Gauss è uguale a 1

- **Distribuzione normale standardizzata:** Una distribuzione di Gauss è standardizzata, quando le variabili casuali che lo costituiscono hanno varianza uguale a 1 e media o valore atteso uguale a 0. La funzione di ripartizione è:

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

Formula excel =DISTRIB.NORM(x;media;deviazione\_standard;cumulativo)

- **Distribuzione di chi-quadro:** è impiegata per stimare la varianza; descrive la somma dei quadrati di alcune variabili aleatorie indipendenti, avente distribuzione normale standardizzata:

$$X^2 = X_1^2 + X_2^2 + X_3^2 + \dots + X_n^2$$

Dove n è il numero di gradi di libertà (prove o variabili aleatorie). Dato il valore atteso  $E[X]=n$  e la varianza  $V[X]=2n$ :

$$\begin{aligned} \text{○ Densità di probabilità} \quad f_x(t) &= \begin{cases} 0 & \text{se } t < 0 \\ \frac{1}{\Gamma(\frac{n}{2})} \left(\frac{1}{2}\right)^{\frac{n}{2}} t^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}t\right) & \text{se } t \geq 0 \end{cases} \\ \text{○ Funzione di ripartizione} \quad \Gamma(\alpha) &= \int_0^{+\infty} e^{-x} x^{\alpha-1} dx \text{ si chiama anche } \textbf{funzione gamma di Eulero} \end{aligned}$$

Formula excel =DISTRIB.CHI(x;n\_gradi\_libertà)

- **Distribuzione t di Student:** È una distribuzione di probabilità continua che governa il rapporto tra due variabili aleatorie, dove la prima variabile ha una distribuzione normale standardizzata e la seconda variabile il cui quadrato aveva distribuzione CHIQUADRATO. Se da una distribuzione normale standardizzata prendo un piccolo campione, si noterà che esso sarà distribuito secondo una t di Student. Essa viene impiegata per la stima della media di una popolazione che segue la distribuzione normale.

$$t = \frac{Z}{\sqrt{\frac{S^2}{n}}}$$

- Z = variabile aleatoria che segue la distribuzione normale standardizzata  $Z = \frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$
- $S^2$  è la distribuzione chi-quadro con n gradi di libertà
- **Densità di probabilità**  $f_x(t) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \frac{1}{\sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-\frac{1}{2}(n+1)}$

Formula excel =DISTRIB.T(x;n\_gradi\_libertà;codà)

## TEOREMI DI CONVERGENZA

Studiano le condizioni sotto cui certe successioni di variabili casuali con comportamento incognito di una certa distribuzione tendono ad altre distribuzioni note.

### Legge dei grandi numeri

La media calcolata a partire da un numero sufficiente di campioni è sufficientemente vicina alla media reale (media campionaria); essa migliora al crescere del numero di campioni (n).

$$\bar{x}_n \Rightarrow^d M$$

Con questa legge, prevedo probabilisticamente, la proporzione di successi in una sequenza di n realizzazioni indipendenti di un evento E

### Teorema del limite centrale

È possibile approssimare la somma di un gran numero di variabili casuali, indipendentemente dalla loro distribuzione, con una distribuzione normale i cui parametri sono  $\mu$  (valore atteso) e scarto quadratico medio pari a

$$se = \frac{\sigma}{\sqrt{n}}$$

Che è l'**errore standard**, dove  $\sigma$  è la deviazione standard. L'errore standard è il grado di certezza con il quale la media, calcolata da un campione casuale, stima la vera media della popolazione dalla quale il campione è stato tratto. La distribuzione della media campionaria può essere approssimata dalla distribuzione normale, con media pari alla media della popolazione  $\mu$  e varianza data dalla varianza  $\sigma^2$  della popolazione su n.

## STATISTICA INFERENZIALE

Insieme di metodi con cui si cerca di trarre una conclusione sulla popolazione in base ad informazioni ricavate da un campione. Si estrae un campione di dati dalla popolazione, vengono eseguiti il calcolo delle statistiche campionario sul campione (media, ecc.) e si stimano i parametri dell'intera popolazione in base ai risultati dei campioni.

### Stime di parametri

Il comportamento della popolazione viene studiato attraverso i dati relativi ad un campione, un sottoinsieme rappresentativo della popolazione, di  $n$  unità. L'**inferenza** è il processo attraverso il quale, da un campione significativo e non dai dati relativi a tutta la popolazione, si effettuano stime e verifiche delle ipotesi e da cui si deducono informazioni. Quando le caratteristiche che si vogliono individuare sono esprimibili numericamente (si possono contare, discrete), allora prendono il nome di **parametri**. I metodi principali della statistica inferenziale (o statistica induttiva) sono due:

- **Stima:** il campione viene usato per stimare un parametro della popolazione e porta alla costruzione di un intervallo di fiducia per il suddetto parametro.
- **Test di ipotesi:** si formula un'ipotesi parametrica o non parametrica, e la si verifica sui dati del campione.

Il **parametro** (la variabile casuale) è un valore numerico che descrive una caratteristica di una popolazione, ed è una grandezza associata ad una sua distribuzione (caratterizzata da un valore atteso e varianza). Una stima del parametro è una misura che descrive una caratteristica del campione. Il processo che porta alla selezione delle  $n$  unità significative (costituiscono il campione) si chiama campionamento.

### Strategie di campionamento

Distinguiamo tre tipi di campionamento:

- **Casuale:** ogni unità della popolazione ha la stessa probabilità di entrare a far parte del campione, dato che la selezione avviene casualmente.
- **Stratificato:** la popolazione viene suddivisa in sottogruppi, detti **strati**, in base ad una o più variabili ausiliarie o di classificazione (età, sesso, etc..). Da ciascun sottogruppo viene estratto in maniera indipendente (e senza ripetizione) un campione di numerosità  $n$ .
- **A Grappoli:** la popolazione viene suddivisa in grappoli, ossia da gruppi di unità omogenei per qualche caratteristica, dove l'unità campionata è il grappolo.

### Statistica campionaria

Variabile casuale che è funzione del campione aleatorio, cioè dipende solo dai dati del campione e non da variabili incognite

$$H_n = f(X_1, \dots, X_n)$$

- **Stimatore:** statistica campionaria usata per stimare (dedurre informazioni) il parametro della popolazione.
- **Distribuzioni campionarie:** sono le distribuzioni delle statistiche campionarie, ovvero delle distribuzioni di probabilità che dipendono dalle variabili casuali presenti nel campione preso in esame. Inoltre sono anche distribuzioni di valori di una statistica del campione (media, ecc.) ottenuta a partire da tutti i campioni possibili.

- **Media campionaria:** è la stima della media della popolazione

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$$

- **Intervallo di confidenza della media campionaria:** è un intervallo di valori intorno alla media campionaria entro il quale si ritiene sia compreso il parametro in esame con un certo grado di confidenza.
- **Varianza campionaria:** è la varianza ristretta al campione preso in esame; la varianza della distribuzione è uguale al valore atteso campionario

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

Formula excel =DEV.ST(valori) per i campioni; su tutta la popolazione =DEV.ST.POP(valori)

## Stime e intervalli puntuali

La stima del parametro di una popolazione

- Se data da un solo numero, è chiamata **stima puntuale** ( $E_s$ : la misura è 5.28 metri).
- Se data mediante due numeri che definiscono un intervallo (o range), si chiama **stima intervallare** ( $E_s$ : 5.28,  $\pm 0.03$ mt). In quest'ultimo caso, il secondo numero definisce l'ampiezza dell'intervallo. Più è ampio questo intervallo, maggiore è la probabilità di trovare il parametro (il range dell'esempio è [5.25, 5.31]).

## Intervallo di fiducia (o di confidenza)

Un intervallo di fiducia è un intervallo di valori che ha una determinata probabilità di contenere il parametro oggetto di stima (valore atteso). Ad ogni intervallo di confidenza viene associato un **livello di confidenza** ( $1-\alpha$ ) che rappresenta il grado di attendibilità del nostro intervallo. Gli intervalli di fiducia del 95% e del 99% sono quelli più comunemente usati. L'intervallo di fiducia è costruito sul valore della media campionaria  $\bar{X}$  ed ha un raggio pari a:

$$Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}}$$

Formula excel =CONFIDENZA(alfa,dev\_standard,dimensioni) successivamente, per ottenere l'estremo sinistro o destro si dovrà sottrarre o aggiungere alla media il risultato appena ottenuto

Distinguiamo 4 possibili scenari vincolati ai parametri media e varianza della popolazione

1. **Caso popolazione distribuita non normalmente e varianza nota**: ricordando che la variabile aleatoria che funge da valore di probabilità  $Z = \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}}$  definiamo l'intervallo di confidenza come

$$p\left(\bar{X} - Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}\right) = 1 - p\left(Z_{\frac{\alpha}{2}}\right)$$

Dove

- ciò che è minore di  $\mu$  è il limite fiduciale inferiore, mentre ciò che è maggiore di  $\mu$  è il limite fiduciale superiore.
  - $\bar{X}$  è la media campionaria e  $\mu$  è la media della popolazione
  - $\frac{\sigma}{\sqrt{n}}$  è l'errore standard della media
  - $\alpha$  è il livello di fiducia
  - $Z_{\frac{\alpha}{2}}$  è il quantile della normale di ordine  $1 - \frac{\alpha}{2}$
2. **Caso popolazione distribuita non normalmente e varianza sconosciuta**: come al caso 1 ma essendo la varianza sconosciuta si sostituisce la deviazione standard con una sua stima
$$S_n = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$
  3. **Caso popolazione distribuita normalmente e varianza nota**: di opera come nel caso 2, con l'unica differenza che non vi sono restrizioni sulla numerosità del campione. Infatti, nel 1° caso, n deve essere  $\geq 30$ .
  4. **Caso popolazione normalmente distribuita e varianza sconosciuta**: si stima la varianza non nota della popolazione con il seguente stimatore
    - **Varianza campionaria corretta**:  $\hat{S}_n^2 = \frac{n}{n-1} S_n^2$Alla luce di ciò, la probabilità che il valore atteso cada nell'intervallo di confidenza è
$$p\left(\bar{X}_n - t_{1-\frac{\alpha}{2}} * \frac{\hat{S}_n}{\sqrt{n}} < \mu < \bar{X}_n + t_{1-\frac{\alpha}{2}} * \frac{\hat{S}_n}{\sqrt{n}}\right) = 1 - \alpha$$
Per il calcolo del quantile  $t_{1-\frac{\alpha}{2}}$  si usa la formula excel =INV.T( $\alpha$ ;  $n - 1$ )

## VERIFICA DI IPOTESI

Il test di verifica d'ipotesi si utilizza per verificare la bontà di un'ipotesi. Terminologia:

- **Ipotesi statistica:** È un'affermazione o congettura riguardante una o più caratteristiche di una distribuzione di probabilità di una variabile casuale. Es: il voto medio degli studenti di un corso è 26
- **Verifica di ipotesi:** È l'analisi statistica che consente di controllare che l'ipotesi statistica sia vera. È un metodo che si basa sullo studio delle differenze tra i risultati osservati e quelli che si aspetterebbe se una possibile ipotesi sulla distribuzione fosse vera.
- **Test di ipotesi:** metodo con cui lo sperimentatore effettua la verifica di ipotesi. Abbiamo due tipi di test:
  - **Test parametrici:** sono test effettuati sulla base di ipotesi relative ai parametri della distribuzione della popolazione, quali la media, varianza, ecc... Tra questi test ricordiamo il test di Student.
  - **Test non parametrici:** sono test che si effettuano quando le caratteristiche della distribuzione della popolazione o parametri, quali ad esempio la media, varianza, ecc... non si conoscono o non sono esprimibili. L'assenza di questi parametri non aiuta ad identificare il tipo di distribuzione, quindi è necessario utilizzare metodi alternativi, ovvero dei test di tipo non parametrico. Tra questi, abbiamo il Test di Kolmogorov-Smirnov e il test del Chi-quadro.

### Caratteristiche generali di un test di ipotesi

Gli elementi fondamentali di un test di ipotesi sono:

- Una popolazione statistica su cui fare i test
- Un **ipotesi nulla**  $H_0$  da convalidare su un campione scelto dalla popolazione statistica su cui viene effettuato il test di ipotesi.
- Un **ipotesi alternativa**  $H_1$  opposta all'ipotesi nulla e rappresenta la conclusione raggiunta quando l'ipotesi nulla è rifiutata.
- Una **distribuzione campionaria**  $T = T(X_1, \dots, X_n)$  di cui è nota la distribuzione quando  $H_0$  è vera. Essa è divisa in due regioni:
  - Una **regione di accettazione**  $\bar{C}$ , ovvero l'insieme dei valori del test statistico che è probabile si verifichino quando l'ipotesi nulla è vera.
  - Una **regione critica**  $C$ , ovvero l'insieme dei valori del test statistico che è probabile si verifichino quando l'ipotesi nulla è falsa.
- Un **Livello di Significatività** (ampiezza del test) indicato con  $\alpha$ , ovvero quel valore tale che l'ipotesi nulla  $H_0$  viene rifiutata, quando questa è vera con probabilità  $\alpha$ .
- Un valore critico o valore soglia

Da ricordare che una verifica di ipotesi non dice se la nostra ipotesi è vera o falsa, ma dice se l'ipotesi fatta è compatibile con i dati raccolti.

### Livello di significatività

Relativamente al test statistico si può respingere o accettare l'ipotesi zero, ovvero l'ipotesi nulla potrebbe essere giusta ma potrebbe essere anche sbagliata. La probabilità di commettere questo errore si chiama **livello di significatività del test**. Questo livello di significatività generalmente viene fissato ai valori 0.05 (5%) o di 0.01(1%).

Es: se  $\alpha = 0.05$ , c'è il 5% di probabilità di rifiutare l'ipotesi quando dovrebbe essere accettata, ovvero si ha il 5% di probabilità di sbagliare e il 95% di prendere una decisione giusta.

Esso rappresenta una stima quantitativa della probabilità che le differenze osservate siano dovute al caso. Per prendere una decisione su  $H_0$ , si dovrà determinare il valore critico (valore soglia). Quest'ultimo parametro separerà la regione di accettazione dalla regione di rifiuto. Il valore critico si evince da un'equazione in cui è presente la **regola di rifiuto**  $|\bar{X}_n - \mu_0| > k$ . Dalla regola di rifiuto si ricaverà questa probabilità condizionata:

$$P(|\bar{X}_n - \mu_0| > k | H_0) = \alpha$$

### Errori

In un test si possono verificare due tipologie di errori, entrambi dipendenti dal livello di significatività  $\alpha$ :

1. **Errori di prima specie:** seguito a errori statistici si rifiuta l'ipotesi  $H_0$  quando questa invece è vera
2. **Errori di seconda specie:** si accetta l'ipotesi  $H_0$  quando questa invece è falsa

A posteriori, ovvero dopo aver estratto il campione e presa una decisione con la statistica test, non ha senso parlare di probabilità di decisioni giuste o errate, in quanto non essendo nota l'ipotesi vera sulla popolazione si è già commesso di fatto un errore o si è già presa di fatto una decisione giusta.

### Test parametrici sulla media di una popolazione statistica

- **Z test bilatero con varianza nota:** Estraiamo un campione aleatorio  $(X_1, \dots, X_n)$  proveniente da una popolazione normale con media  $\mu$  incognita e varianza  $\sigma^2$  nota. Vogliamo fare un test di verifica per trovare l'ipotesi nulla  $H_0$  e quella alternativa  $H_1$ . Data una costante  $\mu_0$ , siccome la popolazione è normale, se fosse vera la  $H_0$

$$U = \frac{\bar{X}_n - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \cong N(0,1)$$

Dove U dovrebbe essere distribuita secondo una normale standardizzata. Fissato un livello di significatività  $\alpha$

$$\begin{cases} se - Z_1 - \frac{\alpha}{2} < U < Z_1 - \frac{\alpha}{2} & \text{accettiamo } H_0 \\ se U < Z_1 - \frac{\alpha}{2} \cup U > Z_1 - \frac{\alpha}{2} & \text{rifiutiamo } H_0 \end{cases}$$

- **P-value:** è il più basso livello di significatività a cui l'ipotesi nulla può essere rifiutata

$$p - value = 2(1 - F_x(|U|))$$

Calcolato il p-value (area delle due code) devo confrontarlo col livello di significatività  $\alpha$  del test

$$\begin{cases} se p - value \geq \alpha & \text{l'ipotesi nulla va accettata al livello } \alpha \\ se p - value \leq \alpha & \text{l'ipotesi nulla va rigettata al livello } \alpha \end{cases}$$

In parole povere, se il p-value è un numero molto piccolo allora si scarcerà l'ipotesi nulla

- **Z test unilatero con varianza nota:** si ragiona come al caso precedente, solo che la regione critica sarà ad una sola coda.
- **Test di student:** lo scopo di questo test è valutare se la differenza fra due o più coppie dati (campioni) è significativa, ossia affermare che la differenza non è dovuta al caso ma esiste una reale diversità tra le medie delle due popolazioni da cui i campioni stessi derivano. L'ipotesi  $H_0$  è che la differenza è dovuta al caso; il test ci aiuta ad accettare o rifiutare questa ipotesi. Il tipo di confronto usato è la media della differenza di due campioni con una differenza data usando la variabile aleatoria t

$$t = \frac{\text{differenza fra medie campionarie}}{\text{errore standard della differenza fra medie campionarie}} = \frac{\bar{X}_A - \bar{X}_B}{\sigma \sqrt{\frac{1}{N_A} + \frac{1}{N_B}}}$$

dove  $\sigma = \sqrt{\frac{N_A S_A^2 + N_B S_B^2}{N_A + N_B - 1}}$ ,  $N_A$  ed  $N_B$  sono la numerosità dei due campioni A e B,  $S_A^2$  ed  $S_B^2$  sono le varianze campionarie corrette.

Formula excel =TEST.T(matrice1;matrice2;coda;tipo), dove coda può essere 1 o 2 a seconda del test a 1 o 2 code; tipo può essere 1 o 2 se accoppiato o moschedastico, 3 per eteroschedastico; se  $P \leq 0.005$  il test si accetta.

### Test non parametrici sulla bontà dell'adattamento

Quando non è noto come sono distribuiti i dati, si utilizzano i test non parametrici, chiamati anche bontà dell'adattamento perché verificano se una certa distribuzione è compatibile con i dati del campione. La nostra ipotesi  $H_0$  è accettata quando una data distribuzione si comporta nel modo da noi specificato. L'ipotesi alternativa  $H_1$  si ha quando la funzione di ripartizione da noi adattata è diversa per almeno un valore dalla funzione di ripartizione reale della distribuzione.

- **Test di Kolmogorv-Smirnov:** questo test confronta la densità di probabilità del campione osservato con quella ipotizzata e "risponde" alla domanda: questa popolazione è distribuita secondo un'assegnata probabilità o funzione di ripartizione? Tale test si svolge con la seguente funzione di ripartizione empirica

$$\hat{F}_{X_n}(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{U}_{[0,t]}(X_i) \forall t \in \mathbb{R} \text{ con } \mathbb{U}_{[0,t]}(X_i) = \begin{cases} 1 & \text{se } X_i \in [0, t] \\ 0 & \text{altrimenti} \end{cases}$$

Che è una **funzione indicatrice**  $f_A: X \rightarrow \{0,1\}$  che restituisce 1 se  $x$  appartiene ad  $A$  altrimenti 0. La funzione di ripartizione di tale test genererà  $n$  differenze. Successivamente, si considererà l'**estremo superiore** di tali differenze, ossia il valore più grande. Questo valore verrà chiamato

$$D_n = \sup_{t \in \mathbb{R}} |F(t) - \hat{F}_{X_n}(t)|$$

il quale restituisce il quantile  $d_{1-\alpha}$

- Se  $d_{1-\alpha} < D_n < 1$  si rifiuta l'ipotesi  $H_0$  con livello di fiducia  $\alpha$
- Se  $D_n < d_{1-\alpha}$  l'ipotesi  $H_0$  viene accettata con livello di fiducia  $\alpha$

- **Test chi-quadro:** Spesso i risultati ottenuti dall'analisi di campioni, non corrispondono con i risultati teorici attesi. Es: lanciando una moneta non truccata 100 volte, teoricamente dovrebbero uscire 50 teste e 50 croci, mentre nella realtà è quasi impossibile che questo accada. Per mezzo del test del Chi-Quadro, si può valutare la discrepanza esistente tra i risultati osservati e quelli teorici. Più precisamente, confronto la distribuzione di dati con una probabilità teorica. Chiameremo **supporto di  $F(t)$**  l'intervallo in cui la funzione di ripartizione è diversa da 0. La formula generale per valutare quantitativamente la bontà dell'adattamento delle frequenze osservate alle frequenze attese è:

$$W = \sum_{k=1}^K \frac{(n_k - np_k)^2}{np_k} = n \sum_{k=1}^K \frac{(\bar{p}_k - p_k)^2}{p_k}$$

- $dn_k$  è la frequenza osservata
- $p_k$  è la probabilità che ogni singolo valore cada in un caso
- $mp_k$  è la frequenza attesa
- $\bar{p}_k = \frac{n_k}{\sum_k n_k} = \frac{n_k}{n}$  è la probabilità empirica.

Formula excel =TEST.CHI(int\_effettivo;int\_previsto) restituisce il p-value il quale se è maggiore di un  $\alpha$  si rifiuta l'ipotesi nulla.

## GENERAZIONE DI NUMERI CASUALI

I generatori di numeri casuali (**RNG** – Random Number Generator) sono fondamentali per varie applicazioni: esperimenti statistici, analisi di algoritmi, simulazione di sistemi stocastici, analisi numerica basata su metodi Monte Carlo, algoritmi probabilistici, computer games (Es: Virtual Casinò), crittografia, protocolli di comunicazione sicura, ...

### Generazione dei numeri casuali con densità di probabilità uniforme

I numeri casuali possono derivare da:

- fenomeni naturali (ad es. tempo di decadimento delle particelle  $\alpha$  nell'uranio)
- algoritmi matematici che simulano i fenomeni naturali

Questi algoritmi hanno carattere **deterministico**, ovvero a parità di seme iniziale la sequenza di output è sempre la stessa. Inoltre la sequenza di numeri generata si ripete con un periodo  $T$ , per questo si parla di numeri pseudocasuali. Un buon RNG deve avere:

1. periodo  $T$  molto grande
2. generazione efficiente dei numeri: veloce e uso di poche risorse
3. sequenza riproducibile
4. portabilità del codice

- **RNG basato su ricorrenze lineari:** dato dalla formula

$$X_{n+1} = (a_0 x_n + a_1 x_{n-1} + \dots + a_j x_{n-j}) \bmod P$$

questo RNG è caratterizzato da un periodo  $T$  che è  $T \leq p^{j+1}$

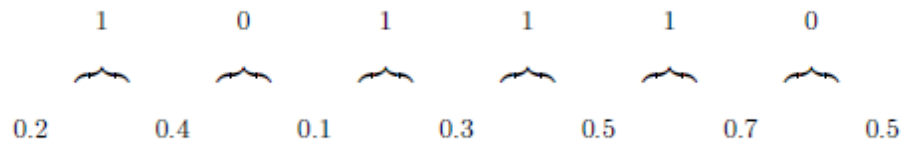
### Test statistici per i numeri casuali

Sono test aventi lo scopo di verificare la corretta generazione dei numeri casuali

- **Uniformità:** divido l'intervallo  $[0,1]$  in  $k$  intervalli disgiunti, l'ipotesi da verificare è che in ogni intervallo cadano lo stesso numero di valori casuali.

- **Test per il gradi di casualità:** si dividono in:

- **Up and down test:** data una sequenza N di numeri casuali, si assegna 1 se il numero incrementa, 0 se decrementa



La sequenza di 1,0 che ho è 1 0 1 1 1 0, ovvero ho

- Run Up di lunghezza 1
- Run down di lunghezza 1
- Run Up di lunghezza 3
- Run down di lunghezza 1

Una buona sequenza di numeri casuali non deve avere una run di numero elevato

- **Gap test:** dati dei numeri casuali  $\{X_i\}$   $i = 1, \dots, N$  diremo che ogni sottosuccessione di  $r+1$  numeri casuali rappresenta un gap o divario di lunghezza  $r$  se

$$\begin{cases} x_j \text{ e } x_{j+r} & \text{stanno tra } \alpha \text{ e } \beta (0 \leq \alpha < \beta \leq 1) \\ x_{j+i} \text{ } i = 1, \dots, r-1 & \text{non stanno in } [\alpha, \beta] \end{cases}$$

Per avere una buona sequenza di numeri casuali si è calcolato che la probabilità di ottenere un gap di lunghezza  $r$  è

$$P(r) = (0.9)^r (0.1)$$

### Generazione di numeri casuali con assegnata densità di probabilità

- **Tecnica diretta:** dato un numero casuale uniforme  $r \in [0,1]$ , in corrispondenza cerco un numero  $x_r \in [a, b]$  tale che

$$r = F(x_r) = \int_a^{x_r} f(t) dt$$

dove  $x_r = F^{-1}(r)$ . Data  $P(x)$  la densità di probabilità associata a  $x$ , avremo

$$P(x) dx = dr = dF = \frac{dF}{dx} = f(x) dx \Rightarrow P(x) = f(x) \quad r = \frac{\int_a^{x_r} f(t) dt}{\int_a^b f(t) dt}$$

- **Tecnica di reiezione:** dati due 2 numeri casuali con distribuzione uniforme in  $[0,1]$  si calcola

$$x_1 = a + (b - a)r_1, \quad f_1 = r_1' C$$

Che sono ancora numeri casuali con distribuzione uniforme; se accade che  $f_1 \leq f(x_1)$  allora si prende  $x_1$  come nuovo numero casuale con distribuzione  $f(x)$ . In pratica si scarta  $x_1$  se cade nella parte che sta fuori della  $f(x)$ , altrimenti si accetta.

- **Tecnica combinata:** Sia  $g(x)$  una seconda densità di probabilità abbastanza semplice per permettere la generazione di numeri casuali distribuita secondo essa e tale che

$$f(x) < k g(x) \quad x \in [a, b]$$

Generiamo  $x_1$  numero casuale con densità di probabilità  $g(x)$  e  $r_1$  con distribuzione uniforme  $\in [0,1]$ ; se  $r_1 k g(x_1) < f(x_1)$  si accetta  $x_1$  altrimenti rigenera  $x_1$ . Al solito si scarta  $x_1$  se cade nella parte evidenziata, ma adesso la zona di scarto è più piccola.

### Metodo Monte Carlo (o metodo MC)

L'algoritmo Monte Carlo è un metodo numerico che viene utilizzato per trovare le soluzioni di problemi matematici, che possono avere molte variabili aleatorie e che non possono essere risolti facilmente, per esempio con il calcolo integrale. L'efficienza di questo metodo aumenta rispetto agli altri metodi quando la dimensione del problema cresce. Fa parte della famiglia dei metodi statistici non parametrici, ed è usato per trarre stime attraverso simulazioni basate su algoritmi.

Non c'è un solo metodo Monte Carlo; tuttavia, tutti gli approcci che adottano MC tendono a seguire un particolare schema:

1. Definire un dominio di possibili dati in input.



2. Generare input casuali dal dominio con una certa distribuzione di probabilità determinate.
3. Eseguire un calcolo deterministico utilizzando i dati in ingresso (input).
4. Aggregare i risultati dei calcoli singoli nel risultato finale.

Es: Il metodo Monte Carlo può essere illustrato come una Battaglia navale. Prima un giocatore fa alcuni colpi a caso. Successivamente il giocatore applica alcuni algoritmi (es. la corazzata è di quattro punti nella direzione verticale o orizzontale). Infine, sulla base dei risultati del campionamento casuale e degli algoritmi il giocatore può determinare le posizioni probabili delle navi degli altri giocatori.

- **Integrazione Monte Carlo:**  $E \propto \frac{1}{\sqrt{n}} \forall d$

Nel caso  $d = 1$  tutti gli algoritmi numerici lavorano meglio del MC, mentre per  $d$  sufficientemente grande il MC ne batte qualcuno

- **Metodo MC Hit or Miss:** sia  $f(x)$  definita in  $[a, b]$  e limitata, cioè  $\exists c \in \mathbb{R} : 0 \leq f(x) \leq c$ , la probabilità  $p$  che il vettore  $(x, y)$  di numeri casuali uniformemente distribuiti in  $\Omega$ , cada sotto la curva  $f(x)$  è

$$p = \frac{\text{area } S}{\text{area } \Omega} = \frac{\int_a^b f(x) dx}{c(b-a)}$$

- Algoritmo MC Hit or Miss:

1. Generare una successione di  $2N$  numeri casuali  $\{U_i\} \ i = 1, 2, \dots, N$
2. Disporre i numeri casuali in  $N$  coppie  $(U_1, U'_1), (U_2, U'_2), \dots, (U_N, U'_N)$  in modo che ogni numero casuale  $\{U_i\}$  è usato una sola volta
3. Calcolare  $x_i = a + U_i(b-a)$  e  $f(x_i) \ i = 1, 2, \dots, N$
4. Contare il n° di volte  $N_H$  in cui  $f(x_i) \geq cU'_i$
5. Stimare l'integrale  $I$  con  $\Theta_1 = c(b-a) \frac{N_H}{N}$
6. Fissato il livello di fiducia  $\alpha$ , calcolare l'intervallo di confidenza con

$$\left[ \Theta_1 - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{N}} (b-a)c, \Theta_1 + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{N}} (b-a)c \right]$$

- **Metodo MC Sample-Man:** si riscrive l'integrale come  $I = \int_a^b \frac{f(x)}{g(x)} g(x) dx$  ne segue che  $I = E \left[ \frac{f(x)}{g(x)} \right]$  dove  $x$  è una variabile casuale distribuita con funzione di distribuzione  $g(x)$ . Assumiamo quindi

$$g(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0 & \text{altrimenti} \end{cases}$$

- Algoritmo MC Sample-Mean:

1. Generare una sequenza  $\{U_i\}$  di  $N$  numeri casuali uniformi  $\in [0, 1]$
2.  $x_i = a + U_i(b-a) \ i = 1, \dots, N$
3. Calcolare  $f(x_i)$  e quindi

$$\Theta_2 = (b-a) \frac{1}{N} \sum_{i=1}^N f(x_i)$$

- **Tecniche di riduzione della varianza:** Esistono diverse tecniche per ridurre la varianza, ma in generale vale la regola che si può ridurre la varianza solo se si hanno informazioni sul problema da trattare.

- **Importance sampling:** l'idea principale di questa tecnica consiste nel concentrare la distribuzione dei punti del campione nella parte di  $D$  che è più importante invece di distribuirli uniformemente in  $D$ .

L'integrale si può quindi scrivere:  $I = \int \frac{f(x)}{g(x)} g(x) dx = E \left[ \frac{f(x)}{g(x)} \right]$ , quindi la stima di  $I$  è

$$\xi = \frac{f(x)}{g(x)}, \quad E[\xi] = I \quad \text{con varianza} \quad V[\xi] = \int \frac{|f(x)|^2}{g(x)} dx - I^2 \quad \text{calcolato con la formula}$$

$$\Theta_3 = \frac{1}{N} \sum_{i=1}^N \frac{f(x_i)}{g(x_i)}$$

- **Control variates:** in questo caso si cerca di migliorare la varianza confrontando con qualche modello analitico noto. Siano  $Y$  e  $C$  due variabili random tali che  $E[C] = \mu_C$  allora  $\forall \beta \in \mathbb{R}$  definisco

$$Y(\beta) = Y - \beta(C - \mu_C) = Y + \beta(\mu_C - C) \quad : \quad E[Y(\beta)] = E[Y]$$

- **Stratified sampling:** Si suddivide la regione di integrazione  $D$  in  $m$  sotto regioni disgiunte  $D_i$ , cioè

$$D_i \cap D_j = \emptyset \quad i \neq j \Rightarrow D = \bigcup_{i=1}^m D_i$$

E quindi 
$$I = \int_D \phi(x)g(x)dx = \sum_{i=1}^m \int_{D_i} \phi(x)g(x)dx$$

- **Antithetic variates:** Con questa tecnica si cercano due stimatori unbiased dell'integrale  $I$ , tale che abbiano una forte correlazione negativa. Siano  $Y'$  e  $Y''$  due stimatori unbiased di  $I$  cioè  $E[Y'] = E[Y''] = I$ ; calcolando la varianza ad essi e troviamo

$$V\left[\frac{1}{2}(Y' + Y'')\right] = \frac{1}{4}V[Y' + Y''] = \frac{1}{4}V[Y'] + \frac{1}{4}V[Y''] + \frac{1}{2}\text{Cov}(Y', Y'')$$

Se  $\text{Cov}(Y', Y'') < 0$  allora la varianza diminuisce.

## CATENE DI MARKOV

Un processo di Markov è un processo stocastico (cioè che varia nel tempo in modo casuale) nel quale la probabilità che determina il passaggio ad uno stato di sistema dipende unicamente dallo stato di sistema immediatamente precedente (proprietà di Markov) e non dal come si è giunti a tale stato (in quest'ultima ipotesi si parla di processo non markoviano); lo stato futuro del sistema dipende solo da quello presente e non dalla sua storia.

- **Proprietà di Markov:** dato lo stato corrente  $X_t$ , la distribuzione di ogni stato futuro  $X_y$  ( $y > t$ ) non dipende dalla sua storia passata  $\{X_u : u < t\}$  ma solo da quello presente (linea temporale  $X_u \rightarrow X_t \rightarrow X_y$ )
- **Spazio degli stati:** è l'insieme dei valori presi dalle variabili casuali  $X_t$
- **Catena di Markov a tempo discreto (DTCM):** Noi saremo interessati a studiare processi stocastici con spazio degli stati discreti  $(0, 1, 2, \dots)$  ed in cui anche il parametro  $t \in \mathbb{N}$  sia discreto. In tal caso il processo sarà detto DTMC. Quindi se il sistema si trova nello stato  $n$ , il suo stato al tempo  $n+1$  non dipende dagli stati  $0, 1, 2, \dots, n-1$

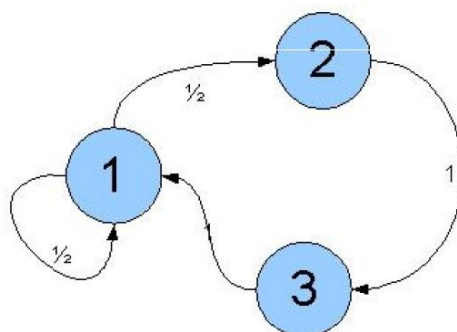
$$P(X_{n+1} = j | X_0, X_1, \dots, X_n) = P(X_{n+1} = j | X_n)$$

- **Probabilità di transizione ad un passo:** descrive l'evoluzione dello schema attraverso una matrice  $Q$  i cui elementi  $q_{i,j}(n)$  ci dicono che il sistema si muoverà nello stato  $j$  al tempo successivo  $n+1$ , supposto che al tempo  $n$  si trovi nello stato  $i$

$$q_{i,j}(n) = P(X_{n+1} = j | X_n = i), \quad i, j, n = 0, 1, 2, \dots$$

- **Definizione temporale omogenea:** se questa probabilità di transizione non dipende dall'istante di tempo, cioè  $q_{i,j}(n) = q_{i,j}$ ,  $i, j, n = 0, 1, \dots$   
Poiché il sistema deve essere in qualche stato ad ogni tempo di osservazione, ne segue che la somma degli elementi di ogni riga della matrice sia uguale a uno.

Un modo per rappresentare una catena di Markov è con un **grafo orientato** (o diretto) in cui i vertici corrispondono agli stati della catena e gli archi mostrano le probabilità di transizione non nulle.



## Calcolo leggi congiunte

Consideriamo adesso la transizione fatta dalla catena di Markov in due passi. Poiché la catena è temporalmente omogenea, le corrispondenti probabilità non dipendono dall'indice temporale. Definiamo

$$q_{i,j}^2 = P(X_{n+2} = j | X_n = i), i, j, n = 0, 1, \dots$$

Se ci vogliamo muovere da uno stato  $i$  ad uno  $j$  in due passi, la catena deve transitare da uno stato intermedio  $k$ , dopo il primo passo. Quindi per la proprietà di Markov diventa

$$q_{i,j}^2 = \sum_{k=0}^{\infty} q_{i,k} q_{k,j}$$

- **Teorema probabilità di transizione da uno stato  $i$  ad uno  $j$  ad  $s$  passi:** in generale vale che

$$q_{i,j}^s = P(X_{n+s} = j | X_n = i), s = 1, 2, \dots$$

Allora la matrice di transizione  $Q^{(s)}$  i cui elementi sono  $q_{i,j}^{(s)}$  è data da  $Q^{(s)} = Q^s$

## Classificazione degli stati

Sia  $E$  l'insieme degli stati di una catena di Markov e  $C \subset E$

- Se  $i, j \in E$  diciamo che  $i$  **comunica con**  $j$  (ovvero  $i \rightarrow j$ ) se esiste  $n > 0$  tale che  $q_{i,j}^n > 0$ . Questo equivale al fatto che esista un percorso che porta dallo stato  $i$  allo stato  $j$  in  $n$  passi.
- Un sottoinsieme di stati  $C$  si dice **chiuso** se gli stati di  $C$  non comunicano con gli stati che stanno fuori di  $C$  (ovvero del complementare di  $C$ ).
- Un sottoinsieme di stati  $C$  chiuso si dice **irriducibile** se tutti i suoi stati comunicano tra loro.
- Una catena di Markov si dice **irriducibile** se tutti i suoi stati comunicano tra loro, ovvero se  $E$  è l'unica classe irriducibile
- Uno stato di una catena di Markov si dice **assorbente** se costituisce da solo una classe irriducibile.
- Uno stato  $i$  di una catena di Markov si dice **ricorrente** se il sistema essendo stato una volta in  $i$ , la catena ritornerà in  $i$  con probabilità uno. Se indichiamo con  $p_{ii}$  questa probabilità, avremo che  $p_{ii} = 1$ .
- Un stato  $i$  di una catena di Markov che non è ricorrente si dice **transitorio**, e in questo caso  $p_{ii} < 1$
- Uno stato  $j$  di una catena di Markov si dice **periodico** con periodo  $m > 1$ , se i ritorni consecutivi allo stato  $j$  avvengono solamente con multipli di  $m$  passi, cioè

$$P(X_{n+s} = j | X_n = j) = 0, \text{ se } s \neq mk \text{ per qualche } k \geq 1$$

Se non esiste qualche  $m > 1$  che soddisfi la relazione precedente, la catena si dirà **aperiodica**.

## Probabilità invarianti

- Sia  $v$  una probabilità su  $E$ , allora essa si dice **invariante o stazionaria** se accade che  $v = vQ$
- Una matrice di transizione  $Q$  si dice **regolare** se esiste un intero positivo  $m$  tale che  $q_{ij}^m > 0, \forall i, j \in E$
- Una catena di Markov si dice regolare se è tale la sua matrice di transizione.
- **Criterio di regolarità:** se tutti gli stati comunicano tra loro e se inoltre esiste  $h \in E$  tale che  $q_{hh} > 0$  allora la catena è regolare.

## Stato stazionario

Vogliamo studiare una catena di Markov per tempi molto grandi. Quando l'istante di osservazione è molto lontano dal punto iniziale, la probabilità di trovare la catena in uno stato  $j$ ,  $p_j$ , non dipende dallo stato iniziale

$$\lim_{n \rightarrow \infty} P(X_n = j | X_0 = i) = \lim_{n \rightarrow \infty} q_{i,j}^n = p_j, j = 0, 1, \dots$$

Quando le probabilità limite  $p_j$  esistono (e la loro somma fa uno), esse si chiamano **distribuzione di equilibrio o distribuzioni dello stato stazionario** di una catena di Markov. Il problema principale della teoria delle catene di Markov è di sapere se esiste una distribuzione di equilibrio ed in tal caso determinarla.

Data una catena di Markov il cui insieme degli stati sia finito e pari a  $N$ , se inoltre la matrice di transizione  $Q$  è regolare, allora esiste ed è unica la probabilità invariante  $p = (p_1, p_2, \dots, p_N)$  di  $Q$  tale che

$$\lim_{n \rightarrow \infty} q_{i,j}^n = p_j$$

## Algoritmo di Metropolis

- **Bilancio dettagliato:** una probabilità  $p$  su  $E$  si dice **reversibile** se accade che  $p_i q_{ij} = p_j q_{ji}, \forall i, j \in E$
- Se  $p$  è reversibile allora è anche invariante.
- **Teorema di Metropolis, Rosenbluth e Teller:** se  $X_n$  è una catena di Markov con matrice di transizione  $P$ , allora  $X_n$  converge a  $\pi$  per  $n \rightarrow \infty$ .