SPAZI CAMPIONARI ED EVENTI

Parliamo di probabilità quando non abbiamo tutti gli eventi per descrivere deterministicamente un determinato evento.

DEF: Chiameremo esperimenti i fenomeni aleatori di cui ci occupiamo e chiameremo esiti i risultati di tali esperimenti.

DEF: Dato un esperimento, è detto spazio campionario l'insieme di tutti gli esiti possibili.

Con Ω indichiamo il generico spazio campionario e con ω i generici esiti.

Esempio:

- il lancio di una moneta: $\Omega = \{\text{Testa}, \text{Croce}\}$
- il lancio di un dado: $\Omega = \{1, 2, 3, 4, 5, 6\}$

Abbiamo che i risultati dell'esperimento associano un insieme e gli esiti di un esperimento sono gli elementi dell'insieme.

RISULTATI $\longleftrightarrow$ INSIEME , ESITI $\longrightarrow$ ELEMENTI INSIEME

DEF: Gli eventi sono sottoinsiemi di Ω

Esempio:

$\Omega = \{1, 2, 3, 4, 5, 6\}$

ci interessa la probabilità che nei primi 2 lanci escano il 3 e il 5, consideriamo dunque $\Omega' = \{3, 5\}$ (evento).

- Ω: evento certo (insieme di partenza)
- $\emptyset$: evento impossibile.

DEF Due eventi si dicono mutuamente esclusivi se:

$E \cap F = \emptyset$ con $E \subseteq \Omega$, $F \subseteq \Omega$.

Questi eventi non possono accadere contemporaneamente.

CORRISPONDENZE

- Ω : spazio campionario certo;
- ω : evento elementare / esito;
- E : si verifica uno degli esiti contenuti in E / si verifica l'evento E.
  [l'evento può essere un solo elemento o l'intero sottoinsieme]
- $\bar{E}$: non si verifica uno degli esiti contenuti in E, oppure non si verifica l'evento E.
  [non pescare una carta rossa quando E = pescare una carta rossa]
- $E \cap F$: si verificano entrambi gli eventi E ed F.
  [se E = rosso e F = fiori $\Rightarrow E \cap F = \emptyset$]
- $E \cup F$: si verifica almeno uno degli eventi E ed F, o entrambi.
- $E \setminus F$: si verifica E e non F
  [E : carta rossa, F : carta pari, $E \setminus F$ = carta rossa dispari]
- $E \subseteq F$: se si verifica E, allora si verifica F ($E \Rightarrow F$)
  [E : cuori, F : carta rossa $\Rightarrow$ se si verifica E allora anche F]
- $E \cap F = \emptyset$: eventi mutuamente esclusivi, se si verifica uno dei due, è impossibile che si verifichi l'altro.
- $\emptyset$: evento impossibile
- Ω : insieme degli esiti
- $\mathcal{F}$: insieme degli eventi / famiglia degli eventi, è l'insieme formato dai sottoinsiemi di Ω

REQUISITI PER $\mathcal{F}$

1. $\Omega \in \mathcal{F}$ (Ω è un evento)
2. $E \in \mathcal{F} \Rightarrow \bar{E} \in \mathcal{F}$ (pescare una carta e la carta che non è pescata)
3. $E, F \in \mathcal{F} \Rightarrow E \cup F \in \mathcal{F}$

$E \cap F$ non è tra i requisiti perché è ricavabile a partire da ① e ②

DIM ($E \cap F \in \mathcal{F}$)

$$E, F \in \mathcal{F} \overset{②}{\Longrightarrow} \bar{E}, \bar{F} \in \mathcal{F} \overset{③}{\Longrightarrow} \bar{E} \cup \bar{F} \in \mathcal{F} \overset{②}{\Longrightarrow} \overline{\bar{E} \cup \bar{F}} \in \mathcal{F}$$

$\{ \overline{E \cup F} = \bar{E} \cap \bar{F} \}$ De morgan

$$\Rightarrow \overline{\bar{E} \cup \bar{F}} = E \cap F \in \mathcal{F}$$

DEF Un insieme $\mathcal{F}$ che soddisfa i requisiti prende il nome di Algebra

REQUISITI PER $\mathcal{F}$ ESTENSIONE

t_n = all'n-esimo lancio esce testa

C_n = all' n-esimo lancio esce croce

E_n = è uscita testa per la prima volta all'n-esimo lancio.

$E_n = C_1 \cap C_2 \cap C_3 \cap \dots \cap t_n$

1. $\Omega \in \mathcal{F}$

2. $E \in \mathcal{F} \Rightarrow \bar{E} \in \mathcal{F}$

3. $E_n \in \mathcal{F} \ \forall n \Rightarrow \bigcup_{n=1}^{\infty} E_n \in \mathcal{F}$

DEF Un insieme $\mathcal{F}$ che soddisfa l'estensione prende il nome di σ-algebra.

Dato Ω, la più piccola σ-algebra è $\mathcal{F} = \{\emptyset, \Omega\}$ mentre la più estesa è $\mathcal{P}(\Omega)$ [insieme delle parti].

DEF Dato un insieme Ω e dati alcuni suoi sottoinsiemi $E_1, E_2, \dots, E_n$ si dice σ-algebra generata dagli E_i la più piccola σ-algebra di cui faranno parte tutti gli E_i

PROBABILITÀ

Esistono diverse definizioni di probabilità, quella più semplice è la definizione classica e ci dice che:

$$\boxed{\cdot\ P = \frac{\#\text{casi favorevoli}}{\#\text{casi possibili}}}$$

Questa definizione però presuppone che conosciamo il sistema, ovvero che tutti gli eventi possibili siano equiprobabili

Un'altra definizione è la definizione frequentista e a differenza della prima (fatta a priori) è fatta a posteriori. La definizione frequentista ci dice che:

$$\boxed{\cdot\ P = \frac{\#\text{successi}}{\#\text{prove effettuate}}}$$

L'ultima definizione è la definizione assiomatica di Kolmogorov

DEF

Sia Ω uno spazio campionario e $\mathcal{F}$ una σ-algebra

- Assioma 1: Per ogni evento $E \in \mathcal{F}$

esiste un num. $P[E] \in \mathbb{R}$, $0 \leq P[E] \leq 1$

- Assioma 2: Per l'intero spazio campionario

$$P[\Omega] = 1$$

- Assioma 3: Presa una qualsiasi successione di eventi tali che $E_i \cap E_j = \emptyset$

se $i \neq j$, $P[\bigcup_i E_i] = \sum_i P[E_i]$

La probabilità è una corrispondenza

$$\boxed{P : \mathcal{F} \longrightarrow [0, 1]}$$

che soddisfa i precedenti assiomi.

Quindi P è una funzione definita in $\mathcal{F}$ a valori in $[0, 1]$

$(\Omega, \mathcal{F}, P)$ è detto spazio delle probabilità

Dati questi assiomi è possibile ricavare quello che ci serve

- Proposizione 1:

Dato $P[E]$, $P[\bar{E}] = ?$

DIM $P[\Omega] = 1$

$\Omega = E \cup \bar{E}$, $E \cap \bar{E} \neq \emptyset$ [non hanno elementi in comune]

$\Rightarrow P[E \cup \bar{E}] = 1 \Rightarrow P[E \cup \bar{E}] = P[E] + P[\bar{E}]$

$\Rightarrow P[E] + P[\bar{E}] = 1 \Rightarrow \boxed{P[\bar{E}] = 1 - P[E].}$

- Proposizione 2:

Se $F \subseteq E \Rightarrow P[F] \leq P[E]$ (se si verifica si verifica F)

DIM

Se $F \subseteq E \Rightarrow E = F \cup (E \setminus F)$

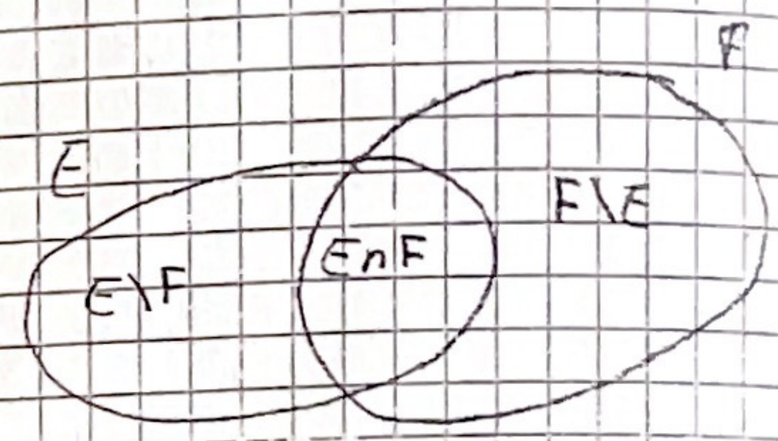
$\Rightarrow P[E] = P[F \cup (E \setminus F)]$

$\overset{A_3}{\Rightarrow} P[F] + P[E \setminus F] \geq P[F]$

- Proposizione 3

$P[E \cup F] = P[E] + P[F] - P[E \cap F]$

DIM



- $P[E] = P[(E \setminus F)] + P[E \cap F]$

$\Rightarrow P[E \setminus F] = P[E] - P[E \cap F]$

- $E \cup F = (E \setminus F) \cup (E \cap F) \cup (F \setminus E)$

$\Rightarrow P[E \cup F] = P[E \setminus F] + P[E \cap F] + P[F \setminus E] =$

$= P[E] - \cancel{P[E \cap F]} + \cancel{P[E \cap F]} + P[F] - P[E \cap F] =$

$= P[E] + P[F] - P[E \cap F]$

- Proposizione 4 (n = numero di eventi)

$$P\left[\bigcup_{i=1}^{n} E_i\right] = \sum_i P[E_i] - \sum_{i<j} P[E_i \cap E_j] + \dots + (-1)^{n+1} P[E_1 \cap E_2 \cap \dots \cap E_n]$$

PROBABILITÀ CONDIZIONATA

La probabilità è legata all'ignoranza che abbiamo rispetto al nostro sistema, ogni volta che aggiungiamo informazioni al nostro sistema, qualcosa cambierà sempre.

DEF Siano E ed F due eventi.

Si chiama probabilità di E dato F e scriveremo $P[E|F]$ la probabilità che si verifichi l'evento E supponendo che si sia verificato l'evento F.

$$\boxed{P[E|F] = \frac{P[E \cap F]}{P[F]}, \quad P[F] > 0}$$

Segue:

- $P[E \cap F] = P[F] \cdot P[E|F]$

- $P[E \cap F] = P[E] \cdot P[F|E]$

TEOREMA: Regola della catena

Dati n eventi $E_1, E_2, \dots, E_n$ la cui intersezione ha probabilità positiva ($P[E_1 \cap E_2 \cap \dots \cap E_n] > 0$), si ha che:

$$P[E_1 \cap E_2 \cap \dots \cap E_n] = P[E_1] \cdot P[E_2 | E_1] \cdot P[E_3 | E_2 \cap E_1] \cdot \dots \cdot P[E_n | E_1 \cap E_2 \cap \dots \cap E_{n-1}]$$

DEF

Si dice partizione di Ω un insieme di eventi E_i (finito o numerabile) con le seguenti caratteristiche:

- 1. Ogni E_i ha probabilità diversa da $\emptyset$;

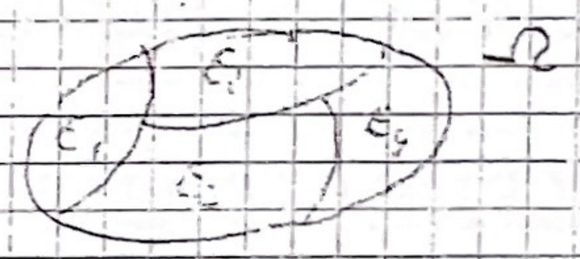
  $0 < P[E_i] \leq 1$

- 2. Gli E_i sono mutuamente esclusivi;

  $P[E_i \cap E_j] = 0 \quad , \quad i \neq j$

- 3. L'unione di tutti gli E_i copre Ω;

  $\bigcup_{i=1}^{n} E_i = \Omega$



TEOREMA: Formula delle probabilità totali o legge delle alternative

Sia $\{E_i\}$ una partizione di Ω

Sia F un evento qualsiasi contenuto in Ω, allora:

$$\boxed{P[F] = \sum_{i=1}^{n} P[F | E_i] \cdot P[E_i]}$$

DIM

- $P[F \cap E_i] = P[F | E_i] \cdot P[E_i]$

- $F = F \cap \Omega = F \cap \left(\bigcup_i E_i\right) = \bigcup_i (F \cap E_i)$

$\Rightarrow P[F] = P\left[\bigcup_i (F \cap E_i)\right] = \sum_i P[F \cap E_i] = \sum_i P[F | E_i] \cdot P[E_i]$

ESEMPIO

Le lampadine di una certa marca vengono prodotte in 3 stabilimenti distinti A, B, C.

- A produce il 50%
- B produce il 30%
- C produce il 20%

L'1% delle lampadine prodotte da A è difettoso mentre le lampadine difettose prodotte da B e C sono rispettivamente 5% e 10%.
Infine tutte le lampadine vengono raggruppate in un unico stabilimento.
Prendendo a caso una lampadina, qual è la probabilità che sia difettosa?

Dati:

$P[A] = 50\%$, $P[B] = 30\%$, $P[C] = 20\%$

$P[F|A] = 1\%$, $P[F|B] = 5\%$, $P[F|C] = 10\%$, con F = lampadine difettose

$$P[F] = P[F|A] \cdot P[A] + P[F|B] \cdot P[B] + P[F|C] \cdot P[C] =$$

$$= 0.01 \cdot 0.5 + 0.05 \cdot 0.3 + 0.1 \cdot 0.3 = 0.04 = 4\%$$

Applicazione della legge delle alternative

TEOREMA: Legge condizionata o delle alternative

Sia $\{E_i\}$ una partizione di Ω e siano G ed F due eventi qualsiasi, allora

$$\boxed{P[F|G] = \sum_i P[F|E_i \cap G] \cdot P[E_i|G]}$$

Dove la sommatoria è estesa unicamente ai casi $P[E_i \cap G] \neq 0$.

TEOREMA: Formula di Bayes

Dato un evento F e data una partizione $\{E_i\}$ di Ω si ha:

$$\boxed{P[E_h|F] = \frac{P[F|E_h] \cdot P[E_h]}{\sum_i P[F|E_i] \cdot P[E_i]}}$$

DIM

- $P[E_h|F] = P[F|E_h] \cdot P[E_h] = P[E_h|F] \cdot P[F]$
- $P[E_h|F] = \dfrac{P[F|E_h] \cdot P[E_h]}{P[F]} = \dfrac{P[F|E_h] \cdot P[E_h]}{\sum_i P[F|E_i] \cdot P[E_i]}$

EVENTI INDIPENDENTI

osservazioni:

- $P[E|F] > P[E]$: l'evento F è favorevole ad E
- $P[E|F] < P[E]$: l'evento F è sfavorevole ad E
- $P[E|F] = P[E]$: l'evento F è ininfluente oppure indipendente da E.

PROPRIETA'

Se E è indipendente da F, allora F è indipendente da E. $P[E|F] = E \Rightarrow P[F|E] = P[F]$

DIM

- $\underline{P[E \cap F]} = P[E|F] \cdot P[F] = \underline{P[E] \cdot P[F]}$
- $\underline{P[E \cap F]} = \underline{P[F|E] \cdot P[E]}$

} stessa cosa

$\Rightarrow \underline{P[F|E] \cdot \cancel{P[E]}} = \underline{\cancel{P[E]} \cdot P[F]} \Rightarrow$

$\Rightarrow P[F|E] = P[F]$

allora

$P[E \cap F] = P[E] \cdot P[F]$ se E ed F sono indipendenti

perché: $P[E \cap F] = \underset{\substack{\parallel \\ P[E]}}{P[E|F]} \cdot P[F]$

Proposizione

Se E ed F sono indipendenti, allora sono indipendenti anche E ed $\bar{F}$.

DIM

$$E = (E \cap F) \cup (E \cap \bar{F})$$

$$P[E] = P[E \cap F] + P[E \cap \bar{F}] \overset{\text{ind}}{=} P[E] \cdot P[F] + P[E \cap \bar{F}]$$

- $P[E \cap \bar{F}] = P[E] - P[E]P[F] = P[E](1 - P[F]) = P[E] \cdot P[\bar{F}]$.

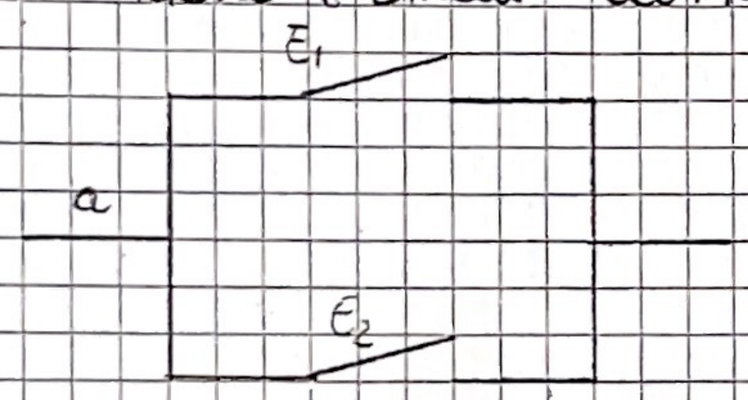
OSSERVAZIONE (indipendenza tra più eventi)

- $P[E \cap F \cap G] = P[E] \cdot P[F] \cdot P[G]$ eventi indipendenti

DEF

Una famiglia di n eventi $E_1, E_2, \dots, E_n$ si dice famiglia di eventi indipendenti se gli eventi $\{E_i\}$ sono a due a due e a tre a tre e così via fino a n ad n indipendenti.

OSSERVAZIONE (circuiti elettrici)



- E_1, E_2 = interruttore chiuso
- A = evento passaggio di corrente
- P_1 = probabilità che passi corrente in 1
- P_2 = probabilità che passi corrente in 2

$$A = E_1 \cup E_2 \quad , \quad \bar{A} = \bar{E}_1 \cap \bar{E}_2$$

$$P[A] = 1 - P[\bar{A}] = 1 - P[\bar{E}_1] \cdot P[\bar{E}_2] = 1 - (1 - P_1)(1 - P_2) = P_1 + P_2 - P_1 P_2$$

CASO 2:

$$P[A] = P[E_1 \cap E_2] = P[E_1] \cdot P[E_2] = P_1 \cdot P_2$$

RICHIAMI AL CALCOLO COMBINATORIO

DEF Si dice che n oggetti sono ordinati (o disposti) in un allineamento quando sono collocati in n posti numerati da 1 a n.

Esempi:

- $a_1, a_2, \dots, a_n$ è un allineamento
- $a_{h_1}, a_{h_2}, \dots, a_{h_n}$ dove $h_1, h_2, \dots, h_n$ sono diverse permutazioni degli indici.

DEF (Permutazione)
Si dice permutazione di n oggetti distinti ogni allineamento degli oggetti stessi: due permutazioni sono distinte quando differiscono per il posto occupato da almeno un oggetto.
Indicheremo con P_n il numero di permutazioni.

TEOREMA

$$P_n = n! \quad , \text{dove } n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot (n-1)n.$$

TEOREMA
Il numero delle permutazioni distinte di n oggetti dei quali q_1 uguali tra loro, q_2 uguali tra loro e distinti dai precedenti, ... e q_h uguali tra di loro e distinti dai precedenti (dove $q_1 + q_2 + \dots + q_n = n$), è:

$$\boxed{P^*_{q_1, q_2, \dots, q_h} = \frac{n!}{q_1! \cdot q_2! \cdot \dots \cdot q_n!}}$$

OSSERVAZIONE
Se gli n oggetti sono solo di due tipi $(k, n-k)$

$$\Rightarrow \quad \boxed{P^*_{k, n-k} = \frac{n!}{k!\,(n-k)!} = \binom{n}{k}}$$

DEF Si dice disposizione semplice di n oggetti di classe k ogni allineamento di k oggetti scelti fra gli n.
Indicheremo le disposizioni con $D_{n,k}$

TEOREMA
$D_{n,k}$ è il prodotto di k interi consecutivi il cui massimo è n.

$$\boxed{D_{n,k} = n(n-1)\cdot\ldots\cdot(n-k+1) = \frac{n!}{(n-k)!}}$$

DEF Si dice disposizione con ripetizione di n oggetti di classe k ogni allineamento di k oggetti scelti fra gli n con la convenzione che ogni oggetto può essere ripetuto una o più volte.

Si dimostra che:

$$\boxed{D_{n,k} = n^k}$$

DEF Si dice combinazione semplice di n oggetti di classe k ogni raggruppamento di k oggetti comunque scelti fra gli n.

TEOREMA

$$\boxed{C_{n,k} = \binom{n}{k} = \frac{n(n-1)\cdot\ldots\cdot(n-k+1)}{k!}}$$

DEF Si dice combinazione con ripetizione di n oggetti di classe k ogni raggruppamento di k oggetti comunque scelti tra gli n assegnati, con la convenzione che un oggetto può essere ripetuto più volte.

TEOREMA

$$\boxed{C^*_{n,k} = \binom{n+k-1}{k}}$$

VARIABILI ALEATORIE

Consideriamo lo spazio di probabilità $(\Omega, \mathcal{F}, P)$

X è una funzione (variabile aleatoria) tale che:

$$\boxed{X : \Omega \longrightarrow \mathbb{R} \text{ ovvero } X(\omega) = x \ , x \in \mathbb{R}}$$

con

$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ che può essere: discreto, infinito numerabile o continuo.

DEF Dato uno spazio di probabilità $(\Omega, \mathcal{F}, P)$, si dice variabile aleatoria una corrispondenza tra gli elementi di Ω e i numeri (reali), tale corrispondenza deve soddisfare la condizione che:

$$\boxed{A_t = \{\omega \mid X(\omega) \leq t\} \in \mathcal{F} \quad , \forall t \in \mathbb{R}}$$

È detto supporto l'insieme dei valori che la variabile aleatoria può assumere.

DEF (funzione di ripartizione)

Data la variabile aleatoria X, è definita funzione di ripartizione di X la funzione:

$$\boxed{F_X(t) = P[X \leq t] = P[A_t]}$$

con

$$\boxed{F_X(t) : \mathbb{R} \longrightarrow [0, 1]}$$

PROPRIETA' DELLA FUNZIONE DI RIPARTIZIONE

1. F_X è monotona decrescente, cioè:

$$F_X(t_1) \leq F_X(t_2) \quad \text{se } t_1 \leq t_2$$

DIM

$$A_{t_1} \subseteq A_{t_2} \Rightarrow P[A_{t_1}] \leq P[A_{t_2}]$$

2. $\lim_{t \to -\infty} F_X(t) = 0$, $\lim_{t \to +\infty} F_X(t) = 1$

3. Per ogni valore $x_0 \in \mathbb{R}$

$$\Rightarrow P[X = x_0] = \lim_{t \to x_0^+} F_X(t) - \lim_{t \to x_0^-} F_X(t)$$

4. La funzione di ripartizione F_X è continua a destra, ossia:

$$\forall x_0 \in \mathbb{R} \Rightarrow F_X[x_0] \left(= P[X \leq x_0]\right) = \lim_{t \to x_0^+} F_X(t)$$

DEF Se la variabile aleatoria X può assumere solo un numero finito (o al più numerabile) di valori, la relativa F_X sarà costante a tratti e la variabile viene detta discreta.

Se X può assumere i valori di un intervallo $I \subseteq \mathbb{R}$, allora F_X è generalmente continua e la variabile è detta continua

DENSITÀ DI VARIABILI ALEATORIE

1. Caso discreto

Una variabile aleatoria X discreta può assumere solo alcuni valori x_i $(i = 1, \dots, n \; ; \; i = 1, \dots)$

$$\boxed{P[X = x_i] \rightarrow \sum_i P[X = x_i] = 1}$$

DEF (Densità di probabilità)

La funzione:

$$\boxed{p_X : \mathbb{R} \rightarrow [0, 1]}$$

definita dalla relazione

$$\boxed{p_X(x) = P[X = x] = \begin{cases} P[X = x_i] & \text{se } x = x_i \text{ per un certo } i \\ 0 & \text{se } x \neq x_i \;\; \forall i \end{cases}}$$

IMPORTANZA DEL SEGNO DI INCLUSIONE

Supponiamo di voler calcolare la probabilità che la variabile aleatoria discreta assuma valori tra a e b.

- $P[a < X \leq b] = ?$
- $P[a < X \leq b] = \sum_i P[X = x_i]$

Possiamo scrivere meglio definendo l'insieme A:

$$A = \{ x_i \mid b < x_i \leq a \}$$

$$\Rightarrow P[a < X \leq b] = \sum_{x_i \in A} P[X = x_i] = \sum_{x_i \in A} p_X(x_i)$$

OSS.

$$P[a < X \leq b] = P[a < X < b] + P[X = b]$$

Dunque avremo che nel caso discreto in generale:

$$\boxed{P[a < X \leq b] \neq P[a < X \leq b]}$$

Nel caso discreto:

$$F_X(t) = P[X \leq t] = \sum_{x_i \leq t} p_X(x_i)$$

osserviamo anche che:

$$P[X \leq x_i] = P[X < x_i] + P[X = x_i]$$

esempio

Sia X la variabile aleatoria "numero di occasioni in cui è uscita testa dopo due lanci della moneta".

$\Omega = \{(T,T); (T,C); (C,T); (C,C)\}$

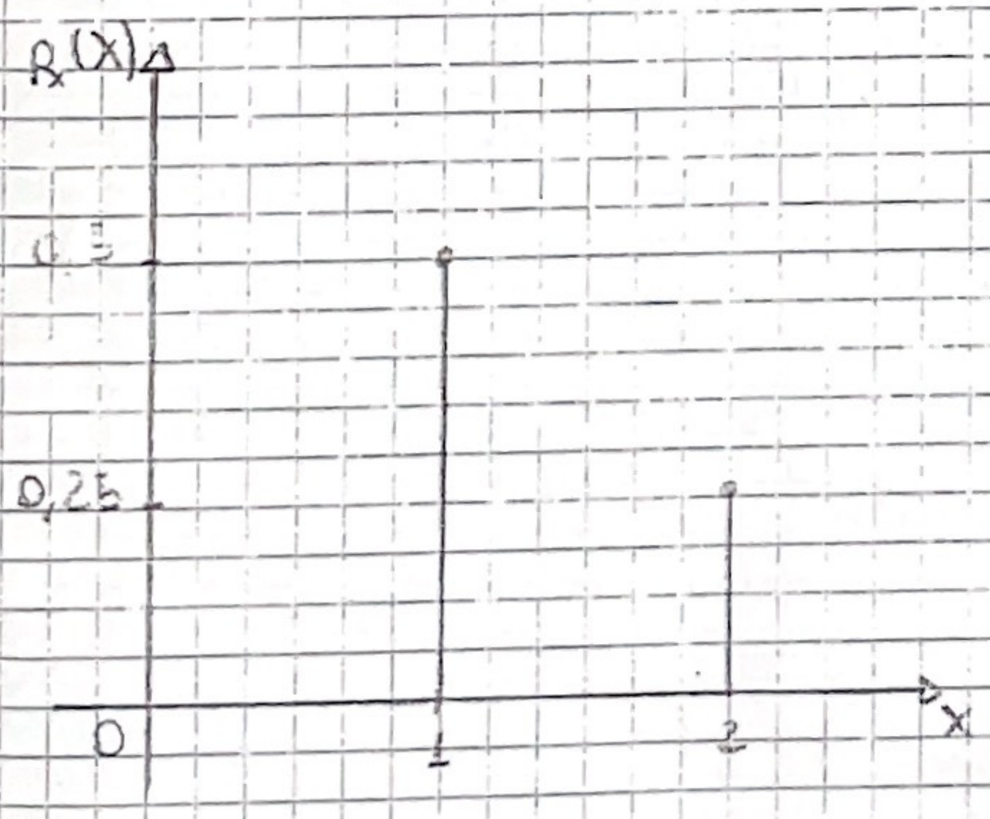
$X : \Omega \to \mathbb{R}$

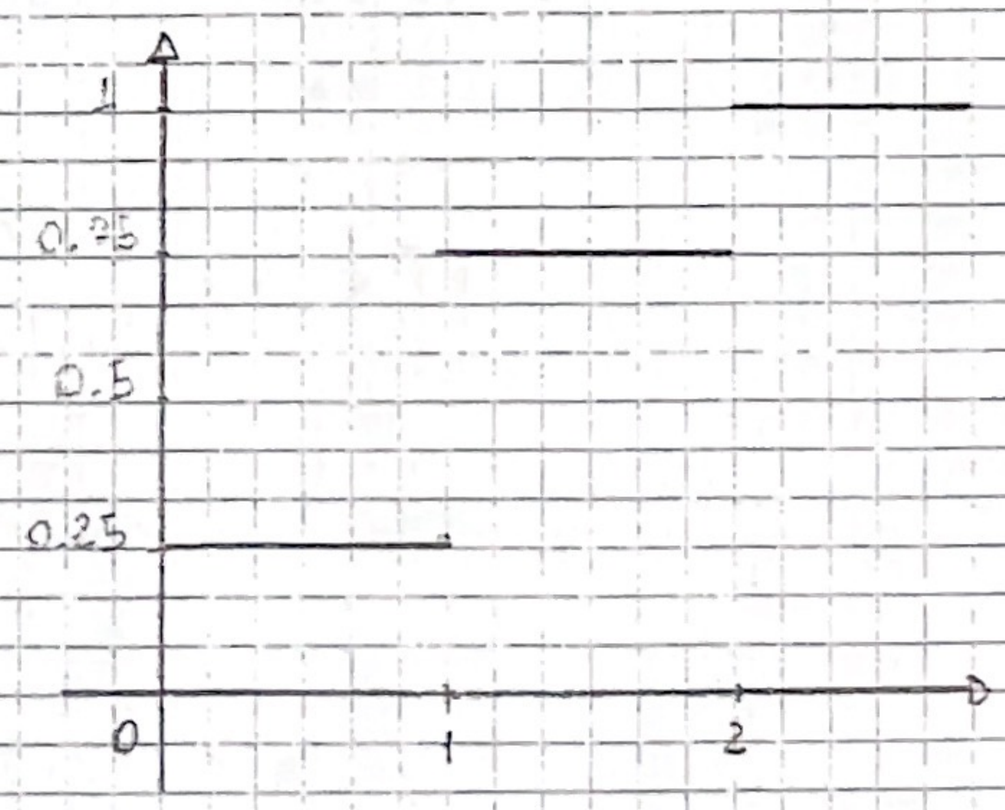
- $X((T,T)) = 2$
- $X((T,C)) = X((C,T)) = 1$
- $X((C,C)) = 0$
- $P[X=2] = \frac{1}{4}$; $P[X=1] = \frac{1}{2}$; $P[X=0] = \frac{1}{4}$

Dimenticandoci di Ω:

- $P[-2 < X < 0.5] = P[X=0] = \frac{1}{4}$
- $P[-2 \leq X \leq 0.5] = P[X=0] = \frac{1}{4}$
- $P[0 < X \leq 2] = P[X=2] + P[X=1] = \frac{3}{4}$
- $P[0 < X < 2] = P[X=1] = \frac{1}{2}$

$$P[X \leq t] = F_X(t) = \begin{cases} 0 & \text{per } t < 0 \\ 1/4 & \text{per } 0 \leq t < 1 \\ 3/4 & \text{per } 1 \leq t < 2 \\ 1 & \text{per } t \geq 2 \end{cases}$$





funzione di ripartizione

2. Caso continuo

In questo caso, il supporto della variabile aleatoria sarà continuo.
Per definire la probabilità consideriamo la funzione di ripartizione
$(F_X(a) = P[X \leq a])$.

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

La funzione $f_X(x)$ è detta funzione densità.

TEOREMA

$$\cdot\ P[a \leq X \leq b] = F_X(b) - F_X(a) = \int_a^b f_X(x)\, dx$$

DIM

$$\begin{cases} I_a = (-\infty, a) \\ I_b = (-\infty, b) \\ I_{ab} = [a, b] \end{cases}$$

$$\cdot\ P[I_{ab}] = P[I_b] - P[I_a]$$

$$= P[X \leq b] - P[X \leq a] = F_X(b) - F_X(a)$$

$$= \int_{-\infty}^{b} f_X(x)\, dx - \int_{-\infty}^{a} f_X(x)\, dx = \int_a^b f_X(x)\, dx$$

DEF

Sia X una variabile aleatoria continua. Se esiste una funzione f_X tale per cui $\forall a, b \in \mathbb{R}$:

$$\boxed{P[a < X \leq b] = \int_a^b f_X(x)\, dx}$$

allora f_X è detta funzione densità di probabilità

OSS.

1. $f_X(x) = 0$ se $x \notin I$, ($I \subseteq \mathbb{R}$ è il supporto di X)
2. $P[a < X \leq b] = P[a < X < b]$

PROPRIETA'

Se f_X è la funzione densità di una variabile aleatoria continua, allora:

1. $f_X(x) \geq 0 \quad \forall x \in \mathbb{R}$ e $f(x) > 0$ se $x \in I$
2. $\int_{-\infty}^{+\infty} f_X(x)\, dx = 1$

esempio

Supponiamo che sia possibile estrarre un numero reale a caso nell'intervallo $[1, 3]$ in modo che nessuna zona dell'intervallo sia privilegiata.

Sia Y la variabile aleatoria "numero estratto".

Definire la densità di probabilità:

Supporto $= I = (1, 3)$

$$f_Y(x) = \begin{cases} 0 & \text{se } x < 1 \\ c & \text{se } 1 \leq x \leq 3 \\ 0 & \text{se } x > 3 \end{cases}$$

$$\int_{-\infty}^{+\infty} f_Y(x)\, dx = \int_1^3 f_Y(x)\, dx = \int_1^3 c\, dx = c \int_1^3 dx = 2c$$

Sappiamo che $\int_{-\infty}^{+\infty} f_Y(x)\, dx = 1$

$\Rightarrow 2c = 1 \Rightarrow c = 1/2$.

Definiamo la funzione di ripartizione

$$F_X(t) = P[X \le t] = \int_{-\infty}^{t} f_X(x)\,dx$$

$$F_Y(1) = \int_{-\infty}^{1} f_Y(x)\,dx = 0 \quad \text{(per definizione)} \quad [t < 1]$$

$$F_Y(t) = \int_{-\infty}^{t} f_Y(x)\,dx = \int_{1}^{t} f_Y(x)\,dx = \frac{1}{2}\int_{1}^{t} dx = \frac{1}{2}(t-1) \quad [1 \le t \le 3]$$

$$F_Y(3) = \int_{-\infty}^{3} f_Y(x)\,dx = \int_{1}^{3} f_Y(x)\,dx = \frac{1}{2}\int_{1}^{3} dx = 1 \quad [t > 3]$$

$$\Rightarrow F_Y(t) = \begin{cases} 0 & \text{se } t < 1 \\ 1/2\,(t-1) & \text{se } 1 \le t \le 3 \\ 1 & \text{se } t > 3 \end{cases}$$

OSSERVAZIONE

- Se X è discreta:

$$P[a < X \le b] = F_X(b) - F_X(a)$$

$$\Rightarrow P[X > a] = (1 - P[X \le a]) = 1 - F_X(a)$$

- se X è continua

VARIABILE ALEATORIA FUNZIONE DI UN'ALTRA VARIABILE ALEATORIA

- <u>Caso discreto</u>

Supponiamo di avere

- $P[X=-1]=0.2$; $P[X=0]=0.3$; $P[X=1]=0.4$; $P[X=2]=0.1$

Determinare la densità della variabile aleatoria $Y=X^2$

$\text{supp}X=\{-1,0,1,2\} \longrightarrow \text{supp}Y=\{0,1,4\}$

- $P[Y=0]=P[X=0]=0.3$
- $P[Y=1]=P[X=1]+P[X=-1]=0.2+0.4=0.6$
- $P[Y=4]=P[X=2]=0.1$

- <u>Caso continuo</u>

TEOREMA

Sia X una variabile aleatoria continua e sia $Y=g(X)$.
Se g è una funzione invertibile nel supporto I di X e h è la sua inversa, allora:

$$\boxed{f_Y(y)=f_X(h(y))\cdot|h'(y)|}$$

DIM

$y=g(x) \Rightarrow x=g^{-1}(y) \quad h(y)$

$$F_Y(t)=P[Y\le t] \overset{x.hyp}{=} P[g(x)\le t]=\int_{g(x)\le t} f_X(x)\,dx$$

$$F_Y(t)=\int_{y\le t} f_X(h(y))\left|\frac{dx}{dy}\right|dy=\int_{y\le t} f_X(h(y))\,|h'(y)|\,dy$$

esempi

Sia X la variabile aleatoria avente densità

$$f_X(x)=\begin{cases}\frac{3}{4}(1-x^2) & \text{se } -1\le x\le 1\\ 0 & \text{altrove}\end{cases}$$

Determinare la densità della variabile aleatoria $Y=e^X$

$[-1,1] \xrightarrow{y=e^x} [e^{-1},e]$, $x=\ln(y)$

$$f_Y(y)=f_X(\ln y)\left|\frac{dx}{dy}\right| \Rightarrow f_Y(y)=\frac{3}{4}\left(1-(\ln(y))^2\right)\left|\frac{1}{y}\right|$$

DISUGUAGLIANZA DI CEBICEV

$$\boxed{P[\,|X-E[X]|\geq \varepsilon\,] \leq \frac{Var[X]}{\varepsilon}\,,\ \forall \varepsilon>0}$$

DEF (momenti di una variabile aleatoria)

Si dice momento K-esimo di una variabile aleatoria X e si indica con μ_K la seguente quantità:

$$\boxed{\mu_K = E[X^K]}$$

avremo che:

- $\mu_K = \sum_i x_i^K \, p_X(x_i)$, nel caso discreto
- $\mu_K = \int_{-\infty}^{+\infty} x^K f_X(x)\,dx$, nel caso continuo

OSS.

- Per K = 1 avremo la media
- $Var[X] = \mu_2 - (\mu_1)^2$

DEF (Funzione generatrice dei momenti)

$$\boxed{m_X(t) = E[e^{tX}]}$$

PROPRIETÀ

$$\mu_1 = \left[\frac{d}{dt} m_X(t)\right]_{t=0}$$

$$\mu_K = \left[\frac{d^K}{dt^K} m_X(t)\right]_{t=0}$$

PROPRIETÀ

Siano X e Y due variabili aleatorie, se

$$\boxed{m_X(t) = m_Y(t)\ ,\ \forall t \in [-t_0, +t_0]}$$

allora X e Y hanno la medesima densità.

PROPRIETÀ

Se X possiede la funzione generatrice dei momenti e $Y = aX + b$, allora, esiste anche $m_Y(t)$ e

$$\boxed{m_Y(t) = e^{tb}\, m_X(at)}$$

DEF

Data una variabile aleatoria X, si dice funzione caratteristica di X la funzione:

$$\boxed{H_X(t) = E[e^{itx}] \quad \text{, con } i \text{ unità imm.}}$$

esercizi

52 carte di un mazzo vengono distribuite a caso tra 4 giocatori.

a. Qual è la probabilità che in una distribuzione ogni giocatore abbia le stesse carte che aveva nella distribuzione precedente?

b. Qual è la probabilità che uno dei giocatori abbia 7 carte di picche?

SVOLGIMENTO A

METODO 1.

- $52!$, sono tutti i possibili casi
- $13!$, modi diversi di prendere le carte per ogni giocatore
- $(13!)^4$, perchè i giocatori sono 4.

$$P = \frac{(13!)^4}{52!} = 1.8 \cdot 10^{-29}$$

METODO 2

- $\binom{52}{13}$ carte I° giocatore
- $\binom{39=52-13}{13}$ carte II° giocatore
- $\binom{26}{13}$ carte III° giocatore
- $\binom{13}{13} = 1$ carte IV° giocatore

I modi possibili sono:

$$\binom{52}{13}\binom{39}{13}\binom{26}{13} \cdot 1 = \frac{52!}{13! \cdot 13! \cdot 13! \cdot 13!}$$

$$\Rightarrow P = \frac{(13!)^4}{52!} \approx 1.8 \cdot 10^{-29}$$

SVOLGIMENTO B

Abbiamo 52! modi di distribuire le carte. Ogni giocatore ha 13 carte

7 devono essere di picche e 6 di qualsiasi altro seme

- Ogni giocatore può avere $\binom{13}{7}$ carte di picche.
- Le altre saranno $\binom{52-13}{6}$
- 13! sono i modi in cui queste carte possono essere pescate

39! dagli altri giocatori

$$\Rightarrow \frac{\binom{13}{7}\binom{39}{6}\,13!\,39!}{52!} = 0.0088$$

METODO 2

- $\binom{52}{13}$ modi di prendere le carte
- $\binom{13}{7}$ modi di prendere le picche
- $\binom{39}{6}$ modi di prendere le non picche

$$\Rightarrow p = \frac{\binom{13}{7}\binom{39}{6}}{\binom{52}{13}} = 0.0088$$

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

ESERCITAZIONE

- Un dado viene truccato in modo tale che il verificarsi di un numero pari è due volte più probabile di un numero dispari. Se E è l'evento che un numero inferiore a 4 si verifichi a un singolo lancio del dado. Si calcoli P[E].

  - $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - $P_{dispari} = w$
  - $P_{pari} = 2w$
  - $\sum_i p_i = 1 \Rightarrow 3P_{pari} + 3P_{dispari} = 1 \Rightarrow 3w + 6w = 1 \Rightarrow w = \frac{1}{9}$
  - La probabilità che esca dispari è $1/9$ e la probabilità che esca pari è $2/9$.
  - $E\{1, 2, 3\} \Rightarrow P[E] = 1/9 + 2/9 + 1/9 = 4/9$

- Considerando l'esercizio precedente, sia A l'evento: "l'esito del lancio è un numero pari" e sia B l'evento: "l'esito del lancio è un numero divisibile per 3". Si determinino $P[A \cup B]$ e $P[A \cap B]$

  - $A = \{2, 4, 6\}$
  - $B = \{3, 6\}$
  - $P[A \cup B] = P[A] + P[B] - P[A \cap B]$
  - $A \cap B = \{6\}$
  - $P[A \cap B] = 2/9$
  - $P[A] = 2/9 + 2/9 + 2/9 = 6/9 = \frac{2}{3}$
  - $P[B] = 1/9 + 2/9 = 3/9 = 1/3$
  - $P[A \cup B]$

- Durante una lezione per ingegneri sono presenti 25 studenti di ingegneria industriale, 10 di ingegneria meccanica, 10 di ingegneria elettrica e 8 di ingegneria civile. Scegliendo uno studente a caso si vuole conoscere la probabilità che lo studente scelto sia:
(a) uno studente di ingegneria industriale
(b) uno studente di ingegneria civile o elettrica

- I, M, E, C

- Tot. studenti $= 25 + 10 + 10 + 8 = 53$

- (a) $P[I] = 25/53$

- (b) $= P[C \cup E] = P[C] + P[E] - P[C \cap E]$

- $C \cap E = \emptyset \Rightarrow P[C \cap E] = 0$

- $P[C] = 8/53$

- $P[E] = 10/53$

- $P[C \cup E] = 8/53 + 10/53 + 0 = 18/53$

- Un giocatore di poker ha in mano 5 carte, si determini la probabilità che abbia 2 assi e 3 Jack.

- Numero di modi in cui si possono distribuire 2 assi su 4 carte.

$$\binom{4}{2} = \frac{4!}{2!(4-2)!} = \frac{1 \cdot 2 \cdot 3 \cdot \cancel{4}}{\cancel{1 \cdot 2} \cdot \cancel{1 \cdot 2}} = 6$$

- Numero di modi in cui si possono distribuire 3 Jack su 4 carte

$$\binom{4}{3} = \frac{4!}{3!(4-3)!} = \frac{\cancel{1 \cdot 2 \cdot 3} \cdot 4}{\cancel{1 \cdot 2 \cdot 3} \cdot 1} = 4$$

- Numero di casi favorevoli $= 6 \cdot 4 = 24$

- Casi possibili $= \binom{52}{5} = \frac{52!}{5!(47)!}$

- $P = \frac{24}{\binom{52}{5}} = 0.9 \cdot 10^{-5}$

- Qual è la probabilità di ottenere 7 o 11 dal lancio di una coppia di dadi equilibrati?

- $A^{7} = \{(1,6), (6,1), (3,4), (4,3), (5,2), (2,5)\}$
- $B^{11} = \{(5,6), (6,5)\}$
- $P[A \cup B] = P[A] + P[B] - P[A \cap B]$
- $A \cap B = \emptyset \Rightarrow P[A \cap B] = 0$
- Casi possibili $= 6 \cdot 6 = 36$
- $P[A] = 6/36 = 1/6$
- $P[B] = 2/36 = 1/18$
- $P[A \cup B] = 1/6 + 1/18$

- Se la probabilità che un meccanico ripari 3, 4, 5, 6, 7, 8 o più automobili in un giorno lavorativo sono rispettivamente: 0.1, 0.12, 0.19, 0.28, 0.24, 0.07. Qual è la probabilità di riparare almeno 5 automobili in un giorno lavorativo?

- E = riparare almeno 5 automobili
- $P[E] = 1 - P[\bar{E}] = 1 - 0.1 + 0.12$

- Si supponga che le specifiche di fabbricazione per la lunghezza di un determinato cavo per computer siano di 2000 ± 10 mm. In questo settore è noto che la probabilità di essere difettoso per un cavo corto è pari a quella per un cavo di dimensioni maggiori.
In altre parole, la probabilità di produrre un cavo di lunghezza inferiore a 1990 mm è uguale alla probabilità di produrre un cavo di lunghezza superiore a 2010 mm. La probabilità che il processo di produzione sia conforme alle specifiche è pari a 0.99

a) Qual è la probabilità che un cavo selezionato a caso ecceda la lunghezza indicata nelle specifiche?

b) Qual è la probabilità che un cavo selezionato a caso sia più lungo di 1990 mm?

- $P[1990 \leq X \leq 2010] = 0.99$

a) $P[X > 2010] = \frac{1}{2}(1 - 0.99)$

b) $P[X > 1990] = P[1990 \leq X \leq 2010] + P[X > 2010] = 0.99 + \frac{1}{2}(1 - 0.99)$

- La probabilità che un volo parta in orario $P[D] = 0.83$ e la probabilità che arrivi in orario è $P[A] = 0.82$
La probabilità che parta e arrivi in orario è $P[D \cap A] = 0.78$
Si determini la probabilità che un aereo:
(a): arrivi in orario dato che è partito in orario
(b): che parta in orario dato che è arrivato in orario

- (a) $P[A|D] = P[D \cap A] / P[D] = 0.78/0.83$

- (b) $P[D|A] = P[D \cap A] / P[A] = 0.78/0.82$

- Si supponga di avere una scatola contenente 20 fusibili di cui 5 difettosi. Se 2 fusibili fossero selezionati a caso e poi rimossi dalla scatola, uno dopo l'altro, senza reinserimento. Quale sarebbe la probabilità che entrambi siano difettosi?

- A = "primo fusibile difettoso"

- B = "secondo fusibile difettoso"

- $P[A] = 5/20 = 1/4$

- $P[B] = 4/19$

- $P[A] \cdot P[B]$

ESERCITAZIONE

- In un impianto di assemblaggio, 3 macchine B_1, B_2 e B_3 assemblano una quantità di prodotti pari rispettivamente a: 30%, 45%, 25%. Dai dati storici è noto che rispettivamente il 2%, il 3% e il 2% dei prodotti finali di ciascuna macchina sono difettosi. Si supponga ora, di scegliere casualmente un prodotto finito. Qual è la probabilità che esso sia difettoso?

- A = "prodotto difettoso"

$$P[A] = P[A|B_1] \cdot P[B_1] + P[A|B_2] \cdot P[B_2] + P[A|B_3] \cdot P[B_3] =$$

$$= 0.02 \cdot 0.3 + 0.03 \cdot 0.45 + 0.02 \cdot 0.25$$

- Un'azienda manufatturiera sviluppa 3 progetti per la realizzazione e lo sviluppo di un particolare prodotto. Per ragioni di costo, tutti e tre i progetti sono implementati in tempi diversi. I progetti 1, 2, 3 sono usati rispettivamente per il 30%, il 20% e il 50% dei prodotti. Il tasso di malfunzionamento è diverso per le tre procedure ed è calcolato come segue:

- $P[D|P_1] = 0.01$; $P[D|P_2] = 0.03$; $P[D|P_3] = 0.02$

Dove $P[D|P_S]$ è la probabilità di un prodotto difettoso dato il piano S.
Se un prodotto a caso venisse osservato difettoso, quale progetto potrebbe essere stato impiegato e che quindi sarebbe responsabile del malfunzionamento

- $$P[P_1|D] = \frac{P[P_1] \cdot P[D|P_1]}{P[P_1] \cdot P[D|P_1] + P[P_2] \cdot P[D|P_2] + P[P_3] \cdot P[D|P_3]} = 0.58$$

- $P[P_2|D] = \ldots = 0.316$

- $P[P_3|D] = \ldots = 0.526$

P_3 viene rigettato.

- Andrea e Bruno giocano d'azzardo con la seguente regola: lanciano un dado, se esce 1 o 2, Andrea da a Bruno 2€, se esce 3 o 4, andrea da a Bruno 4€, se esce 5 Andrea da a Bruno 8€. Se esce 6, Bruno da ad Andrea una somma di α €. Quanto deve valere α affinché il gioco sia equo?

- X = "guadagno di Bruno"

$P[X=2] = 2/6$

$P[X=4] = 2/6$

$P[X=8] = 1/6$

$P[X=-\alpha] = 1/6$

$E[X] = 2 \cdot P[X=2] + 4 \cdot P[X=4] + 8 \cdot P[X=8] - \alpha \cdot P[X=-\alpha] =$

Expectation value

$= 2 \cdot 1/3 + 4 \cdot 1/3 + 8 \cdot 1/6 - \alpha \cdot 1/6 = \frac{20-\alpha}{6}$

Per essere equo $\Rightarrow \alpha = 20$

SCHEMA DI BERNOULLI

Consideriamo un esperimento che si può concludere solo in due modi chiamati convenzionalmente "successo" e "insuccesso".
Un esperimento di questo tipo è detto di Bernoulli.
Sia p la probabilità che si verifichi il successo e q quella dell'insuccesso.

$$\boxed{q = 1 - p}$$

DEF (Distribuzione di Bernoulli)

Una variabile aleatoria X che vale 1 se la prova di Bernoulli si è conclusa con il successo e 0 altrimenti, è detta variabile aleatoria di Bernoulli di parametro p, in simboli:

$$\boxed{X \sim Ber(p)}$$

DENSITÀ DELLA DISTRIBUZIONE DI BERNOULLI

- $P[X=1] = p$
- $P[X=0] = q$
- $P[X=k] = p^k \cdot q^{1-k}$

MOMENTI DELLA DISTRIBUZIONE DI BERNOULLI

- $m_x(t) = E[e^{xt}] \Rightarrow m_x(t) = pe^t + q$
- $E[X] = \frac{d}{dt} m_x(t) \Big|_{t=0} \Rightarrow E[X] = pe^t \Big|_{t=0} = p$
- $Var[X] = E[(X - E[X])^2] = E[X^2] - E^2[X]$
  - $E[X^2] = \frac{d^2}{dt^2} m_x(t) \Big|_{t=0} = pe^t \Big|_{t=0} = p$

$\Rightarrow Var[X] = p - p^2 = p(1-p) = p \cdot q$

DEF (schema di Bernoulli)

Una sequenza di prove di Bernoulli, tra di loro indipendenti per le quali la probabilità del successo si mantiene invariata è detto schema di Bernoulli.

DISTRIBUZIONE BINOMIALE

DEF

La variabile aleatoria X che conta il numero di successi in n prove di uno schema di Bernoulli, dove p è la probabilità del singolo successo, è detta binomiale, in simboli:

$$\boxed{X \sim Bin(n,p)}$$

PROPRIETA' (Densità della distribuzione binomiale)

- $P[X=k] = \binom{n}{k} p^k \cdot q^{n-k}$

DIM.

- n: tentativi
- k: successi
- n-k: insuccessi

- probabilità di k successi: p^k
- probabilità di n-k insuccessi: q^{n-k}

$\Rightarrow P[X=k] = \binom{n}{k} p^k q^{n-k}$

MOMENTI DELLA DISTRIBUZIONE BINOMIALE

- $m_x(t) = (q + pe^t)^n$
- $E[X] = np$
- $Var[X] = n \cdot p \cdot q$

DIM

- $m_x(t) = \sum_{k=0}^{n} \binom{n}{k} p^k \cdot q^{n-k} \cdot e^{tk} =$

$$= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k q^{n-k} =$$

$$= (pe^t + q)^n$$

- $E[X] = \frac{d}{dt} m_x(t) \Big|_{t=0} = n(pe^t + q)^{n-1} \cdot pe^t \Big|_{t=0} =$

$$= np \overbrace{(p+q)}^{1} = np$$

- $\frac{d^2}{dt^2} m_x(t) \Big|_{t=0} = np\left[(n-1)(pe^t+q)^{n-2} pe^{2t} + (pe^t+q)^{n-1} e^t\right]_{t=0} =$

$$= np\left[(n-1)p + 1\right] = np\left[np - p + 1\right].$$

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

$$= n^2p^2 - np^2 + np - n^2p^2 = np(1-p) = npq$$

DISTRIBUZIONE GEOMETRICA

DEF

La variabile aleatoria che conta il numero di prove di uno schema di Bernoulli fino alla comparsa del primo successo è detta geometrica di parametro p, dove p rappresenta la probabilità del singolo successo, in simboli:

$$\boxed{X \sim Geom(p)}$$

PROPRIETÀ (Densità della distribuzione geometrica)

- $P[X=k] = pq^{k-1}$ $\qquad k = 1, 2, 3 \dots$

DIM

$\{X=k\}$: la situazione in cui si sono verificati $k-1$ insuccessi e 1 successo.

- $P[X=k] = \underbrace{P[ins] \cdot P[ins] \cdot \dots \cdot P[ins]}_{k-1 \text{ volte}} \cdot P[successo] = q^{k-1} p$

MOMENTI DELLA DISTRIBUZIONE GEOMETRICA

- $m_x(t) = \dfrac{pe^t}{1-qe^t}$, per $t < \log 1/q$

- $E[X] = 1/p$

- $Var[X] = q/p^2$

DIM

- $m_x(t) = E[e^{xt}] = \sum_{k=1}^{\infty} e^{tk} \cdot q^{k-1} p =$

$$= p \sum_{k=1}^{\infty} e^{tk} q^{k-1} =$$

$$= p \sum_{k=1}^{\infty} e^{tk} \cdot e^{-t} \cdot e^{t} \cdot q^{k-1} =$$

$$= pe^t \sum_{k=1}^{\infty} e^{tk-1} \cdot q^{k-1} =$$

$$= pe^t \sum_{k=1}^{\infty} (qe^t)^{k-1} =$$

$$= pe^t \sum_{k=0}^{\infty} (qe^t)^k = pe^t \cdot \frac{1}{1-qe^t}$$

OSS

$\sum_{n=0}^{\infty} x^n = 1/(1-x)$, $|x|<1 \Rightarrow qe^t < 1 \Rightarrow \dfrac{pe^t}{1-qe^t}$, $t < \log 1/q$

$$E[X] = \frac{d}{dt} m_x(t) \Big|_{t=0} =$$

$$= \frac{pe^t \cdot (1-qe^t) - pe^t(-qe^t)}{(1-qe^t)^2} \Bigg|_{t=0} =$$

$$= \frac{pe^t(1-qe^t) + pq\,e^{2t}}{(1-qe^t)^2} \Bigg|_{t=0} =$$

$$= \frac{p(1-q) + pq}{(1-q^2)} =$$

$$= \frac{p - pq + pq}{p^2} = \frac{1}{p}$$

PROPRIETÀ (Assenza di memoria)

Sia $X \sim Geom(p)$, allora:

$$\boxed{P[X = i+j \mid X > i] = P[X = j]}$$

DIM

$X = i + j \Rightarrow X \geq i$

$\Rightarrow (X = i+j) \cap (X > i) = (X = i+j)$

$$\Rightarrow P[X = i+j \mid X > i] = \frac{P[(X = i+j) \cap (X > j)]}{P[X > i]} =$$

$$= \frac{pq^{i+j-1}}{q^j} = pq^{i-1} = P[X = i]$$

DEF (Distribuzione binomiale negativa)

La variabile aleatoria che conta il numero di tentativi necessari al verificarsi dell' n-esimo successo di uno schema di Bernoulli, è detta Binomiale negativa di parametri n e p, in simboli:

$$\boxed{X \sim \overline{Bin}(n, p)}$$

PROPRIETA' (Densità della distribuzione binomiale negativa)

- $P[X=k] = \binom{k-1}{m-1} p^m q^{k-m}$

MOMENTI DELLA DISTRIBUZIONE BINOMIALE NEGATIVA

- $m_x(t) = \left(\frac{pe^t}{1-qe^t}\right)^m \quad , t < \ln \frac{1}{q}$

- $E[X] = m/p$

- $Var[X] = m \frac{q}{p^2}$

DISTRIBUZIONE DI POISSON

Data la distribuzione Binomiale

$$Bin = P[X=k] = \binom{m}{k} p^k q^{m-k} \quad , E[X] = mp$$

La distribuzione di Poisson è un caso limite in cui:

- $m \to \infty \; ; \; p \to 0 \; ; \; mp \to \mu \neq 0$

$$\Rightarrow P[X=k] = \frac{m!}{k!(m-k)!} p^k \frac{q^m}{q^k} =$$

$$= \frac{1}{k!} \cdot \frac{m!}{(m-k)!} \cdot p^k \cdot \frac{(1-p)^m}{(1-p)^k} =$$

$$= \frac{1}{k!} \cdot \frac{m!}{(m-k)!} \cdot \frac{1}{m^k} (mp)^k (1-p)^m \cdot \frac{1}{(1-p)^k} =$$

Per $m \to \infty \; ; \; p \to 0 \; ; \; mp \to \mu \neq 0$

- $\frac{m!}{(m-k)!} \cdot \frac{1}{m^k} \sim 1 \Rightarrow \frac{m!}{(m-k)!} \cdot \frac{1}{m^k} \cdot (mp)^k \longrightarrow \mu^k$

- $(1-p)^m \longrightarrow \left(1 - \frac{mp}{m}\right)^m \longrightarrow \left(1 - \frac{\mu}{m}\right)^m \longrightarrow e^{-\mu}$

- $\frac{1}{(1-p)^k} \longrightarrow 1$

$$\Rightarrow P[X=k] = \frac{\mu^k e^{-\mu}}{k!}$$

DEF

La variabile aleatoria avente densità:

$$\boxed{P[X=k] = \frac{\mu^k e^{-\mu}}{k!}}$$

È detta variabile di Poisson di parametro μ, in simboli:

$$\boxed{X \sim Poi(\mu)}$$

MOMENTI DELLA DISTRIBUZIONE DI POISSON

- $m_x(t) = e^{\mu(e^t-1)}$
- $E[X] = \mu$
- $Var[X] = \mu$

DIM

- $m_x(t) = \sum_{k=0}^{\infty} e^{tk} \frac{e^{-\mu}\mu^k}{k!} = e^{-\mu} \sum_{k=0}^{\infty} \frac{(e^t \cdot \mu)^k}{k!} =$

$$= e^{-\mu} \cdot e^{\mu e^t} = e^{\mu(e^t-1)}$$

- $E[X] = \frac{d}{dt} m_x(t) \Big|_{t=0} = e^{\mu(e^t-1)} \cdot \mu e^t \Big|_{t=0} = \mu$

- $E[X^2] = \frac{d^2}{d^2t} m_x(t) \Big|_{t=0} = \mu \left[e^{\mu(e^t-1)} \mu e^t e^t + e^{\mu(e^t-1)} e^t \right]_{t=0} =$

$$= \mu \left[e^{\mu(e^t-1)} \mu e^{2t} + e^t \cdot e^{\mu(e^t-1)} \right]_{t=0} =$$

$$= \mu \left[(1 \cdot \mu) + 1 \right] = \mu^2 + \mu$$

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

ESEMPIO - DISTRIBUZIONE DI POISSON

Il numero di automobili che transitano per un determinato passaggio a livello è distribuito secondo una variabile di Poisson. In media, transitano 36 automobili in un'ora.

Calcolare la probabilità che nell'arco di 5 minuti transitino 2 automobili, al massimo 2 e almeno 2.

- X = "numero di automobili che transitano nell'arco di 5 minuti".
- $P[X=k] = \frac{e^{-\mu}\mu^k}{k!}$, $\mu = \frac{36}{60} \cdot 5 = 3$
- $P[X=2] = \frac{e^{-3} \cdot 3^2}{2!}$
- $P[X \leq 2] = P[X=0] + P[X=1] + P[X=2]$
- $P[X \geq 2] = 1 - P[X=0] + P[X=1]$

SCHEMA DI POISSON

In un certo intervallo di tempo (o di spazio) si manifestano degli eventi chiamati "arrivi", che soddisfano le seguenti condizioni:

1. Non vi sono più arrivi simultanei (non ci sono sovrapposizioni);
2. Il numero medio di arrivi nell'unità di tempo è costante (intensità costante);
3. Ciascun arrivo avviene in modo indipendente dagli altri.

DEF (Processo di Poisson)

La variabile aleatoria N_T che conta, in un certo intervallo di tempo t, il numero di arrivi in uno schema di Poisson di intensità λ è detta Processo di Poisson.

CARATTERISTICHE

1. $N_0 = 0$
2. $N_{t+s} = N_t + N_s$
3. se $\Delta t \to 0$:
   - $P[N_{t+\Delta t} - N_t = 1] = \lambda \Delta t + o(\Delta t)$
   - $P[N_{t+\Delta t} - N_t \geq 1] = o(\Delta t)$
   - $P[N_{t+\Delta t} - N_t = 0] \approx 1 - \lambda \Delta t + o(\Delta t)$

PROPRIETA'

Sia N_T un Processo di poisson di intensità λ, allora $\Rightarrow N_T \sim Poi(\mu = \lambda t)$

DEF (Distribuzione esponenziale)
La variabile aleatoria "Tempo di attesa del primo arrivo in un processo di Poisson di intensità λ" è detta esponenziale di variabile λ, in simboli:

$$\boxed{X \sim Esp(\lambda)}$$

PROPRIETÀ (Densità della distribuzione esponenziale)

$$F_X(x) = \begin{cases} 0 & , x<0 \\ \lambda e^{-\lambda x} & , x>0 \end{cases}$$

DIM

$F_X(t) = P[X \leq t]$ "probabilità del primo arrivo entro t"

$P[X > t] = 1 - P[X \leq t] = 1 - F_X(t)$

$\Rightarrow F_X(t) = 1 - P[X > t] = 1 - P[N_T = 0] =$

$= 1 - e^{-\lambda t}$

$\Rightarrow f_X(t) = \frac{d}{dt} F_X(t) = \lambda e^{-\lambda t}$

MOMENTI DELLA FUNZIONE ESPONENZIALE

- $m_X(t) = \frac{\lambda}{\lambda - t}$, $t < \lambda$

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

- $Var[X] = \frac{1}{\lambda^2}$

DIM

$$m_X(t) = \int_{-\infty}^{+\infty} e^{tx} f(x)\, dx = \int_0^{+\infty} e^{tx} \lambda e^{-\lambda x}\, dx =$$

$$= \lambda \int_0^{+\infty} e^{(t-\lambda)x}\, dx = \lambda \left[\frac{e^{(t-\lambda)x}}{t-\lambda} \right]_0^{\infty} =$$

$$= \lambda \left[\frac{e^{(t-\lambda)\infty}}{t-\lambda} - \frac{1}{t-\lambda} \right] = \frac{\lambda}{\lambda - t}$$

$0 < t < \lambda$

PROPRIETA' (ASSENZA DI MEMORIA DELLA LEGGE ESPONENZIALE)

$$P[X > t + s \mid X > t] = P[X > t + s] \quad \forall t, s \in \mathbb{R}^+$$

DEF (Distribuzione gamma)
La variabile aleatoria "tempo di attesa dell'ennesimo arrivo di un processo di Poisson" è detta gamma di parametri n e λ, in simboli:

$$\boxed{X \sim \Gamma(n, \lambda)}$$

PROPRIETA' (Densità della distribuzione gamma)

$$f_X(x) = \begin{cases} 0 & , x < 0 \\ \frac{1}{(n-1)} \cdot \lambda^n x^{n-1} e^{-\lambda x} & , x > 0 \end{cases}$$

DENSITA' DI WEIBUL (FUNZIONE RISCHIO)
DEF (Distribuzione di Weibul)
Sia $Y \sim Esp(\lambda)$ e sia $\lambda > 0$. È detta variabile di Webul la variabile aleatoria:

$$\boxed{X = Y^{\frac{1}{\alpha}}}$$

in simboli

$$\boxed{X \sim Wei(\lambda, \alpha)}$$

DENSITA' DELLA DISTRIBUZIONE DI WEIBUL

$$\cdot f_X(x) = \begin{cases} 0 & , x \leq 0 \\ \lambda \alpha x^{\alpha - 1} e^{-\lambda x^\alpha} & , x > 0 \end{cases}$$

$$\cdot F_X(x) = \begin{cases} 0 & , x < 0 \\ 1 - e^{-\lambda x^\alpha} & , x \geq 0 \end{cases}$$

DIM

$Y \geq 0 \; ; \; X = Y^{1/\alpha}$

$F_X(x) = P[X \leq x] = P[Y^{\frac{1}{\alpha}} \leq X] =$

$= P[Y \leq x^{\alpha}] = F_Y(x^{\alpha}) = 1 - e^{-\lambda x^{\alpha}}$

$f_X(x) = F_X'(x)$

OSS.

- $E[X] = \alpha^{-\frac{1}{\lambda}} \Gamma(1 + \frac{1}{\lambda})$

- $Var[X] = \alpha^{-\frac{1}{\lambda}} [\Gamma(1 + \frac{2}{\lambda}) - \Gamma(1 + \frac{1}{\lambda})]$

FUNZIONE DI RISCHIO

Variabile aleatoria T = "tempo di funzionamento".

La probabilità che, condizionatamente al fatto che fino all'istante $\bar{t}$ l'apparecchio fosse in funzione, esso si guasti immediatamente dopo, ossia, all'istante $\bar{t} + \Delta t$ è:

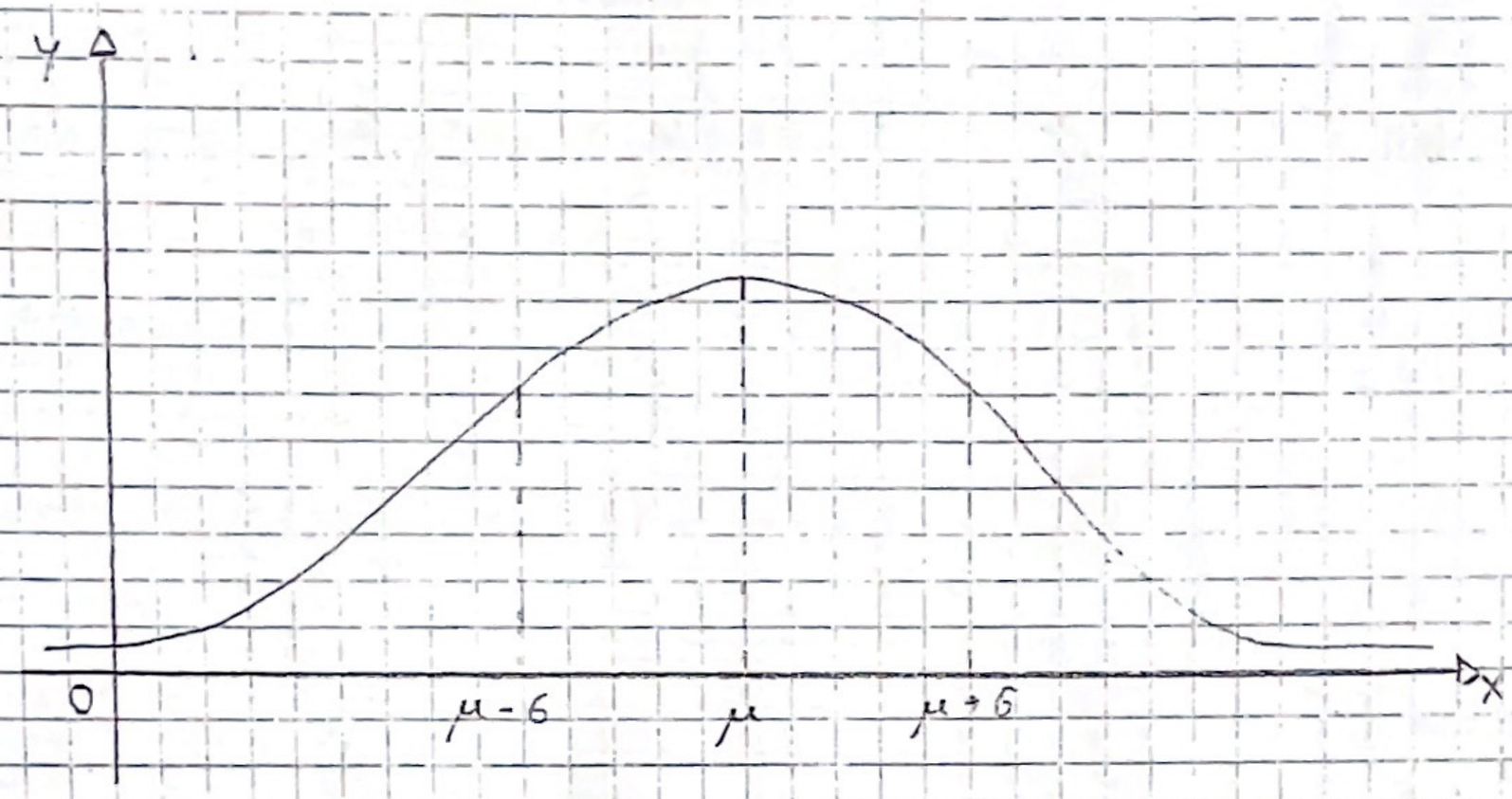
$$\boxed{P[T \leq \bar{t} + \Delta t \mid T > \bar{t}]}$$

DISTRIBUZIONE GAUSSIANA (O NORMALE)

DEF.

Una variabile aleatoria X è detta distribuita secondo una normale (o gaussiana) di parametri $\mu \in \mathbb{R}$ e $\sigma \in \mathbb{R}^+$ ($X \sim N(\mu, \sigma)$) se ha densità di probabilità:

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



PROPRIETÀ

1. $f_X(x)$ è simmetrica rispetto alla retta $x = \mu$
2. ha un massimo assoluto in $\left(\mu, \frac{1}{\sigma\sqrt{2\pi}}\right)$
3. ha asintoto orizzontale $y = 0$
4. ha due flessi in $\left(\mu - \sigma, \frac{1}{\sigma\sqrt{2\pi e}}\right)$, $\left(\mu + \sigma, \frac{1}{\sigma\sqrt{2\pi e}}\right)$
5. $\int_{-\infty}^{+\infty} f_X(x)\, dx = 1$
6. all'aumentare di σ, il massimo diminuisce
7. se $\mu = 0$, è simmetrica rispetto all'asse y
8. $E[X] = \mu$

   $Var[X] = \sigma^2$

La funzione di ripartizione associata è:

$$F_X(t) = P[X \leq t] = \int_{-\infty}^{t} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot \frac{(x-\mu)^2}{\sigma^2}}\, dx$$

DEF (Distribuzione normale standardizzata)

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

DIM

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

cambio le variabili: $z = \frac{x-\mu}{\sigma} \Rightarrow x = \sigma z + \mu \Rightarrow dx = \sigma dz$

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}z^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}z^2} dx$$

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

VETTORI DI VARIABILI ALEATORIE

DEF.

È detto vettore aleatorio, o variabile aleatoria multidimensionale, un vettore le cui componenti sono variabili aleatorie:

$$X = [X_1 \cap X_2 \cap \ldots \cap X_m]$$

Se abbiamo le variabili aleatorie X e Y.

- $\sum_i \sum_j P[X=i, Y=j] = 1$; $\sum_i P[X=i] = 1$; $\sum_j P[Y=j] = 1$
- I valori $P[X=i, Y=j]$ formano la densità discreta congiunta del vettore (X, Y).
- La densità $P[X=i]$, $P[Y=j]$ sono dette densità marginali.

OSS.

È sempre possibile ricavare le probabilità marginali dalle probabilità congiunte ma non è vero il viceversa.

DEF (funzione di ripartizione congiunta)
Dato un vettore aleatorio (X, Y), è detta funzione di ripartizione congiunta la funzione:

$$\boxed{F_{XY}(x, y) = P[X \leq x, Y \leq y]}$$

$$(P[(X \leq x) \cap (Y \leq y)])$$

DEF (Funzione di ripartizione originale)
Data una funzione di ripartizione congiunta F_{XY}, è definita la funzione di ripartizione marginale come:

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

PROPRIETÀ

- $\forall y \in \mathbb{R}$, $\lim_{x \to -\infty} F_{XY}(x, y) = 0$; $\lim_{x \to +\infty} F_{XY}(x, y) = F_Y(y)$

- $\forall x \in \mathbb{R}$, $\lim_{y \to -\infty} F_{XY}(x, y) = 0$; $\lim_{y \to +\infty} F_{XY}(x, y) = F_X(x)$

DENSITÀ CONGIUNTE (CASO DISCRETO)
Si dice densità congiunta del vettore aleatorio discreto (X, Y) e si indica con P_{XY}, l'insieme dei valori:

$$\boxed{P_{XY}(x_i, y_j) = P[X = x_i, Y = y_j]}$$

DEF (Densità marginali)

- $P_X[x_i] = \sum_j P_{XY}(x_i, y_j)$

- $P_Y[x_j] = \sum_i P_{XY}(x_i, y_j)$

DENSITA' CONGIUNTE (CASO CONTINUO)

Dato un vettore aleatorio continuo (X, Y), è detta funzione densità di probabilità la funzione $f_{XY}(x, y)$ per la quale, per ogni scelta di $a, b, c, d \in \mathbb{R}$:

$$P[a < X < b,\ c < Y < d] = \int_a^b \left(\int_c^d f_{XY}(x,y)\, dy \right) dx = \int_c^d \left(\int_a^b f_{XY}(x,y)\, dx \right) dy$$

$$= \int_a^b dx \int_c^d dy\, f_{XY}(x,y)$$

Si dimostra che:

- $F_{XY}(s,t) = P[X < s,\ Y < t] = \int_{-\infty}^{s} dx \int_{-\infty}^{t} dy\, f_{XY}(x,y)$
- $f_X(x) = \dfrac{dF_X(x)}{dx}$
- $f_{XY}(x,y) = \dfrac{\delta^2}{\delta x\, \delta y} \cdot F_{XY}(x,y)$

DEF (Funzioni densità marginali)

Sono dette funzioni densità marginali

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)\, dy \quad ; \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)\, dx$$

DEF

Date n variabili aleatorie X_i, con $i = 1, \dots, n$, esse si dicono indipendenti se per ogni scelta degli intervalli A_i, si ha:

$$P[X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n] = P[X_1 \in A_1] \cdot P[X_2 \in A_2] \cdot \dots \cdot P[X_n \in A_n]$$

Si dimostra che (nel caso di variabili indipendenti)

- $P_{XY}[x_i, y_i] = P_X[x_i] P_Y[y_i]$ (caso discreto)

- $f_{XY}(x_i, y_i) = f_X(x_i) \cdot f_Y(y_i)$ (caso continuo)

TEOREMA

Dato un vettore aleatorio (X, Y) [discreto o continuo], la somma $U = X + Y$ delle sue componenti ha densità:

- $P_U(u) = \sum_k P_{XY}(u, u-k)$ (caso discreto)

- $f_U(u) = \int_{-\infty}^{+\infty} f_{XY}(t, u-t)\, dt$ (caso continuo)

DEF

Sia (X, Y) un vettore aleatorio e $g(x,y)$ una funzione. $E[g(X,Y)]$, se esiste è dato da:

- $E[g(x,y)] = \sum_i \sum_j g(x_i, y_j) P_{XY}(x_i, y_j)$ (caso discreto)

- $E[g(x,y)] = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy\, g(x,y) \cdot f_{XY}(x,y)$ (caso continuo)

PROPRIETÀ

Per ogni coppia di variabili aleatorie X, Y, il valore atteso della somma è uguale alla somma dei valori attesi.

$$\boxed{E[X+Y] = E[X] + E[Y]}$$

DEF (covarianza)

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

PROPRIETÀ

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

PROPRIETÀ

X, Y, V, W variabili aleatorie

$\alpha, \beta \in \mathbb{R}$

1. $Cov[X, \alpha] = 0$

2. $Cov[X,Y] = Cov[Y,X]$

3. $Cov[\alpha X, \beta Y] = \alpha\beta\, Cov[X,Y]$

5. $Cov[X+Y, V+W] = Cov[X,V] + Cov[X,W] + Cov[Y,V] + Cov[Y,W]$

DEF. (matrice di covarianza)

Dato $X = [X_1, X_2, \dots, X_n]$, la matrice di covarianza C_X è la matrice di elementi:

$$\boxed{(C_X)_{ij} = Cov[X_i, X_j]}$$

DEF. (coefficiente di correlazione)

Date due variabili aleatorie X, Y, è detto coefficiente di correlazione (ρ_{XY})

$$\boxed{\rho_{XY} = \frac{Cov[X,Y]}{\sqrt{Var[X]\cdot Var[Y]}}}$$

PROPRIETÀ

1. $-1 < \rho_{XY} < 1$

2. $|\rho_{XY}| = 1$ se unicamente tra X e Y esiste un valore deterministico lineare $Y = aX + b$, con $a, b \in \mathbb{R}$ che hanno lo stesso segno

- La statistica ci dice se esiste correlazione tra due eventi, non è suo compito trovare il nesso causale, la 2. ci dice che se esiste una forte correlazione, il coefficiente di correlazione sarà uguale a 1.

DEF.

Si dicono incorrelate due variabili aleatorie X, Y per cui $Cov[X,Y]=0$, ovvero, il coefficiente di correlazione è nullo.

PROPRIETÀ

Se X e Y sono indipendenti, allora sono incorrelate.

DIM

$$Cov[X,Y] = E[XY] - E[X] \cdot E[Y]$$

$$E[XY] = \sum_i \sum_j x_i y_j P[x_i, y_j] =$$

$$= \sum_i \sum_j x_i y_j P[x_i] \cdot P[y_j] =$$

$$= \left(\sum_i x_i P[x_i]\right)\left(\sum_j y_j P[y_j]\right) = E[X]E[Y]$$

$\Rightarrow$ non è vero il viceversa.

PROPRIETÀ

- $Var[X+Y] = Var[X] + Var[Y] + 2Cov[X,Y]$
- $Var[X-Y] = Var[X] - Var[Y] - 2Cov[X,Y]$
- Se X_i sono indipendenti $\Rightarrow Var[\sum_i x_i] = \sum_i Var[X_i]$

TRASFORMAZIONE DI VETTORI ALEATORI

$$(X,Y) \rightarrow (U,V)$$

$$\begin{cases} u = g_1(x,y) \\ v = g_2(x,y) \end{cases} \Leftrightarrow \begin{cases} x = h_1(u,v) \\ y = h_1(u,v) \end{cases}$$

TEOREMA

Sia (X,Y) un vettore aleatorio continuo e sia $(U,V) = g(X,Y)$ una trasformazione invertibile con trasformazione inversa $h(U,V)$, allora:

$$\boxed{f_{u,v}(u,v) = f_{x,y}(h_1(u,v), h_2(u,v)) \cdot |\det J|}$$

$$J = \begin{bmatrix} \frac{\delta h_1}{\delta u} & \frac{\delta h_1}{\delta v} \\ \frac{\delta h_2}{\delta u} & \frac{\delta h_2}{\delta v} \end{bmatrix} \equiv \left(\frac{\delta(h_1, h_2)}{\delta(u,v)} \right)$$

DISTRIBUZIONE UNIFORME

È la distribuzione che assume un valore costante in un intervallo $[a,b]$

DEF

Una variabile aleatoria X ha distribuzione uniforme in $[a,b] \in \mathbb{R}$ se è assolutamente continua con densità di probabilità:

$$\boxed{f_X(t) = \begin{cases} \frac{1}{b-a} & \text{se } t \in (a,b) \\ 0 & \text{altrimenti} \end{cases}}$$

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

Avremo che

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

OSSERVAZIONE

$$\boxed{Ber(n,p) \xrightarrow{n \to \infty} N(\mu, \sigma) \xleftarrow{\lambda \to \infty} Pois(\lambda)}$$

DISTRIBUZIONE χ^2 (CHI-QUADRO)

Siano X_i, $i = 1, \dots, n$, n variabili aleatorie con distribuzione normale $N(0,1)$ indipendenti tra loro.

Sia X la variabile aleatoria:

$$\boxed{X = X_1^2 + X_2^2 + \dots + X_n^2}$$

allora si prova che questa variabile aleatoria è distribuita secondo una distribuzione χ^2.

DEF

Diremo che la variabile aleatoria X è distribuita secondo una $\chi^2(m)$, con gradi di libertà m, se è assolutamente continua con densità di probabilità:

$$\boxed{f_X(t) = \begin{cases} 0 & \text{se } t<0 \\ \frac{1}{\Gamma(\frac{m}{2})} \cdot \left(\frac{1}{2}\right)^{m/2} \cdot t^{\frac{m}{2}-1} \cdot e^{-\frac{1}{2}t} & \text{se } t \geq 0 \end{cases}}$$

Con $\Gamma(\alpha) = \int_0^{+\infty} e^{-x} x^{\alpha-1} dx$ (Gamma di Eulero)

Si prova che:

- $E[X] = m$
- $Var[X] = 2m$

DISTRIBUZIONE T-STUDENT

Siano $Z \simeq N(0,1)$ e $Y \simeq \chi^2(m)$ due variabili aleatorie indipendenti.
Definisco $X = Z / \sqrt{Y/m}$ e si prova che X è una variabile aleatoria distribuita secondo una t di Student con m gradi di libertà.

DEF

Diremo che la variabile aleatoria X è distribuita secondo una T-STUDENT con m gradi di libertà se è assolutamente continua con densità di probabilità:

$$\boxed{f_X(t) = \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})} \cdot \frac{1}{\sqrt{m\pi}} \cdot \left(1 + \frac{t^2}{m}\right)^{-\frac{1}{2}(m+1)}}$$

Si prova che

- $E[X] = 0$ se $m > 1$
- $Var[X] = \frac{m}{m-1}$ se $m > 2$
- Per $m \geq 30$ la t di Student approssima $N(0,1)$

Consideriamo $X_1, X_2, \dots, X_n$ una n-upla di variabili indipendenti e identicamente distribuite (iid).

DEF (CAMPIONE CASUALE)
Data una variabile aleatoria X, chiamiamo campione casuale (o più brevemente campione) di ampiezza n, una n-upla di variabili aleatorie tra di loro indipendenti, ciascuna delle quali distribuita come la popolazione X.

OSS.
- $X_1, X_2, \dots, X_n$ è una n-upla di variabili.
- $x_1, x_2, \dots, x_n$ è una n-upla di valori (realizzazione).

DEF (MEDIA CAMPIONARIA)
La media campionaria di un campione casuale di ampiezza n $(X_1, X_2, \dots, X_n)$ è data da:

$$\boxed{\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

DEF (MOMENTO CAMPIONARIO DI ORDINE K)

$$M_n^{(k)} = \frac{X_1^k + X_2^k + \dots + X_n^k}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

DEF (VARIANZA CAMPIONARIA)

1. valore atteso noto $(E[X] = m)$:

$$S_0^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - m)^2$$

2. valore atteso incognito

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

PROPRIETA'

Se è nota la media della popolazione X, allora:

$$S_0^2 = M_n^{(2)} - 2m\bar{X}_n + m^2$$

Se la media della popolazione è incognita, allora:

$$S_n^2 = \frac{1}{n-1}\left(n M_n^{(2)} - n\bar{X}_n^2\right)$$

PROPOSIZIONE (VALORE ATTESO E VARIANZA DELLA MEDIA CAMPIONARIA)

Se la popolazione è dotata di media e di varianza ($E[X] = m$ e $Var[X] = v$), allora, dato un campione iid: $X_1, X_2, \dots, X_n$ si ha

- $E[\bar{X}_n] = m$
- $Var[\bar{X}_n] = v/n$

DIM

$$E[\bar{X}_n] = E\left[\frac{X_1 + X_2 + \dots + X_n}{n}\right] = \frac{1}{n}E[X_1 + X_2 + \dots + X_n] = \frac{1}{\cancel{n}}\cancel{n}E[X] = m$$

$$Var[\bar{X}_n] = Var\left[\frac{X_1 + X_2 + \dots + X_n}{n}\right] = \frac{1}{n^2}\left(Var[X_1] + Var[X_2] + \dots + Var[X_n]\right) =$$

$$= \frac{1}{n^{\cancel{2}}}\cancel{n}\,Var[X] = v/n$$

PROPOSIZIONE (VALORE ATTESO DELLA VARIANZA CAMPIONARIA)

Se la popolazione X è dotata di media e varianza ($E[X] = m$, $Var[X] = v$), allora, dato un campione $X_1, X_2, \dots, X_n$, si ha:

$$\boxed{E[S_0^2] = v \quad , E[S_n^2] = v}$$

Esaminiamo il caso in cui la popolazione sia distribuita secondo la legge normale:

$$\boxed{X \sim Norm(\mu = m, \sigma^2 = v)}$$

PROPRIETÀ (DISTRIBUZIONE DELLA MEDIA CAMPIONARIA)

Sia $X_1, X_2, \dots, X_n$ un campione estratto da una popolazione:

$$\boxed{X \sim Norm(\mu = m, \sigma^2 = v)}$$

allora, detta $\bar{X}_n$ la media campionaria, si ha:

$$\boxed{\bar{X}_n \sim Norm(\mu = m, \sigma^2 = v/n)}$$

PROPRIETÀ

Sia $X_1, X_2, \dots, X_n$ un campione estratto da una popolazione $X \sim Norm(\mu = m, \sigma^2 = v)$, allora, detta S_0^2 la varianza campionaria (o media nota), si ha:

$$\boxed{S_0^2 \sim \frac{v}{n} \chi^2 (v = n)}$$

Invece, con media incognita si ha:

$$\boxed{S_n^2 \sim \frac{v}{n-1} \chi^2 (v = n-1)}$$

PROPRIETÀ

Sia $X_1, X_2, \dots, X_n$ un campione estratto da una popolazione $X \sim Norm(\mu = m, \sigma^2 = v)$, allora $\bar{X}_n$ è indipendente da S_n^2.

PROPRIETÀ

Sia $X_1, X_2, \dots, X_n$ un campione estratto da una popolazione $X \sim Norm(\mu = m, \sigma^2 = v)$, allora:

$$\boxed{\frac{\bar{X}_n - m}{\sqrt{S_n^2 / n}} \sim T(v = n-1)}$$

[t-student]

DEF (Convergenza di una distribuzione)

Si dice che la successione di variabili aleatorie $\{X_n\}$ converge in distribuzione alla variabile aleatoria X, $X_n \xrightarrow{d} X$ se:

$$\boxed{\lim_{n\to\infty} F_{X_n}(x) = F_X(x)}$$

per tutti i valori x ove $F_X(x)$ è continua.

DEF (Convergenza in probabilità)

Si dice che la successione di variabili aleatorie $\{X_n\}$ converge in probabilità alla variabile aleatoria X se:

$$\boxed{\forall \varepsilon>0,\ \lim_{n\to\infty} P[|X_n - X| < \varepsilon] = 1}$$

$X_n \xrightarrow{P} X$ in simboli.

DEF (Convergenza quasi certa)

Si dice che la successione di variabili aleatorie $\{X_n\}$ converge quasi certamente alla variabile aleatoria X se:

$$\boxed{\forall \varepsilon>0,\ P[\lim_{n\to\infty} |X_n - X| < \varepsilon] = 1}$$

$X_n \xrightarrow{q.c} X$, in simboli.

DEF (Convergenza in media quadratica)

Si dice che la successione di variabili aleatorie $\{X_n\}$ converge in media quadratica alla variabile aleatoria X se:

$$\boxed{\lim_{n\to\infty} E[(X_n - X)^2] = 0}$$

$X_n \xrightarrow{m.q.} X$, in simboli.

Si dimostra che:

$$\left.\begin{array}{l} X_n \xrightarrow{q.c} X \\ X_n \xrightarrow{m.q.} X \end{array}\right\} \Rightarrow X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$$

Quindi q.c. e m.q. sono più forti perché da esse posso ricavare le altre.

TEOREMA (Legge debole dei grandi numeri)
Sia X_n una successione di variabili aleatorie i.i.d proveniente da una popolazione X dotata di media ($E[X] = m$), allora,

$$\boxed{\forall \varepsilon > 0, \quad \lim_{n \to \infty} P[|\bar{X}_n - m| < \varepsilon] = 1}$$

cioè, la successione della media campionaria converge in probabilità ad m, il valore atteso di tutte le X_n.

TEOREMA (Legge forte dei grandi numeri)
Date le ipotesi della legge debole:

$$\boxed{P[\lim_{n \to \infty} |\bar{X}_n - m| < \varepsilon] = 1}$$

questa conferma l'ipotesi precedente e ci dice che effettuando il procedimento la riuscita è certa

TEOREMA (Centrale del limite)
Sia X_n una successione di variabili aleatorie i.i.d provenienti da una popolazione X dotata di media e di varianza ($E[X] = m$, $Var[X] = v$), allora:

$$\boxed{\forall x \in \mathbb{R}, \quad \lim_{n \to \infty} P\left[\frac{\bar{X}_n - m}{\sqrt{v/n}} \leq x\right] = \phi(x)}$$

cioè la successione delle medie campionarie standardizzate converge in legge ad una $Norm(\mu = 0, \sigma^2 = 1)$.

$$\boxed{\bar{X}_n \sim Norm\left(\mu = m, \sigma^2 = \frac{v}{n}\right)}$$

<u>per n che tende a ∞.</u>

OSSERVAZIONI
1. $Bin(n, p) \approx Norm(\mu = np, \sigma^2 = np)$

   se $n \geq 30$, $np \geq 5$, $nq \geq 5$

2. $\overline{Bin}(n, p) \approx Norm\left(\mu = n/p, \sigma^2 = \frac{n(1-p)}{p^2}\right)$

   se $n \geq 30$

3. $Poi(\lambda) \approx Norm(\mu = \lambda, \sigma^2 = \lambda)$

   se $\lambda \geq 30$

DEF (Statistica)

Si dice statistica di una variabile aleatoria T, funzione nota delle variabili aleatorie $X_1, \dots, X_n$:

$$\boxed{T(X_1, \dots, X_n) = T(x)}$$

DEF (Stimatore)

Data una popolazione dipendente da un parametro θ e un campione casuale $X_1, X_2, \dots, X_n$ si dice stimatore di una statistica T che non dipenda da θ e che serva per stimare θ.

DEF (Correttezza e correttezza asintotica)

Uno stimatore si dice corretto (o non distorto) se il suo valore atteso coincide con la quantità da stimare.

$$\boxed{E[T_n] = \theta}$$

Si dirà asintoticamente corretto se:

$$\boxed{E[T_n] \xrightarrow{n \to \infty} \theta}$$

DEF (Consistenza)

Uno stimatore si dice consistente in media quadratica se:

$$\boxed{E[(T_n - \theta)^2] \xrightarrow{n \to \infty} 0}$$

Si dirà consistente in probabilità se:

$$\boxed{P[|T_n - \theta| < \varepsilon] \xrightarrow{n \to \infty} 0 \ , \forall \varepsilon > 0}$$

DEF (Asintotica normalità)

$$\boxed{\frac{T_n - E[T_n]}{\sqrt{Var[T_n]}} \overset{n \to \infty}{\sim} Norm(0,1)}$$

esempio

Sia $\theta > 0$. Supponiamo che X abbia densità:

$$f_X(x) = \begin{cases} \theta x^{\theta-1} & \text{se } 0 \le x \le 1 \\ 0 & \text{altrimenti} \end{cases}$$

Vogliamo stimare θ tramite il metodo dei momenti

$$E[X] = \int_0^1 x \theta x^{\theta-1} dx = \frac{\theta}{\theta+1}$$

$$E[X] = \frac{\theta}{\theta+1} \Rightarrow \theta = \frac{E[X]}{1-E[X]}$$

↳ stimatore di θ è $\hat{\theta} = \frac{\bar{X}_n}{1-\bar{X}_n}$

esempio

$$f_X(x) = \begin{cases} \frac{2\theta+1}{2} (x^2)^\theta & \text{se } -1 \le x \le 1 \\ 0 & \text{altrimenti} \end{cases} \qquad \theta \ge 0$$

$E[X] = 0 \Rightarrow$ non va bene come stimatore

$$E[X^2] = \int_{-1}^1 x^2 \frac{2\theta+1}{2} (x^2)^\theta dx = \frac{2\theta+1}{2\theta+3}$$

$$\theta = \frac{3E[X^2]-1}{2(1-E[X^2])}, \qquad M_n^2 = \frac{1}{n}\sum_{i=1}^n X_i^2$$

$$\hat{\theta} = \frac{3M_n^2 - 1}{2(1-M_n^2)}$$

METODO DELLA MASSIMA VEROSIMIGLIANZA

$X_1, \dots, X_n$

$$\mathcal{L}(X_1, \dots, X_n, \theta) = p_x(x_1, \theta) \cdot p_x(x_2, \theta) \cdot \dots \cdot p_x(x_n, \theta) = \prod_{i=1}^{n} p_x(x_i, \theta)$$

Scegliamo θ parametro θ che minimizza questa quantità

esempio:

$$f_x(x) = \begin{cases} \theta x^{\theta-1} & 0 \le x \le 1 \\ 0 & \text{altrimenti} \end{cases}, \quad \theta > 0$$

$$\mathcal{L}(\theta) = [\theta x_1^{\theta-1}] \cdot [\theta x_2^{\theta-1}] \cdot \dots \cdot [\theta x_n^{\theta-1}] = \theta^n \left[\prod_i x_i\right]^{\theta-1}$$

Calcolo θ attraverso la discesa del gradiente

$$\frac{d}{d\theta} \mathcal{L}(\theta) = \frac{n}{\theta} + \ln\left[\prod_i x_i\right] = 0$$

$$\theta = -\frac{n}{\ln\left[\prod_i x_i\right]} \quad \text{o} \quad \theta = -\frac{n}{\sum_i \ln x_i}$$

$P_x(x) = \theta^x (1-\theta)^{1-x}$

$Poi(\mu)$

$Esp(\lambda)$

$Norm(\mu, \sigma^2)$

VERIFICA DI IPOTESI

Quando effettuiamo una verifica, è nota la distribuzione della nostra variabile aleatoria mentre potrebbero essere incognite la media, la varianza, ecc. Facciamo i test perchè i dati che abbiamo a disposizione non ci permettono di arrivare a delle conclusioni.

STRUTTURA DI UN TEST DI IPOTESI

- ipotesi

  L'ipotesi di partenza è detta ipotesi nulla ed è indicata con H_0

  esempio

  $H_0: \mu = 5$ (ipotesi semplice)

  $H_0: \mu \leq 5$ oppure $\mu \geq 5$ (ipotesi composta)

  il contrario dell'ipotesi nulla è l'ipotesi alternativa e verrà indicata con H_1 o H_a, si tratta di ciò che dovremmo accettare per vero nel caso in cui rigettassimo l'ipotesi nulla.

- statistica test ($\mathcal{U}$)

  si tratta di una variabile aleatoria della quale, almeno approssimativamente possiamo specificare la distribuzione.
  Indicheremo con u il valore di $\mathcal{U}$.

- regione critica (C)

  si rifiuta H_0 se $u \in C$

  non si rifiuta H_0 se $u \notin C$

- errori possibili

  gli errori che si possono commettere sono i seguenti:

  H_0 è vera $\begin{cases} \text{se non si rifiuta } H_0 & \text{non si sbaglia} \\ \text{se si rifiuta } H_0 & \text{errore del I tipo} \end{cases}$

  H_0 è falsa $\begin{cases} \text{se non si rifiuta } H_0 & \text{errore del II tipo} \\ \text{se si rifiuta } H_0 & \text{non si sbaglia.} \end{cases}$

- Livello di significatività (α)

Livello di soglia α, tale che la probabilità di commettere l'errore del I tipo sia non superiore ad α.

Solitamente, $\alpha = 10\%, 5\%, 1\%$

- Funzione potenza

Se il test è un test parametrico, cioè riguarda un parametro, la probabilità di rifiutare l'ipotesi dipende dal valore incognito del parametro (θ).

Questa probabilità è detta funzione potenza del test:

$$\boxed{\pi(\theta) = P_\theta[\text{rifiutare } H_0]}$$

oppure

$$\boxed{\pi(\theta) = P_\theta[u \in C]}$$

- p-value

Ci si può chiedere a quale livello di significatività corrisponderebbe una regione critica la cui frontiera fosse costituita proprio da u, cioè per quale livello il valore osservato di u valica la frontiera di C. Questo valore prende il nome di p-value.

$$\boxed{\text{p-value} = \sup\{\alpha : u \notin C\} = \inf\{\alpha : u \in C\}}$$

Dato un test, rifiuteremo H_0 per tutti i livelli maggiori del p-value e non lo rifiuteremo per tutti i livelli inferiori.

TEST RIGUARDANTI UNA SOLA POPOLAZIONE

- test riguardanti la media di una normale

Supponiamo $X \sim Norm(\mu, \sigma^2)$ e sia n il numero di osservazioni di tale variabile aleatoria.
Possiamo fare diverse ipotesi su μ:

Tipo I $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$

μ_0 valore dato

Tipo II $\begin{cases} H_0 : \mu = \mu_0 \text{ o } \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$

Tipo III $\begin{cases} H_0 : \mu = \mu_0 \text{ o } \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$

La statistica test U utilizzata è diversa se σ^2 è nota o incognita. Nei due casi variano anche la regione critica e il p-value.

CASO 1: σ^2 NOTA: Z-TEST

$$\boxed{\bar{X} \overset{H_0}{\sim} Norm(\mu, \sigma^2 = v)}$$

equivale a:

$$\frac{\bar{X}_n - \mu}{\sqrt{\frac{v}{n}}} \sim Norm(0,1)$$

Scegliamo come statistica la standardizzazione che sarebbe corretto eseguire, se fosse vera H_0, ossia scegliamo per μ il valore μ_0:

$$\boxed{U = \frac{\bar{X}_n - \mu_0}{\sqrt{v/n}}}$$

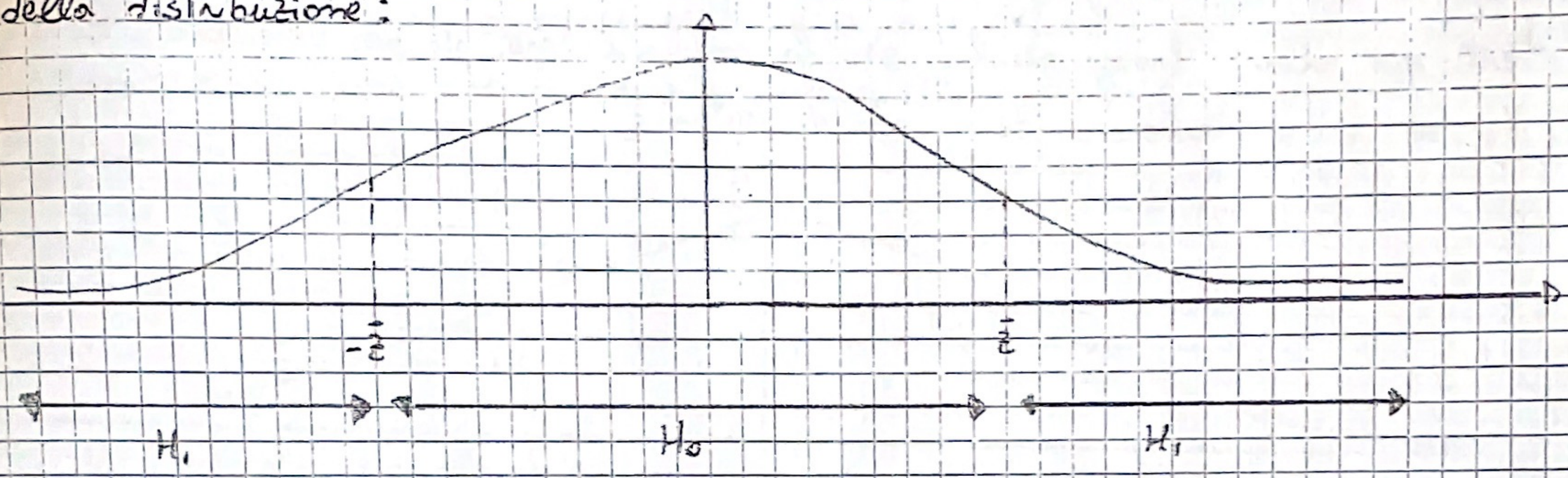
Scriveremo

$$\boxed{U = \frac{\bar{X}_n - \mu_0}{\sqrt{v/n}} \overset{H_0}{\sim} Norm(0,1)}$$

- Tipo I, dobbiamo decidere tra:

$$\boxed{H_0 : \mu = \mu_0 \quad e \quad H_1 : \mu \neq \mu_0}$$

propenderemo per H_0 se il valore osservato u di U sarà "vicino a zero" mentre propenderemo per H_1 se u si troverà in una delle due code della distribuzione:



Si deve stabilire il valore di $\bar{z}$ in corrispondenza del quale si inizia a rifiutare H_0. Poiché il nostro test ha livello di significatività α, vogliamo che:

$$\boxed{\alpha = \max P[\text{errore di I tipo}] = \max P[|U| > \bar{z} \mid H_0 \text{ vera}]}$$

cioè α deve essere il valore massimo per cui u, che nell'ipotesi è una normale standard finisca in una delle due code.

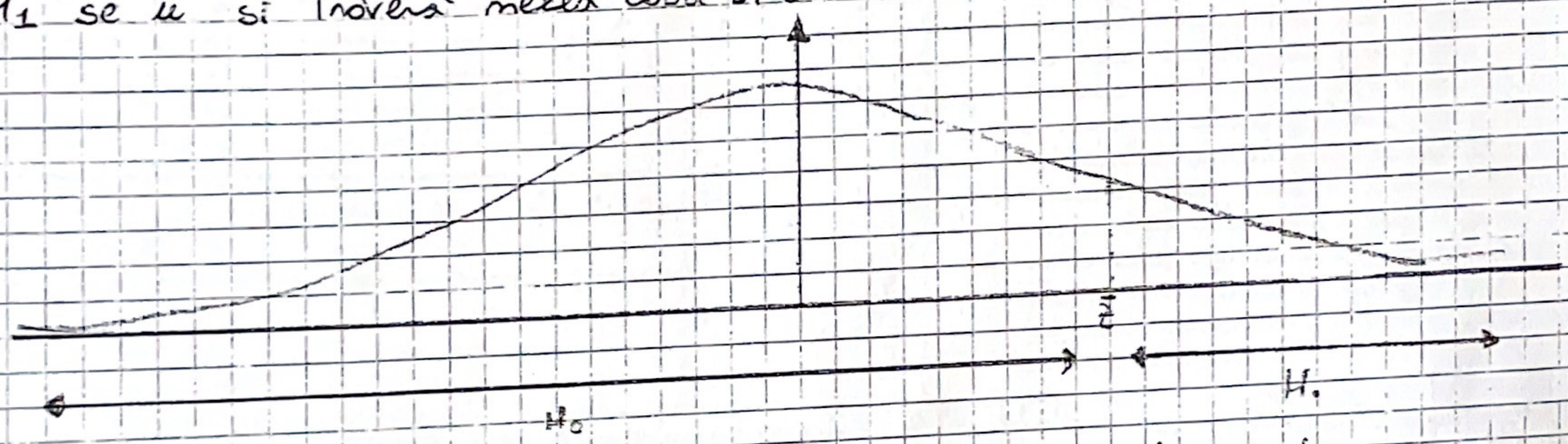
$$\boxed{\begin{array}{l} U \sim Norm(0,1) \\ P[|U| > \bar{z} > \alpha] \end{array} \Rightarrow \bar{z} = z_{1-\frac{\alpha}{2}}}$$

Quindi rifiuteremo H_0 se $|u| > z_{1-\frac{\alpha}{2}}$

- Se (tipo II) dobbiamo decidere tra:

$$\boxed{H_0 : \mu = \mu_0 \quad e \quad H_1 : \mu > \mu_0}$$

propenderemo per H_0 se il valore osservato u di U sarà "vicino a zero" mentre il fatto di aver utilizzato nell'esecuzione della statistica test un valore "troppo piccolo" per la media a farà propendere per H_1 se u si troverà nella coda di destra.



con conti analoghi ai precedenti, si trova $z = z_{1-\alpha}$ e pertanto rifiuteremo H_0 se $u > z_{1-\alpha}$.

- Se (tipo III) dobbiamo decidere tra:

$$\boxed{H_0 : \mu = \mu_0 \quad e \quad H_1 : \mu < \mu_0}$$

i conti si svolgono in maniera esattamente speculare al caso precedente e rifiuteremo H_0 se $u < -z_{1-\alpha}$ che equivale a $-u > z_{1-\alpha}$.

TABELLA RIASSUNTIVA

TIPO	Rifiuto H_0 se
I	$\lvert u \rvert > z_{1-\frac{\alpha}{2}}$
II	$u > z_{1-\alpha}$
III	$u < -z_{1-\alpha}$

ESEMPIO

Le batterie di una certa marca hanno durata nominale di 22 ore. Si sa che la deviazione standard del loro tempo di vita è $\sigma = 3,5$ ore.
Esaminando 20 batterie, è stata riscontrata una media di 20,7 ore.
Supponendo che le durate delle batterie siano variabili aleatorie normalmente distribuite, al livello di significatività del 5%, possiamo concludere che la durata media delle batterie sia inferiore a quanto dichiarato?

- X = "durata di una batteria"
- $X \sim Norm(\mu, \sigma^2 = (3,5)^2)$
- $H_0: \mu = 22$, $H_1: \mu < 22$

Con il livello di significatività assegnato H_0 viene rifiutata se $u < -z_{0.95}$ $(z_{1-\alpha})$. Si ricorda che $z_{0.95} = 1.645$ (consideriamo -1.645).

$$u = \frac{20.7 - 22}{\sqrt{\frac{12.25}{20}}} \approx 1.661 < -z_{0.95}$$

Quindi rifiutiamo H_0 e accettiamo H_1. Possiamo dire che il valore nominale non è corretto.

CASO 2: σ^2 INCOGNITA: T-TEST

$$\boxed{u = \frac{\bar{X}_n - \mu_0}{\sqrt{\frac{S_n^2}{n}}} \overset{H_0}{\sim} \text{T-student}(v = n-1)}$$

TIPO	Rifiuto H_0 se
I	$\lvert u \rvert > t_{1-\frac{\alpha}{2}}$
II	$u > t_{1-\alpha}$
III	$u < -t_{1-\alpha}$

ESEMPIO PRECEDENTE MA CON INCOGNITA

$$u = \frac{20.7 - 22}{\sqrt{12.25/20}} \approx -1.661 \quad ; \quad T_{0.95} = 1.7291 \quad ; \quad u \not< T_{0.95} \Rightarrow \text{accettiamo } H_0$$

da calcolare

TESTI RIGUARDANTI LA VARIANZA DI UNA NORMALE

Le tipologie di ipotesi sono sempre le seguenti:

- Tipo I $\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{cases}$

- Tipo II $\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \quad o \quad \sigma^2 \leq \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2 \end{cases}$

- Tipo III $\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \quad o \quad \sigma^2 \geq \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2 \end{cases}$

CASO 1: μ NOTA

$$\boxed{S_0^2 = \frac{1}{n} \sum_i (X_i - m)^2} \qquad S_n^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X}_n)^2$$

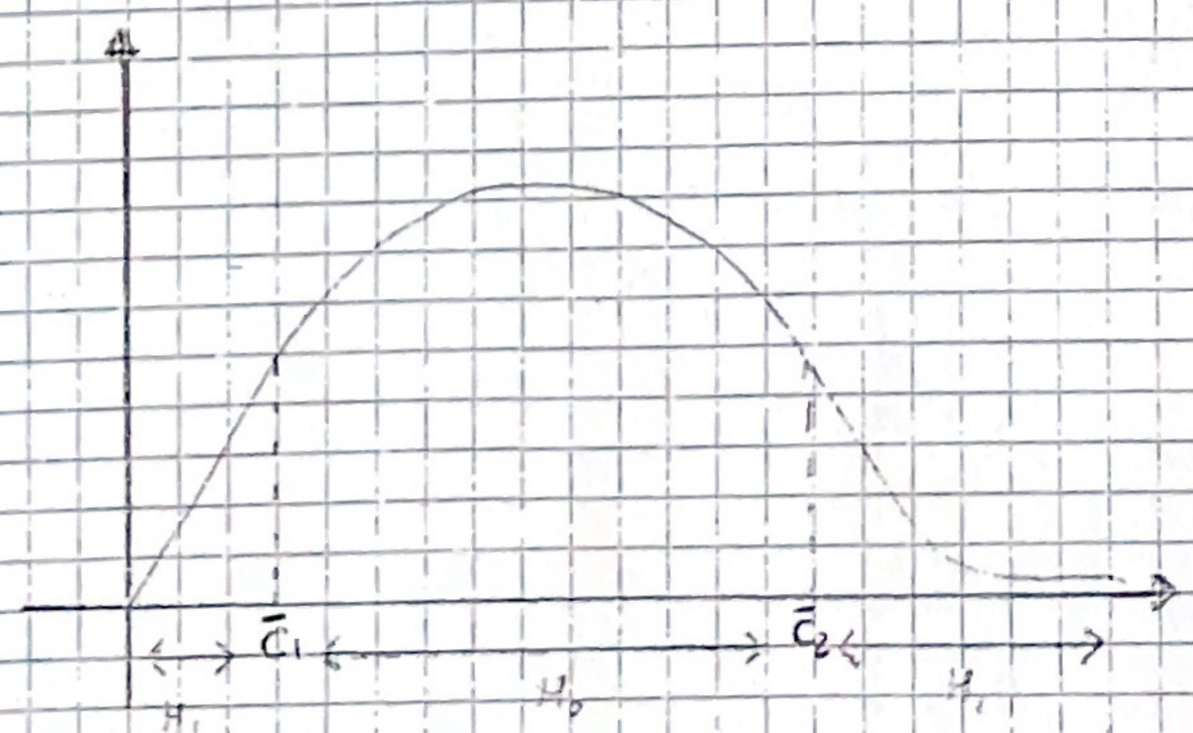
sappiamo che: $S_0^2 \overset{\sigma^2}{\sim} \chi^2(v=n) \iff \frac{n}{\sigma^2} S_0^2 \sim \chi^2(v=n)$

$$\boxed{U = \frac{n}{\sigma^2} S_0^2 \overset{H_0}{\sim} \chi^2(v=n)}$$

Se (Tipo I) dobbiamo decidere tra

$$\boxed{H_0 : \sigma^2 = \sigma_0^2 \quad e \quad H_1 : \sigma^2 \neq \sigma_0^2}$$

propenderemo per H_0 se il valore osservato u di U si troverà in corrispondenza della regione dove la densità della χ^2 è maggiore, mentre propenderemo per H_1 se u si troverà in una delle due code.



$$\boxed{\begin{aligned} \bar{c}_1 &= \chi^2_{\frac{\alpha}{2};n} \\ \bar{c}_2 &= \chi^2_{1-\frac{\alpha}{2};n} \end{aligned}}$$

rifiuteremo quindi H_0 in uno dei due casi

$$u < \chi^2_{\frac{\alpha}{2}} \qquad \text{oppure} \qquad u > \chi^2_{1-\frac{\alpha}{2}}$$

- Se (tipo II) dobbiamo decidere tra

$$\boxed{H_0 : \sigma^2 = \sigma_0^2 \quad e \quad H_1 : \sigma^2 > \sigma_0^2}$$

rifiuteremo H_0 se $u > \chi^2_{1-\alpha}$

- Se (tipo III) dobbiamo decidere tra

$$\boxed{H_0 : \sigma^2 = \sigma_0^2 \quad e \quad H_1 : \sigma^2 < \sigma_0^2}$$

rifiuteremo H_0 se $u < \chi^2_{\alpha}$

CASO 2: μ INCOGNITA

$$\boxed{\frac{(n-1)}{\sigma^2} S_n^2 \sim \chi^2(\nu = n-1)}$$

$$\boxed{U = \frac{(n-1)}{\sigma^2} S_n^2 \overset{H_0}{\sim} \chi^2(\nu = n-1)}$$

TIPO	Rifiuto H_0 se
I	$u < \chi^2_{\frac{\alpha}{2}}$ o $> \chi^2_{1-\frac{\alpha}{2}}$
II	$u > \chi^2_{1-\alpha}$
III	$u < \chi^2_{\alpha}$

esempio

Una pressa produce dischi di ferro il cui diametro X è distribuito come una normale, se la pressa non presenta problemi sappiamo che:

$$X \sim Norm(\mu = 7, \sigma^2 = 0.0001)$$

Per un controllo si misurino 10 dischi e si ottengono i seguenti valori:
7.064, 6.974, 7.036, 7.002, 6.976, 6.972, 7.041, 7.026, 7.026, 7.021, 6.978
è lecito concludere che la pressa funzioni correttamente?

$$\begin{cases} H_0: \mu = 7 \\ H_1: \mu \neq 7 \end{cases} \qquad \begin{cases} H_0: \sigma^2 = 0.0001 \\ H_1: \sigma^2 \neq 0.0001 \end{cases}$$

$\bar{X}_{10} = 7.009 \Rightarrow u_x = \dfrac{\bar{X}_m - \mu_0}{\sqrt{S_m^2 / m}} = 0.8581$

$S_{10}^2 = 0.0011 \Rightarrow u_S = \dfrac{(m-1)}{\sigma_0^2} S_m^2 = 99$

$\chi(v=9)$ H_0 non si rifiuta

$\chi^2_{0.99} = 21.6660$ e quindi rifiutiamo l'ipotesi nulla

TEST ASINTOTICO SU P, PROBABILITÀ DI SUCCESSO DI UN PROCESSO DI BERNOULLI

Se $X \sim Ber(p)$ per "n grandi":

$$\boxed{\bar{X}_n \approx Norm\left(\mu = p,\ \sigma^2 = \frac{p(1-p)}{n}\right)}$$

- Tipo I $\begin{cases} H_0: p = p_0 \\ H_1: p \neq p_0 \end{cases}$

- Tipo II $\begin{cases} H_0: p = p_0 \text{ oppure } p \leq p_0 \\ H_1: p > p_0 \end{cases}$

- Tipo III $\begin{cases} H_0: p = p_0 \text{ oppure } p \geq p_0 \\ H_1: p < p_0 \end{cases}$

Statistica:

$$\boxed{u = \frac{\bar{X}_m - p_0}{\sqrt{\frac{p_0 \times (1-p_0)}{n}}} \overset{H_0}{\sim} Norm(0,1)}$$

moltiplicazioni

Z-TEST

TIPO	Rifiutiamo H_0 se
I	$\lvert u \rvert > z_{1-\frac{\alpha}{2}}$
II	$u > z_{1-\alpha}$
III	$u < -z_{1-\alpha}$

TEST PARAMETRICI DI CONFRONTO TRA DUE POPOLAZIONI

Supponiamo di avere due popolazioni:

$$\boxed{X_1, X_2, \dots, X_n \quad e \quad Y_1, Y_2, \dots, Y_m}$$

con:

$$X \sim Norm(\mu = \mu_x, \sigma^2 = \sigma_x^2) \quad e \quad Y \sim Norm(\mu = \mu_y, \sigma^2 = \sigma_y^2)$$

I test saranno

Tipo I: $\begin{cases} H_0: \mu_x = \mu_y \\ H_1: \mu_x \neq \mu_y \end{cases}$

tipo II $\begin{cases} H_0: \mu_x = \mu_y \text{ oppure } \mu_x \leq \mu_y \\ H_1: \mu_x > \mu_y \end{cases}$

tipo III $\begin{cases} H_0: \mu_x = \mu_y \text{ oppure } \mu_x \geq \mu_y \\ H_1: \mu_x < \mu_y \end{cases}$

Osserviamo che:

$$\bar{X}_n \sim Norm(\mu = \mu_x, \sigma^2 = \frac{\sigma_x^2}{n})$$

$$\bar{Y}_m \sim Norm(\mu = \mu_y, \sigma^2 = \frac{\sigma_y^2}{m})$$

Quindi:

$$\boxed{\bar{X}_n - \bar{Y}_n \sim Norm(\mu = \mu_x - \mu_y, \sigma^2 = \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m})}$$

CASO 1: VARIANZA NOTA

$$\frac{(\bar{X}_n - \bar{Y}_n) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \sim Norm(0,1)$$

Allora:

$$\boxed{u = \frac{(\bar{X}_n - \bar{Y}_n) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \overset{H_0}{\sim} Norm(0,1)}$$

Tipo	Rifiuto H_0 se
I	$\lvert u \rvert > z_{1-\alpha/2}$
II	$u > z_{1-\alpha}$
III	$u < -z_{1-\alpha}$

ESEMPIO

Confronto delle prestazioni sportive prima e dopo le vacanze natalizie.
Prima delle vacanze 5 corse, tempo medio 53,82 secondi.
Dopo le vacanze 6 corse, tempo medio 54,4 secondi.
Supponendo che il tempo impiegato per percorrere 400 metri sia distribuito normalmente e la varianza sia uguale ($\sigma^2 = 0.1$), le vacanze hanno contribuito a peggiorare le prestazioni? $\alpha = 5\%$

X = "tempo prima delle vacanze" $\sim Norm(\mu_x, \sigma^2 = 0.1)$
Y = "tempo dopo le vacanze" $\sim Norm(\mu_y, \sigma^2 = 0.1)$

$H_0: \mu_x = \mu_y$, le vacanze non hanno influito.
$H_1: \mu_x \neq \mu_y$, le vacanze hanno influito.

$$u = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} = -1.6611$$

$z_{1-\alpha} = 1.6449 \Rightarrow u < -z_{1-\alpha} \Rightarrow$ rifiutiamo H_0

CASO 2: VARIANZE INCOGNITE UGUALI ($\sigma_x^2 = \sigma_y^2 = V$)

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

$$S_y^2 = \frac{1}{m-1} \sum_{i=1}^{m} (y_i - \bar{y}_m)^2$$

$$\hat{V} = S_p^2 = \frac{(n-1) S_x^2 + (m-1) S_y^2}{n+m+2}$$

$\hat{V}$ prende il nome di varianza combinata

La statistica test è data da:

$$\boxed{u = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{S_p^2 \left(1/n + 1/m\right)}} \overset{H_0}{\sim} T(v = n+m-2)}$$

TIPO	Rifiuto di Ho se
I	$\lvert u \rvert > T_{1-\frac{\alpha}{2}; n+m-2}$
II	$u > T_{1-\alpha; n+m-2}$
III	$u < -T_{1-\alpha; n+m-2}$

CASO 3: VARIANZE INCOGNITE NON UGUALI

$$\boxed{U = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} \overset{H_0}{\sim} T(v = v^*)}$$

con $v^* =$ Parte intera $\left(\frac{\left(\frac{S_X}{n} + \frac{S_Y}{m}\right)^2}{\frac{1}{n-1}\left(\frac{S_X^2}{n}\right)^2 + \frac{1}{m-1}\left(\frac{S_Y^2}{m}\right)^2} \right)$

CONFRONTO DI DUE NORMALI A CAMPIONI ACCOPPIATI

Abbiamo un campione di ampiezza n, cioè formato da n coppie di osservazioni

(X_1, Y_1) ; (X_2, Y_2) ; . . . (X_n, Y_n)

X e Y sono normali e la loro differenza ($W = X - Y$) sarà anch'essa normalmente distribuita.

$$\boxed{W \sim Norm(\mu_W, \sigma_W^2)}$$

Con $\mu_W = \mu_X - \mu_Y$ e σ_W^2 incognita

Tipo I: $\begin{cases} H_0 : \mu_W = 0 \\ H_1 : \mu_W \neq 0 \end{cases}$

Tipo II: $\begin{cases} H_0 : \mu_W = 0 \text{ oppure } \mu_W \leq 0 \\ H_1 : \mu_W > 0 \end{cases}$

Tipo III: $\begin{cases} H_0 : \mu_W = 0 \text{ oppure } \mu_W \geq 0 \\ H_1 : \mu_W < 0 \end{cases}$

$S_W^2 = \frac{1}{n-1} \sum_{i=1}^{n} (W_i - \bar{W}_n)^2$; $\boxed{U = \frac{\bar{W}_n}{\sqrt{\frac{S_W^2}{n}}} \overset{H_0}{\sim} T(v = n-1)}$

TIPO	Rifiuto di H_0
I	$\lvert U \rvert > T_{1-\frac{\alpha}{2}; n+m-2}$
II	$U > T_{1-\alpha; n+m+2}$
III	$U < -T_{1-\alpha; n+m+2}$

Esempio

PAZIENTE	1	2	3	4	5	6
GLICEMIA PRIMA	170.3	162.2	210.3	158.3	180.1	192.3
GLICEMIA DOPO	166.5	165.3	171.1	163.2	152.2	161.5

REGRESSIONE LINEARE

Siano Y e X due variabili aleatorie, dobbiamo capire se esiste una relazione lineare tra di loro.

- Y prende il nome di variabile risposta (variabile dipendente);
- X prende il nome di variabile esplicativa o predittore (variabile indipendente)

Avremo:

$$\boxed{Y = \beta_0 + \beta X + \varepsilon}$$

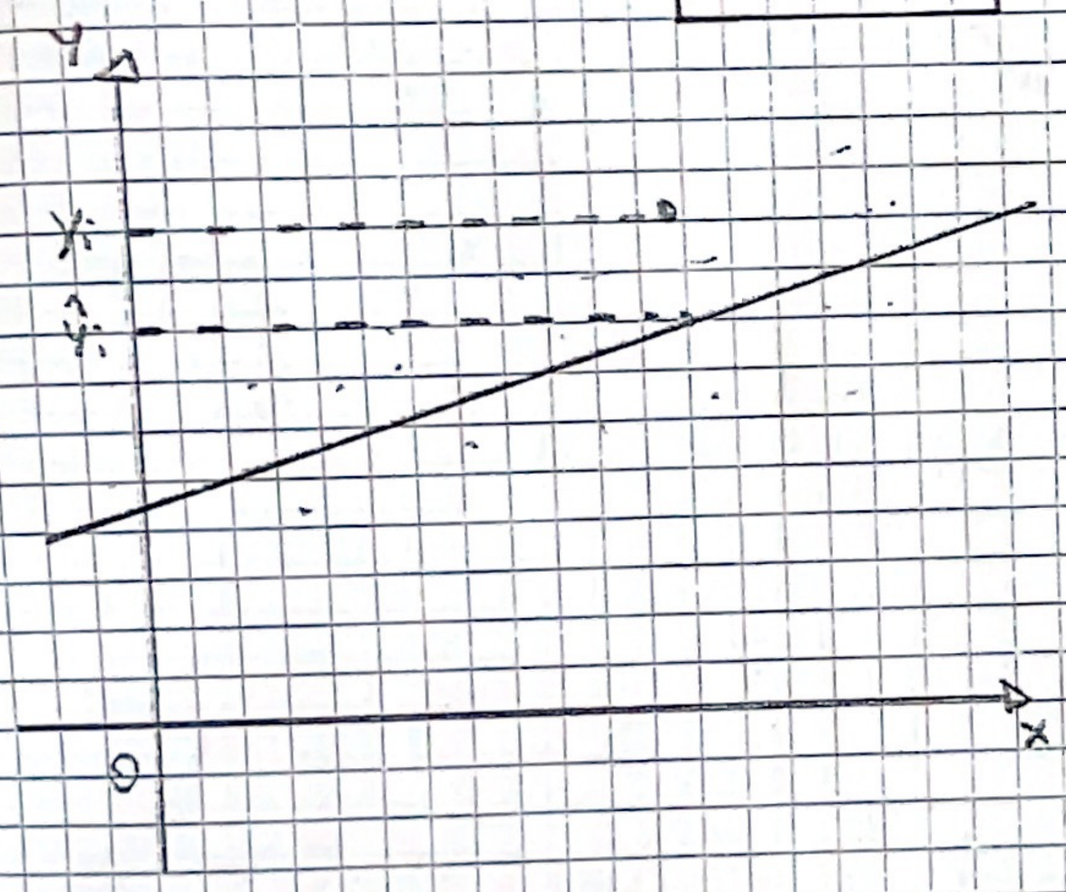
è errore

Vengono raccolte n osservazioni (x_i, y_i) e l'obiettivo è quello di stimare i coefficienti β_0 e β i cui stimatori verranno indicati con $\hat{\beta}_0$ e $\hat{\beta}$.

METODO DEI MINIMI QUADRATI

Questo metodo permette di stimare i valori di β_0 e β.

$$\boxed{y = \hat{\beta}_0 + \hat{\beta} X}$$



$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta} X$$

$$y_i - \hat{y}_i = y_i - (\beta_0 + \beta x_i)$$

Scegliamo come stimatori quei particolari valori che minimizzano la somma dei quadrati delle differenze.

$$\boxed{\sum_j (y_j - (\beta_0 + \beta x_j))^2}$$

Ovvero, vogliamo trovare la retta che minimizza la distanza tra y_i e $\hat{y}_i$:

$$\boxed{f(\beta_0, \beta) = \sum_j (y_j - (\beta_0 + \beta x_j))^2}$$

PROPOSIZIONE

La coppia di valori $(\hat{\beta}_0, \beta)$ che minimizza la funzione $f(\beta_0, \beta)$ è data da:

$$\boxed{\hat{\beta} = \frac{\sum_j (x_j - \bar{x})(y_j - \bar{y})}{\sum_j (x_j - \bar{x})^2}}$$

con $\bar{x}$ e $\bar{y}$ valori reali.

$$\boxed{\hat{\beta}_0 = \bar{y} - \hat{\beta}\bar{x}}$$

DIM

$$f(\beta_0, \beta) = \sum_j (y_j - (\beta_0 + \beta x_j))^2$$

① $\frac{\delta f}{\delta \beta_0} = -2 \sum_j (y_j - (\beta_0 + \beta x_j)) = 0$

cerchiamo i punti stazionari

② $\frac{\delta f}{\delta \beta} = -2 \sum_j x_j (y_j - (\beta_0 + \beta x_j)) = 0$

① $\Rightarrow \sum_j y_j - \sum_j \beta_0 - \beta \sum_j x_j = 0$

$\bar{y} = \frac{1}{m} \sum_j y_j \quad ; \quad \bar{x} = \frac{1}{m} \sum_j x_j$

$m\bar{y} - m\beta_0 - m\beta\bar{x} = 0$

$\beta_0 = \bar{y} - \beta\bar{x}$

② $\Rightarrow \sum_j x_j y_j - \beta_0 \sum_j x_j - \beta \sum_j x_j^2 = 0$

sostituiamo β_0

$\Rightarrow \sum_j x_j y_j - (\bar{y} - \beta\bar{x}) \sum_j x_j - \beta \sum_j x_j^2 = 0$

$\sum_j x_j y_j - \bar{y} \sum_j x_j + \beta\bar{x} \sum_j x_j - \beta \sum_j x_j^2 = 0$

$\sum_j (y_j - \bar{y}) x_j - \beta \sum_j (x_j - \bar{x}) x_j = 0$

$$\beta = \frac{\sum_j (y_j - \bar{y}) x_j}{\sum_j (x_j - \bar{x}) x_j}$$

Si osserva che:

$\sum_j \bar{x} (x_j - \bar{x}) = 0$

$$\Rightarrow \beta = \frac{\sum_j (y_j - \bar{y}) x_j - \sum_j (y_j - \bar{y}) \bar{x}}{\sum_j (x_j - \bar{x}) x_j - \sum_j (x_j - \bar{x}) \bar{x}}$$

$$\Rightarrow \frac{\sum_j (y_j - \bar{y})(x_j - \bar{x})}{\sum_j (x_j - \bar{x})^2} = \hat{\beta}$$

$$\hat{y}_j = \hat{\beta}_0 + \hat{\beta} x_j$$

$$z_j = y_j - \hat{y}_j = y_j - (\hat{\beta}_0 + \hat{\beta} x_j) \quad [\text{residuo}]$$

$$\hat{V} = \frac{1}{n-2} \sum_j (Y_j - \hat{Y}_j)^2$$

$$\hat{V} = \frac{1}{n-2} \sum_j (y_j - \hat{y}_j)^2 = \frac{1}{n-2} \sum_j z_j^2$$

NOTAZIONE

$$S_{xx} = \sum_j (x_j - \bar{x})^2 = \sum_j x_j^2 - n\bar{x}$$

$$S_{yy} = \sum_j (y_j - \bar{y})^2$$

$$S_{xy} = \sum_j (x_j - x)(y_j - \bar{y}) = \sum_j x_j y_j - n\bar{x}\bar{y}$$

DEF

$$SS_{Reg} = \sum_j ((\hat{\beta}_0 + \beta x_j) - \bar{y})^2 \quad \text{varianza spiegata}$$

$$SS_{Res} = \sum_j (Y_j - (\hat{\beta}_0 + \hat{\beta} x_j))^2 \quad \text{varianza residua}$$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\beta}_0 = \bar{Y} - \beta \bar{X}$$

$$\hat{V} = \frac{SS_{Res}}{n-2}$$

$$SS_{Reg} = \frac{(S_{xy})^2}{S_{xx}}$$

$$S_{xx} = \frac{(S_{xy})^2}{S_{yy}}$$

OSSERVAZIONE

Le medie campionarie per y_j e $\hat{y}_j$ coincidono

$$\boxed{\sum_j y_j = \sum_j (\hat{\beta}_0 + \beta x_j) = \sum_j (\bar{y} - \beta \bar{x} + \beta x_j) = \sum_j \bar{y} = m\bar{y}} \; ; \text{ con } \bar{y} = \frac{1}{m} \sum_j y_j$$

VARIANZE PER y

$$S_y^2 = \frac{1}{m-1} \sum_j (y_j - \bar{y}) = \frac{1}{m-1} S_{yy}$$

$$S_{\hat{y}}^2 = \frac{1}{m-1} \sum_j ((\hat{\beta}_0 + \hat{\beta} x_j) - \bar{y}) = \frac{1}{m-1} SS_{Reg}$$

TEOREMA

La varianza totale è la somma della varianza spiegata più la varianza residua.

$$\boxed{S_{yy} = SS_{Reg} + SS_{Res}}$$

$$\Rightarrow SS_{Res} = S_{yy} - SS_{Reg} = S_{yy} - \frac{(S_{xy})^2}{S_{xx}} = \frac{(S_{xx} S_{yy}) - (S_{xy})^2}{S_{xx}}$$

COEFFICIENTE DI DETERMINAZIONE (R^2)

$$\boxed{R^2 = \frac{SS_{Reg}}{SS_{yy}} \approx 1}$$

coefficiente di correlazione e coefficiente di Pearson coincidono e in particolare:

$$\text{Correlazione} = \sqrt{\text{DETERM VARIOUE}}$$

oppure

$$R^2 = \frac{S_{yy} - SS_{Reg}}{S_{yy}} = 1 - \frac{SS_{Res}}{S_{yy}} \approx 0$$

Si dimostra che

$$\boxed{R^2 = \left(\frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} \right)^2}$$

$$\boxed{r = \hat{\beta} \sqrt{\frac{S_{xx}}{S_{yy}}}}$$

COEFFICIENTE DI CORRELAZIONE CORRETTO

$$\boxed{R^2_{corr} = 1 - \frac{SS_{Res} / m-2}{S_{yy} / (m-1)}}$$

TEST RELATIVI AI PARAMETRI

IPOTESI DI NORMALITÀ

$$\boxed{E \sim Norm(\mu = 0, \sigma^2 = v)}$$

PROPRIETÀ

$$\hat{\beta} \sim Norm(\mu = \beta, \sigma^2 = v/S_{xx})$$

$$\hat{\beta}_0 \sim Norm\left(\mu = \beta_0, \sigma^2 = v\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\right)$$

Si ha che

$$\frac{(n-2)}{v}\hat{v} = \frac{1}{v}\sum_j (y_j - \hat{y}_j) \sim \chi^2(\nu = n-2)$$

TEST SU β

$$\frac{\hat{\beta} - \beta}{\sqrt{\frac{v}{S_{xx}}}} \sim Norm(\mu = 0, \sigma^2 = 1)$$

v è incognita quindi consideriamo la stima $\hat{v}$

$$\boxed{\frac{\hat{\beta} - \beta}{\sqrt{\frac{\hat{v}}{S_{xx}}}} \sim T(\nu = n-2)}$$

TEST DI IPOTESI

Tipo I $\begin{cases} H_0: \beta = \beta^* \\ H_1: \beta \neq \beta^* \end{cases}$

Tipo II $\begin{cases} H_0: \beta = \beta^* \text{ oppure } \beta \leq \beta^* \\ H_1: \beta > \beta^* \end{cases}$

Tipo III $\begin{cases} H_0: \beta = \beta^* \text{ oppure } \beta \geq \beta^* \\ H_1: \beta < \beta^* \end{cases}$

per la normale con u
per Student T
per χ^2 = χ^2

STATISTICA TEST

$$U = \frac{\tilde{\hat{\beta}}}{\sqrt{\frac{\hat{v}}{S_{xx}}}} \overset{H_0}{\sim} T(\nu = n-2) \quad \text{con } \tilde{\hat{\beta}} = \hat{\beta} - \beta^*.$$

Valgono le tabelle

ESERCIZIO 5 - PROBLEMI DI STIMA

Si supponga che 10 ratti sono utilizzati per uno studio biomedico. A ciascuno di essi vengono iniettate delle cellule tumorali e poi viene somministrato un farmaco che permette di aumentare il tasso di sopravvivenza. I tempi di sopravvivenza in mesi sono [14, 17, 27, 18, 12, 8, 22, 13, 19, 12]. Si supponga che si possa applicare una distribuzione esponenziale e si trovi una stima di massima verosimiglianza del tempo medio di sopravvivenza.

$$f(x;\beta) = \frac{1}{\beta} e^{-\frac{x}{\beta}} \quad \text{per } x > 0$$

$$\mathcal{L}(\underline{x};\beta) = \prod_{i=1}^{n} f(x_i;\beta) = \prod_{i=1}^{n} \frac{1}{\beta} e^{-\frac{x_i}{\beta}} = \left(\frac{1}{\beta} e^{-\frac{x_1}{\beta}}\right) \cdot \left(\frac{1}{\beta} e^{-\frac{x_2}{\beta}}\right) \cdot \ldots \cdot \left(\frac{1}{\beta} e^{-\frac{x_n}{\beta}}\right) =$$

$$= \frac{1}{\beta^n} e^{-\frac{x_1+x_2+\ldots+x_n}{\beta}} = \frac{1}{\beta^n} e^{-\frac{n\bar{x}}{\beta}} \quad \text{con } \bar{x} = \frac{x_1+x_2+\ldots+x_n}{n}$$

$$\mathcal{L}(x,\beta) = \frac{1}{\beta^n} e^{-\frac{n\bar{x}}{\beta}}$$

Calcoliamo la derivata parziale rispetto a β

$$\frac{\partial L}{\partial \beta} = -n\beta^{-n-1} e^{-\frac{1}{\beta}n\bar{x}} + \frac{1}{\beta^n} e^{-\frac{1}{\beta}n\bar{x}} \left(n\bar{x}\frac{1}{\beta^2}\right) = -\frac{n}{\beta^{n+1}} e^{-\frac{1}{\beta}n\bar{x}} + \frac{1}{\beta^n} n\bar{x} e^{-\frac{1}{\beta}n\bar{x}}$$

Imponiamo $\frac{\partial L}{\partial \beta} = 0$

$$\frac{n}{\beta^{n+1}} e^{-\frac{1}{\beta}n\bar{x}} \left(-1 + \frac{\bar{x}}{\beta^2}\right) = 0$$

$$\downarrow$$

$$\frac{\bar{x}}{\beta} - 1 = 0 \Rightarrow \beta = \bar{x}$$

Quindi il valore che massimizza β è la media $\bar{x}$.

Ricordiamo che $n = 10$ e $\bar{x} = 16.20$ calcolata sperimentalmente.

ESERCIZIO 6 - PROBLEMI DI STIMA

È noto che un campione costituito dai valori $x = [12, 11.2, 13.5, 12.3, 13.8, 11.9]$ proviene da una popolazione con funzione di densità:

$$f(x;\theta) = \frac{\theta}{x^{\theta+1}} \quad \text{per } x \geq 1 \text{ e } f(x;\theta) = 0 \text{ altrove}$$

Dove $\theta > 0$, si trovi la stima della funzione di massima verosomiglianza θ.

$\bar{x} = 12,45$

$$\mathcal{L}(\underline{x};\theta) = \frac{\theta}{x_1^{\theta+1}} \cdot \frac{\theta}{x_2^{\theta+1}} \cdot \ldots \cdot \frac{\theta}{x_n^{\theta+1}} = \frac{\theta}{x_1^{\theta+1} \cdot x_2^{\theta+1} \cdot \ldots \cdot x_n^{\theta+1}} = \frac{\theta}{n\,\bar{x}^{\theta+1}}$$

$$\frac{\partial \mathcal{L}}{\partial \theta}\left(\frac{\theta}{n\,\bar{x}^{\theta+1}}\right) \Rightarrow \theta \cong 0.39 \quad \text{giusto}$$

TEST ASINTOTICO SU P, PROBABILITA' DI SUCCESSO DI UN PROCESSO DI BERNOULLI

Se $X \sim Ber(p)$

Tipo I $\begin{cases} H_0: p = p_0 \\ H_1: p \neq p_0 \end{cases}$

Tipo II $\begin{cases} H_0: p = p_0 \text{ oppure } p \leq p_0 \\ H_1: p > p_0 \end{cases}$

Tipo III $\begin{cases} H_0: p = p_0 \text{ oppure } p \geq p_0 \\ H_1: p < p_0 \end{cases}$

Quando facciamo il test dobbiamo fare in modo che il nostro dubbio sia in H_1 così da poterlo eventualmente non rifiutare

ad es, se vogliamo testare $\mu \geq \mu_0$ facciamo

$\begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$

$$u = \frac{\bar{X}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \overset{H_0}{\sim} Norm(0,1)$$

TIPO	RIFIUTO H0 SE
I	$\lvert u \rvert > z_{1-\alpha/2}$
II	$u > z_{1-\alpha}$
III	$u < -z_{1-\alpha}$

TEST PARAMETRICI DI CONFRONTO TRA DUE POPOLAZIONI

$X \sim Norm(\mu = \mu_x, \sigma^2 = \sigma_x^2)$, $Y \sim Norm(\mu = \mu_y, \sigma^2 = \sigma_y^2)$

Tipo I $\begin{cases} H_0: \mu_x = \mu_y \\ H_1: \mu_x \neq \mu_y \end{cases}$

Tipo II $\begin{cases} H_0: \mu_x = \mu_y \text{ oppure } \mu_x \leq \mu_y \\ H_1: \mu_x > \mu_y \end{cases}$

Tipo III $\begin{cases} H_0: \mu_x = \mu_y \text{ oppure } \mu_x \geq \mu_y \\ H_1: \mu_x < \mu_y \end{cases}$

VARIANZA NOTA

$$u = \frac{(\bar{X}_n - \bar{Y}_n) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \overset{H_0}{\sim} Norm(0,1)$$

TIPO	RIFIUTO H0 SE
I	$\lvert u \rvert > z_{1-\alpha/2}$
II	$u > z_{1-\alpha}$
III	$u < -z_{1-\alpha}$

VARIANZE INCOGNITE UGUALI

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

$$S_Y^2 = \frac{1}{m-1}\sum_{i=1}^{m}(y_i - \bar{y}_m)^2$$

$$\hat{V} = S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m+2}$$

$$u = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{S_p^2\left(1/n + 1/m\right)}} \overset{H_0}{\sim} T(\nu = n+m-2)$$

TIPO	RIFIUTO DI H_0 SE
I	$\lvert u\rvert > t_{1-\frac{\alpha}{2};\, n+m-2}$
II	$u > t_{1-\alpha;\, n+m-2}$
III	$u < -t_{1-\alpha;\, n+m-2}$

VARIANZE INCOGNITE NON UGUALI

$$u = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} \overset{H_0}{\sim} T(\nu = \nu^*) \quad \text{con } \nu^* = P.I.\left(\frac{\left(\frac{S_X}{n} + \frac{S_Y}{m}\right)^2}{\frac{1}{n-1}\left(\frac{S_X^2}{n}\right)^2 + \frac{1}{m-1}\left(\frac{S_Y^2}{m}\right)^2}\right)$$

CONFRONTO DI DUE NORMALI A CAMPIONI ACCOPPIATI

$W \sim Norm(\mu_W, \sigma_W^2)$ con $\mu_W = \mu_X - \mu_Y$ e σ_W^2 incognita

Tipo I $\begin{cases} H_0: \mu_W = 0 \\ H_1: \mu_W \neq 0 \end{cases}$

Tipo II $\begin{cases} H_0: \mu_W = 0 \text{ oppure } \mu_W \leq 0 \\ H_1: \mu_W > 0 \end{cases}$

Tipo III $\begin{cases} H_0: \mu_W = 0 \text{ oppure } \mu_W \geq 0 \\ H_1: \mu_W < 0 \end{cases}$

TIPO	RIFIUTO H_0 SE
I	$\lvert u\rvert > T_{1-\alpha/2;\, \cancel{n+m-2}\, n-1}$
II	$u > T_{1-\alpha;\, \cancel{n+m-2}\, n-1}$
III	$u < -T_{1-\alpha;\, \cancel{n+m-2}\, n-1}$

media di W (diff)

$$u = \frac{\bar{W}_n}{\sqrt{\frac{S_W^2}{n}}} \overset{H_0}{\sim} T(\nu = n-1) \; ; \; S_W^2 = \frac{1}{n-1}\sum_{i=1}^{n}(w_i - \bar{w}_n)^2$$

METODI MATEMATICI E STATISTICI

MEDIA CAMPIONARIA

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

MOMENTO CAMPIONARIO DI ORDINE K

$$M_n^{(k)} = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

$\alpha = 1\%, 2\%, 5\%$

z = norminv.

VARIANZA CAMPIONARIA

- Valore atteso noto ($E[X] = m$)

$$S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2$$

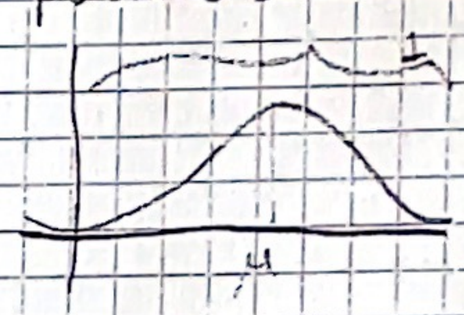
- valore atteso incognito

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

norminv (prob) ⇒ torna il punto

normcdf (punto) ⇒ torna la prob.

- partono da $-\infty$



PROPRIETÀ

$$S_0^2 = M_n^{(2)} - 2m\bar{X}_n + m^2$$

$$S_n^2 = \frac{1}{n-1}\left(nM_n^{(2)} - n\bar{X}_n^2\right)$$

VALORE ATTESO E VARIANZA DELLA MEDIA CAMPIONARIA

$$E[\bar{X}_n] = m$$

$$Var[\bar{X}_n] = 1/n \cdot V$$

VALORE ATTESO DELLA VARIANZA CAMPIONARIA

Se $Var[X] = V \Rightarrow E[S_0^2] = V$; $E[S_n^2] = V$

TEST RIGUARDANTI LA MEDIA DI UNA NORMALE

$X \sim Norm(\mu, \sigma^2)$

Tipo I $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$

Tipo II $\begin{cases} H_0 : \mu = \mu_0 \text{ oppure } \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$

Tipo III $\begin{cases} H_0 : \mu = \mu_0 \text{ oppure } \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$

σ^2 NOTA (Z-TEST)

$$u = \frac{\bar{X}_n - \mu_0}{\sqrt{\frac{V}{n}}} \overset{H_0}{\sim} Norm(0,1)$$

TIPO	RIFIUTO H_0 SE
I	$\lvert u \rvert > z_{1-\alpha/2}$
II	$u > z_{1-\alpha}$
III	$u < -z_{1-\alpha}$

μ INCOGNITA (T-TEST)

$$u = \frac{\bar{X}_n - \mu_0}{\sqrt{S_n^2/n}} \overset{H_0}{\sim} T\text{-Student}(\nu = n-1) \qquad t = t_{inv.}$$

TIPO	RIFIUTO H0 SE
I	$\lvert u \rvert > t_{1-\alpha/2}$
II	$u > t_{1-\alpha}$
III	$u < -t_{1-\alpha}$

TEST RIGUARDANTI LA VARIANZA DI UNA NORMALE

Tipo I $\begin{cases} H_0: \sigma^2 = \sigma_0^2 \\ H_1: \sigma^2 \neq \sigma_0^2 \end{cases}$

Tipo II $\begin{cases} H_0: \sigma^2 = \sigma_0^2 \text{ oppure } \sigma^2 \leq \sigma_0^2 \\ H_1: \sigma^2 > \sigma_0^2 \end{cases}$

Tipo III $\begin{cases} H_0: \sigma^2 = \sigma_0^2 \text{ oppure } \sigma^2 \geq \sigma_0^2 \\ H_1: \sigma^2 > \sigma_0^2 \end{cases}$

μ NOTA

$$u = \frac{n}{\sigma^2} S_0^2 \overset{H_0}{\sim} \chi^2(\nu = n) \qquad \chi^2_\alpha = \text{chi2 inv}$$

TIPO	RIFIUTO H0 SE
I	$\lvert u \rvert < \chi^2_{\frac{\alpha}{2}}$ o $\lvert u \rvert > \chi^2_{1-\frac{\alpha}{2}}$
II	$u > \chi^2_{1-\alpha}$
III	$u < \chi^2_\alpha$

μ INCOGNITA

$$u = \frac{(n-1)}{\sigma^2} S_n^2 \overset{H_0}{\sim} \chi^2(\nu = n-1)$$

TIPO	RIFIUTO H0 SE
I	$u < \chi^2_{\alpha/2}$ oppure $u > \chi^2_{1-\alpha/2}$
II	$u > \chi^2_{1-\alpha}$
III	$u < \chi^2_\alpha$