

STATISTICA

STATISTICA DESCRITTIVA

Introduzione	1
Frequenza	1
Diagramma a barre	2
Indice statistico	2
Media	2
Mediana	2
Moda	2
Media Sfrondata	3
Indici di variabilità	3
Varianza campionaria	3
Varianza empirica	3
Rapporto tra $\text{var}(x)$ e $\text{var}_e(x)$	3
Deviazione standard	3
Sample Skewness	3
Funzione di ripartizione empirica	3
K-simo percentile/B-quantile	4
Box Plot e Outlier	5
K-uple di vettori	5
Diagramma a dispersione	5
Problema della regressione	5
Covarianza	5
Covarianza campionaria	5
Covarianza empirica	5
Relazione tra $\text{cov}(x,y)$ e $\text{cov}_e(x,y)$	5
Coefficiente di correlazione	6
Funzione retta	6
Retta di regressione	6

PROBABILITÀ E INDEPENDENZA

Introduzione	7
Teoria degli insiemi	7
Probabilità	7
Definizione classica (Laplace)	7
Definizione frequentista	8
Definizione assiomatica	8

Spazio di probabilità	8
Proprietà di una probabilità	8
Formula di inclusione-esclusione	8
Modello uniforme	9
Calcolo combinatorio	9
Probabilità condizionata	9
Probabilità condizionata di A dato B	9
Regola dei prodotti	9
Condizionamento ripetuto	9
Rappresentazione ad albero	10
Regola della catena	10
Regola della fattorizzazione	10
Sistema di alternative per esperimento con spazio campionario Ω	10
Formula di fattorizzazione	10
Formula di Bayes	11

EVENTI INDIPENDENTI

Eventi indipendenti	12
Indipendenza di 3 eventi	12
Indipendenza di n eventi	12
Schema di Bernoulli	13

VARIABILI ALEATORIE

Variabile aleatoria	14
Legge di probab. di una variabile aleatoria	14
V.A. Equidistribuite	15
Notazione	15
V.A. discrete	15
Funzione di massa	15
Calcolo delle probabilità di X usando P_x	16
V.A. assolutamente continua	16
Proprietà della funzione densità	16
Calcoli al variare di $A \subseteq \mathbb{R}$	17
Interpretazioni delle V.A.	17
Funzione di ripartizione	17
Proprietà di F_x	17
F_x per V.A. discrete	18
F_x per V.A. continue	18

Calcolo della probabilità di intervalli	18
B-quantile	18
V.A. notevoli	19
Variabile casuale di Bernoulli	19
Variabile casuale binomiale	19
Esempi di V.A. binomiali	20
Variabile casuale geometrica	20
Proprietà assenza memoria	20
Var. di Poisson di parametro $\lambda > 0$	20
V.A. ipergeometriche	21
V.A. uniformi	21
V.A. esponenziali di parametro $\lambda > 0$	22
Trasformazione della V.A. con densità	22
Proposizione cambio di variabile	23
V.A. Gaussiane	23
N(0,1)	23
Funzione di ripartizione di N(0,1)	23
B-quantile di N(0,1)	23
Proprietà di N(0,1)	23
Variabile Gaussiana generale	24
Calcoli e standardizzazione	24
Proprietà di riproducibilità	24

STATISTICA DESCRITTIVA NELLA PROBABILITÀ

E[x] Valore atteso	26
Significato di E[x]	26
Quando E[x] è definito?	26
Proposizione	27
Proprietà del valore atteso	27
E[x ⁿ] Momento di ordine n di X	27
Disuguaglianza di Markov	27
Var(x)	27
$\sigma(x)$ Deviazione standard di X	27
Significato di var(x)	27
Proprietà di scaling	28
Disuguaglianza di Chebyshev	28
Corollario	28
Varianza e valore atteso delle V.A. notevoli	28

T ~ Geom(P)	28
T ~ Geom(P)	28
X ~ Poiss(λ)	29
X ~ IP(N,F,n)	29
X ~ U([a,b])	29
X ~ Exp(λ)	29
X ~ N(m, σ^2)	30
Variabile casuale doppia (bivariata)	30
Legge congiunta di X e Y	30
Distribuzioni marginali di una V.A. doppia	30
V.A. doppia discreta	31
Funziona di massa congiunta di (X,Y)	31
V.A. doppia con densità	32
V.A. indipendenti	32
Criterio di indipendenza per V.A. discrete	32
Stabilità della indipendenza per composizione	32
Proprietà di riproducibilità	32
Valore atteso del prodotto di V.A. indipendenti	33
Covarianza di 2 V.A. con momento secondo	33
Coefficiente di correlazione	33
Variabili scorrelate	33
Scorrelazione vs Indipendenza	33
Proprietà della covarianza bilineare e simmetrica	34
Varianza della somma	34
Teorema	34
(i.i.d) V.A. indipendenti ed equamente distribuite	34
Media aritmetica	34
Convergenza	34
LGN Legge dei grandi numeri (debole)	35
S ²	35
Proprietà	35
Convergenza in legge	35
(TCL/TLC) Teorema central del limite	35
Approssimazione normale della binomiale	36
Caso particolare	36
Regola empirica	36

STATISTICA INFERENZIALE

Statistica inferenziale	37
Introduzione	37
Statistica parametrica	37
Statistica campionaria	37
Stimatori di parametro β	37
Stimatore corretto (non distrorto)	38
Stimatore consistente	38
Efficienza di uno stimatore	38
Metodo della massima verosimiglianza	38
Funzione di verosimiglianza	38
(MLE) Stima massima della verosimiglianza	38
Intervalli di fiducia	39

STATISTICHE RIASSUNTIVE E GRAFICI

Statistica Descrittiva

Dati analizzati senza assunzioni esterne
 => scopo: organizzare dati, evidenziarne la struttura

Statistica Inferenziale

Studio dei dati tramite un modello probabilistico
 => scopo: usare il modello di tale stima per fare previsioni

Popolazione

Insieme di oggetti o fenomeni da studiare
 es. studenti università di Pisa → Popolazione reale
 tutti i possibili lanci di un dado → Popolazione ideale

Carattere

Caratteristica degli individui della popolazione, ottenuta con stesso procedimento
 es. altezza studenti

Modalità

Possibili valori che può assumere il carattere
 es. 180 cm

Campione

Sottoinsieme della popolazione
 es. 10 italiani

Dati

Esiti delle misure

Problema

Capire quando un campione è rappresentativo (problema complesso)

CARATTERI

Quantitativi

Se gli esiti delle misure (modalità) sono paragonabili fra loro
 es. 180 cm, 173 cm, 165 cm

Qualitativi

Se le modalità non sono paragonabili fra loro
 es. associare un numero a un nome e confrontarli:

Discreto

Assume una quantità finita (e piccola) di valori
 es. esiti dei lanci di un dado (1, 2, 3, 4, 5, 6)

Continuo

Assume una grande quantità di valori
 es. altezza degli italiani

FREQUENZA

Frequenza Assoluta

Numero di volte che una modalità compare nei dati

Frequenza Relativa

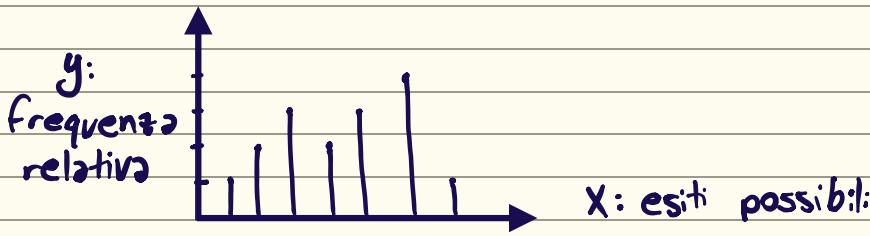
Frazione di volte che una modalità compare nei dati

es. 8 Lanci TCTTCTTT

Frequenza assoluta di T = 6

Frequenza relativa di T = 6/8 = 75%

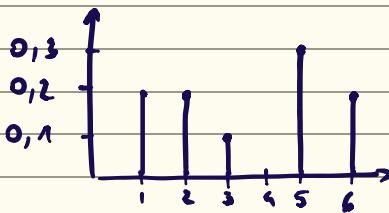
DIAGRAMMA A BARRE



Esempio di Esercizio

10 Lanci di dado, esito: 6, 2, 3, 1, 2, 5, 6, 5, 5, 1

- popolazione (ideale): tutti i possibili lanci del dado (∞)
- campione: i 10 lanci effettuati
- carattere: esito del lancio di un dado
- modalità: possibili valori che puo' prendere il carattere (1, 2, ..., 6)



INDICI

Indice
statistico

Quantitativo numerico che riassume qualche proprietà significativa di una distribuzione di dati
 \Rightarrow servono per effettuare sintesi dei dati

Di centralità

Indica dove si trova il centro di distribuzione dati

Media

Media aritmetica dei dati $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

Mediana

Il dato x_i tale che metà degli altri valori è $\leq x_i$ e l'altra metà è $\geq x_i$
 Per calcolarla:

1) Ordino i dati $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$

2) Divido il vettore

- Se (n) è dispari \Rightarrow mediana = $x_{(\frac{n+1}{2})}$

- Se (n) è pari \Rightarrow mediana = $\frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}$

es. $x = (8, 3, 5, 12, 10) \in \mathbb{R}^5$

$x_{(5)} = 8$

\Rightarrow media: $\bar{x} = (8+3+5+12+10)/5 = 38/5 = 7,6$

\Rightarrow mediana: ordino $x = (3, 5, 8, 10, 12) \Rightarrow n=5$ è dispari $\Rightarrow \frac{5+1}{2} = 3$

Moda

Dato più comune nel vettore $x = (x_1, \dots, x_n)$

Media Sfondata

Media effettuata sui dati che rimangono dopo aver tolto una certa frazione dei dati più grandi e una certa frazione di quelli più piccoli
es. si elimina il voto più alto e il più basso

Indici di variabilità

Misurano la dispersione dei dati intorno ai valori "tipici" individuati dagli indici di posizione (media, mediana, moda)

Varianza campionaria

Si calcola quando si ha un campione della popolazione

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

Varianza empirica

Si calcola quando si ha l'intera popolazione

$$\text{var}_e(x) = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 \xrightarrow{\text{se si conosce f. relativa}} \text{var}_e(x) = \sum_{i=1}^n x_i^2 \cdot \frac{f_i}{n} - \bar{x}^2$$

Rapporto tra $\text{var}(x)$ e $\text{var}_e(x)$

$$\text{var}(x) = \frac{n}{n-1} (\text{var}_e(x))$$

Esempio varianza campionaria ed empirica

10 Lanci di dado, esito: 6, 2, 3, 1, 2, 5, 6, 5, 5, 1

$$\begin{aligned} \text{var}_e(x) &= 1^2 \cdot \frac{2}{10} + 2^2 \cdot \frac{2}{10} + 3^2 \cdot \frac{1}{10} + 4^2 \cdot \frac{0}{10} + 5^2 \cdot \frac{3}{10} + 6^2 \cdot \frac{1}{10} - (3,6)^2 \\ &= 16,6 - 12,96 = 3,64 \end{aligned}$$

$$\text{var}(x) = \frac{10}{9} \cdot \text{var}_e(x) = \frac{10}{9} \cdot 3,64 \approx 4,04$$

Deviazione Standard

Chiamata anche Scarto quadratico medio, si divide in 2 tipi:

1) Campionario: $\sigma(x) = \sqrt{\text{var}(x)}$

2) Empirico: $\sigma_e(x) = \sqrt{\text{var}_e(x)}$

Sample Skewness

Indica quanto è asimmetrica la distribuzione

$$b = \frac{1}{\sigma^3} \cdot \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^3$$

• Esalta con segno positivo gli $x_i >> \bar{x}$

• Esalta con segno negativo gli $x_i << \bar{x}$

Funzione di ripartizione empirica (c.d.f.)

Restituisce il numero di dati presenti sull'asse

$$F_e: \mathbb{R} \rightarrow \mathbb{R} \implies F_e(t) = \frac{\#\{i \mid x_i \leq t\}}{n} \quad \text{f. relativa dei valori} \leq t$$

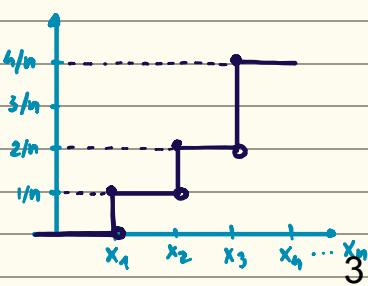
• Si parte da (x_1, \dots, x_n) e si ordinano i dati

• Per $t < x_{(1)}$, $F_e(t) = 0$

$$F_e(x_{(1)}) = \frac{\#\{i \mid x_i \leq x_{(1)}\}}{n} = \frac{1}{n}$$

$F_e(t) = F_e(x_{(1)}) \quad \forall x < t < x_{(2)}$ e poi fa un salto di $\frac{1}{n}$

$$\text{Se } x_{(3)} = x_{(4)} \Rightarrow F_e(x_{(3)}) = \frac{\#\{1, 2, 3, 4\}}{n} = \frac{4}{n}$$



Esempio Funzione di ripartizione empirica

Consideriamo gli esiti di 10 lanci di un dado: 2, 3, 1, 2, 5, 5, 1, 5

Ordiniamo i dati: 1, 1, 2, 2, 3, 5, 5, 5

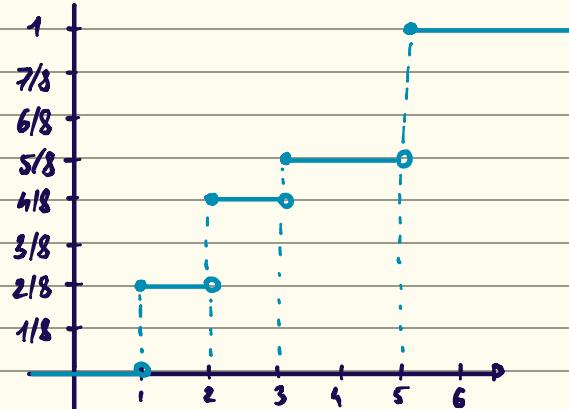
. Per $t < 1$ $F_E(t) = 0 \quad \left. \begin{array}{l} \\ \end{array} \right\}$ salto di
 $F_E(1) = 2/8 \quad \left. \begin{array}{l} \\ \end{array} \right\} 1/8 \cdot 2$

. Per $1 \leq t < 2$ $F_E(t) = 2/8 \quad \left. \begin{array}{l} \\ \end{array} \right\}$ salto di
 $F_E(2) = 4/8 \quad \left. \begin{array}{l} \\ \end{array} \right\} 1/8 \cdot 2$

. Per $2 \leq t < 3$ $F_E(t) = 4/8 \quad \left. \begin{array}{l} \\ \end{array} \right\}$ salto di
 $F_E(3) = 5/8 \quad \left. \begin{array}{l} \\ \end{array} \right\} 1/8$

. Per $3 \leq t < 5$ $F_E(t) = 5/8 \quad \left. \begin{array}{l} \\ \end{array} \right\}$ salto di
 $\downarrow \quad F_E(5) = 8/8 = 1 \quad \left. \begin{array}{l} \\ \end{array} \right\} 1/8 \cdot 3$
 Salta il 4 perché non c'è

. Per $t \geq 5$ $F_E(t) = 1$



K-simo percentile
o
B-quantile

Dato K numero naturale con $0 < k < 100$, $B = \frac{k}{100}$
allora il B-quantile è il dato x_i tale che:

- almeno il $k\%$ dei dati è $\leq x_i$
- almeno il $(1-k)\%$ dei dati è $\geq x_i$

In caso due dati soddisfino questa condizione, si prende la media.
La mediana è quindi il 50simo percentile ($\Rightarrow \frac{50}{100} = \frac{1}{2}$ quantile)

Notazione
Percentili e
Quantili:

$$K = 25 \Rightarrow 25^{\circ} \text{ percentile} \leftrightarrow B = 25/100 \rightarrow \text{I Quartile}$$

$$K = 50 \Rightarrow 50^{\circ} \text{ percentile} \leftrightarrow B = 50/100 \rightarrow \text{mediana} \rightarrow \text{II Quartile}$$

$$K = 75 \Rightarrow 75^{\circ} \text{ percentile} \leftrightarrow B = 75/100 \rightarrow \text{III Quartile}$$

Calcolare B-quantile

$B \cdot n$ ($n = \text{numero dati}$) può essere

- NON INTERO: arrotondando al numero successivo $\Rightarrow x_{BN}$
- INTERO: prendo la media tra x_{BN} e x_{BN+1}

Esempio B-quantili:

Considero la tabella dei voti degli esami di 14 studenti:

A	25	Calcola I, II e III quartile			$n = 14$
B	23				
C	25				
D	25				
E	28	20	22	22	
F	28	23	25	25	
G	26	25	25	25	
H	25	25	25	25	
I	26				
L	22				
M	2				
N	22				
O	20				
P	27				

$$\text{I quartile: } B \cdot n = \frac{25}{100} \cdot 14 = \frac{350}{100} = [3, 5] = 4 \Rightarrow x_4 = 23$$

$$\text{II quartile: } \frac{50}{100} \cdot 14 = 7 \Rightarrow \frac{x_7 + x_8}{2} = 25$$

$$\text{III quartile: } \frac{75}{100} \cdot 14 = [10, 50] = 11 \Rightarrow x_{11} = 26$$

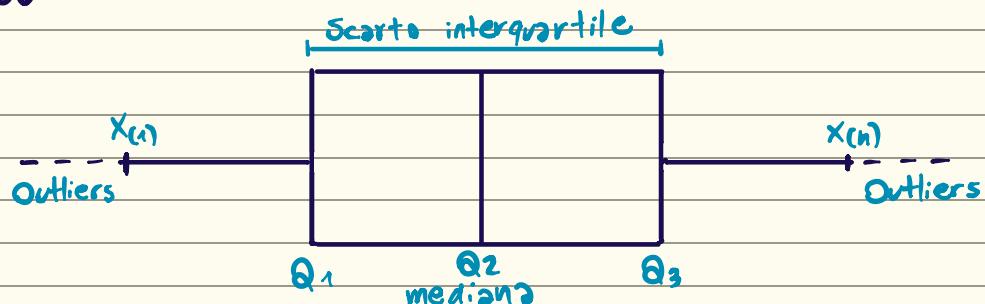
Box Plot

Grafico che fornisce informazioni su: posizione, variabilità e forme di una distribuzione dati

- Una linea in corrispondenza della mediana, rappresenta centro della distribuzione
- Box che va dal I al III quantile, la cui altezza indica la variabilità
- Segmenti che indicano gli estremi dei dati

Valore Anomalo (Outlier)

Valore che differisce in modo significativo dalla grande maggioranza dei dati



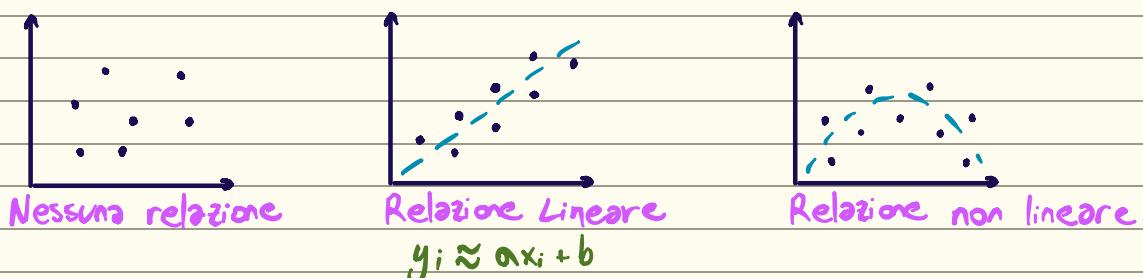
DATI MULTIVARIATI

K-uple di vettori

Capita di avere più dati misurati per lo stesso campione. In questo caso i dati sono K-uple di vettori: $(x_i, y_i, z_i, \dots) \in \mathbb{R}^{k \times n}$
es. (x_i, y_i) = altezza e peso iscritti: informatica

Diagramma a dispersione

Rappresenta con un punto ciascuna coppia di dati



Problema della regressione

Studio dei dati mirato a misurare la dispersione, tramite:

- Componente qualitativa: quale curva approssima meglio i dati (x_i, y_i)
- Componente quantitativa: quanto è buona l'approssimazione con tale curva

Covarianza

Di quanto si distaccano (differiscono) le coppie di dati

Covarianza campionaria

$$\text{cov}(x, y) = \sum_{i=1}^n \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

$$\text{o ss: } \text{cov}(x, y) = \text{cov}(y, x)$$

Covarianza empirica

$$\text{cov}_e(x, y) = \sum_{i=1}^n \frac{(x_i - \bar{x})(y_i - \bar{y})}{n}$$

Relazione tra $\text{cov}(x, y)$ e $\text{cov}_e(x, y)$

$$\text{cov}(x, y) = \frac{n}{n-1} \cdot \text{cov}_e(x, y)$$

$$\text{cov}(x, y) \cdot (n-1) = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = n \cdot \text{cov}_e(x, y) \Rightarrow \text{cov}(x, y) = \frac{n}{n-1} \cdot \text{cov}_e(x, y)$$

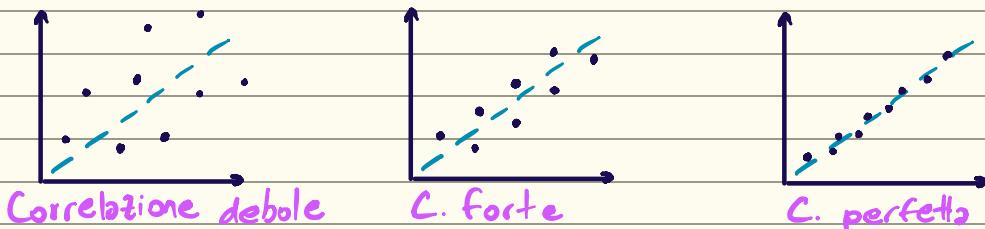
Coefficiente di correlazione

Si può calcolare tramite $r(x,y) \doteq \frac{\text{cov}(x,y)}{\sigma(x) \cdot \sigma(y)}$

OSS: coefficiente non cambia sostituendo con $\text{cov}_e(\dots)$ e $\sigma_e(\dots)$

La correlazione può essere:

- **Debole**: r vicino a 0
- **Forte**: r vicino a -1, oppure vicino a 1
- **Perfecto**: $r=1$ o $r=-1$



Funzione retta

La funzione equivalente alla retta è $y = q + mx$

- I punti della retta $y = q + mx$ verificano perfettamente l'equazione.
- I punti della distribuzione, di coordinate (x_i, y_i) , non la verificano perfettamente a priori, perché non perfettamente allineati. Ci sarà uno scarto tra y_i e il punto della retta corrispondente a x_i , ovvero $q + mx_i$.
- L'idea è quindi scegliere q e m in modo tale da minimizzare gli scarti:
 \Rightarrow cerco $\min_{q, m \in \mathbb{R}} \sum_{i=1}^n (y_i - q - mx_i)^2$

Teorema

Se $\sigma(x) \neq 0$ e $\sigma(y) \neq 0$, allora la $\sum_{i=1}^n (y_i - q - mx_i)^2$ ha uno e un solo minimo, che si ottiene prendendo:

$$m^* = \frac{\text{cov}(x,y)}{\text{var}(x)} \quad \text{e} \quad q^* = \bar{y} - m^* \cdot \bar{x}$$

Inoltre $\min_{q, m \in \mathbb{R}} \sum_{i=1}^n (y_i - q - mx_i)^2$ è uguale a $\sum_{i=1}^n (y_i - \bar{y})^2 \cdot [1 - 2(\bar{x}, \bar{y})^2]$

Retta di regressione

La retta $y = m^*x + q^*$ con m^* e q^* come sopra si dice retta di regressione tra x e y , per cui:

- è la retta che meglio approssima i dati (x_i, y_i) $i=1, \dots, n$
- più il coefficiente di correlazione $r(x,y)$ è vicino a 1, migliore è l'approssimazione
- se $r(x,y) = \pm 1$ allora $m = 0$, ovvero tutti i punti (x_i, y_i) sono sulla retta
- il segno del coefficiente angolare coincide con quello di $r(x,y)$

PROBABILITÀ E INDEPENDENZA

Esperimento aleatorio

Esperimento di cui non si conosce con esattezza il risultato
es. lancio di una moneta, estrazione biglia, ...

Esito

Possibile risultato di un esperimento aleatorio
es. T, biglia blu, ...

Spazio campionario

Rappresentalo con Ω , è l'insieme di tutti i possibili esiti
es. $\Omega = \{T, C\}$, $\Omega = \{\text{bocciato}, 18, 19, \dots, 30\}$

Evento

Sottoinsieme dello spazio campionario
es. affermazione $A = \{2, 4, 6\}$ (esce un numero pari)

Evento elementare

Sottoinsieme dello spazio campionario costituito da un singolo elemento
es. nel lancio di un dado ci sono 6 eventi elementari: $E_1 = \{1\}, \dots, E_6 = \{6\}$

Evento certo

Sottoinsieme improprio di Ω , quindi Ω stesso

Evento impossibile

È \emptyset . Visto che uno degli esiti si verifica sicuramente, \emptyset è impossibile

TEORIA DEGLI INSIEMI

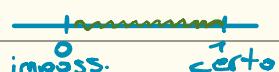
$$\rightarrow A = \{2, 4, 6\} \quad B = \{4, 5, 6\}$$

Operazione insiem.	Operazione Logica	Esempio
$A \cup B$	accade A oppure B	$A \cup B = \{2, 4, 5, 6\}$
$A \cap B$	accade A e contemporan. B	$A \cap B = \{4, 6\}$
A^c	non accade A	$A^c = \{1, 3, 5\}$
$A \setminus B = A \cap B^c$	accade A ma non B	$A \setminus B = \{2\}$
Relazione Insiem.	Relazione Logica	Esempio
$A \subseteq B$	se accade A \Rightarrow accade B	$\{5\} \subset \{5, 6\}$
$A \cap B = \emptyset$	non possono accadere contemporan. A e B	$C = \text{esce un numero} \leq 2$ $C \cap B = \emptyset$

PROBABILITÀ

Probabilità

Misura il grado di fiducia di A su una scala $P(A) \in [0, 1]$ dove
lo 0 indica l'impossibile e l'1 la certezza



Definizione classica (Laplace)

$$P(A) = \frac{\text{caso favorevole di } A}{\text{caso possibile}} = \frac{\# A}{\# \Omega}$$

Pro e Contro:

- Adatta per misurazioni sulla popolazione reale
- Non adatta a popolazione ideale
- Gli eventi devono essere equiprobabili e finiti

$$\text{es. } E = \text{"in un solo lancio di dado si presentano 3 o 4"} \quad P(E) = \frac{\# E}{\# \Omega} = \frac{2}{6} = \frac{1}{3}$$

Definizione Frequentista

$P(A) = \lim_{n \rightarrow +\infty} (\text{frequenza relativa di } A \text{ in } n \text{ prove ripetute})$

Presi quindi:

- w_1, \dots, w_n eventi elementari: incompatibili: non equiprobabili:
- n_1, \dots, n_k frequenza assoluta

$$P(w_i) = \lim_{n \rightarrow +\infty} \frac{n_i}{n}$$

Pro e contro:

- Adatta a popolazioni: ideal
- Non sempre è possibile effettuare un numero enorme di prove
- Non è detto che esista sempre il limite
es. 1000 lanci di una moneta, esce testa 529 volte $\Rightarrow F_{\text{rel}} = \frac{529}{1000} = 0,529$
altri 1000 lanci, viene testa 493 volte $\Rightarrow \frac{493}{1000} = 0,493$
Al tendere all'infinito delle prove dovrà raggiungere il 50%

Definizione assiomatica

Insieme di assiomi che definiscono formalmente una misura di probabilità tramite oggetti e postulati.

Oggetti: esperimento aleatorio, esito, evento, spazio campionario

Postulati:

$$1) 0 \leq P(A) \leq 1 \quad \forall A \subseteq \Omega$$

$$2) P(\Omega) = 1$$

$$\rightarrow (A_i \cap A_j = \emptyset \quad \forall i \neq j)$$

$$3) \text{ se } A_1, \dots, A_n \text{ è una successione di eventi incompatibili:}$$

$$\text{allora } P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i) \text{ additività}$$

$$\text{e } P(\bigcup_{i=1}^{\infty} A_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(A_i) \text{ se la sommatoria è infinita}$$

Oss: la probabilità frequentista soddisfa gli assiomi:

Spazio di probabilità

La coppia (Ω, P) è lo spazio di probabilità

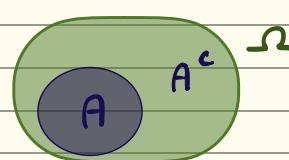
Sia (Ω, P) uno spazio di probabilità $P: P(\Omega) \rightarrow \mathbb{R}$, allora:

$$1) P(A^c) = 1 - P(A)$$

dimostrazione:

$$P(A \cup A^c) = P(\Omega) = 1$$

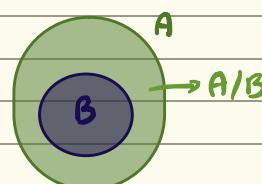
$$\stackrel{!}{=} P(A) + P(A^c)$$



$$2) B \subseteq A \Rightarrow P(A \setminus B) = P(A) - P(B)$$

dimostrazione:

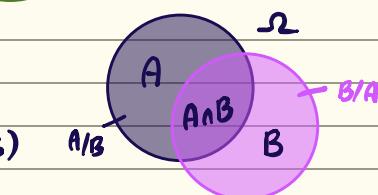
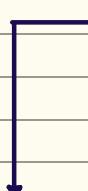
$$A = B \cup A \setminus B \quad P(A) = P(B \cup A \setminus B) = P(B) + P(A \setminus B)$$



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

dimostrazione:

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



Formula di inclusione-esclusione

Potrei non avere due insiemi ma n che si intersecano fra loro:

$$P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k})$$

ESERCIZI PROBABILITÀ

- 1) A = "cliente chiede uno sconto"
 B = "cliente chiede rateizzato"
 $A \cap B$ = "cliente chiede sconto e rateizzato"
 $P(A) = 0,5$ $P(B) = 0,37$ $P(A \cap B) = 0,2$

Calcolare la probabilità che il cliente chieda il pagamento rateizzato oppure lo sconto e quella che non chieda né sconto né rate.

- $A \cup B$ = "cliente chiede sconto o rateizzato"
 $(A \cup B)^c$ = "non chiede né sconto né rate"

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,5 + 0,37 - 0,2 = 0,67$$

$$P(A \cup B)^c = 1 - P(A \cup B) = 1 - 0,67 = 0,33$$

Modello Uniforme

(Ω, P) tale che Ω è finito e ogni esito $\omega \in \Omega$ è equiprobabile
 es. lancio di un dado non truccato

Se (Ω, P) è un modello uniforme $\Rightarrow P(A) = \frac{\#\Omega}{\#\Omega}$
 Inoltre, con un insieme $\{1, \dots, n\}$:

- il $\#\Omega$ di sequenze ordinate con ripetizione di k oggetti è n^k
- il $\#\Omega$ di sequenze ordinate senza ripetizione di k oggetti è $\frac{n!}{(n-k)!}$
- il $\#\Omega$ di sottoinsiemi di $\{1, \dots, n\}$ formato da k oggetti è $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Probabilità condizionata

Quando si è a conoscenza della realizzazione di un evento, cambia la valutazione di probabilità di ogni altro evento

es. So che il numero uscito è pari e mi chiedo la probabilità che sia uscito il numero 6.
 Se non sapessi nulla sarebbe $\frac{1}{6}$ ma, sapendo che è uscito un numero pari, cambio il denominatore $\Rightarrow \frac{P(A \cap B)}{P(B)} = \frac{1}{3}$

Probabilità condizionata di A dato B

Dato (Ω, P) , B evento non trascurabile ($P(B) \neq 0$) e A evento, si calcola:

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

probabilità di A sapendo che si è verificato B

Regola dei prodotti

Se A e B sono non trascurabili, allora $P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$

Dimostrazione

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \text{e} \quad P(A|B) = \frac{P(A \cap B)}{P(B)} \implies P(A) \cdot P(B|A) = P(A \cap B) = P(B) \cdot P(A|B)$$

Condizionamento ripetuto

Siano A_1, \dots, A_n eventi in Ω se $A_1 \cap \dots \cap A_n$ non trascurabile, allora:

$$P(A_1 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \dots \cdot P(A_n|A_1 \cap \dots \cap A_{n-1})$$

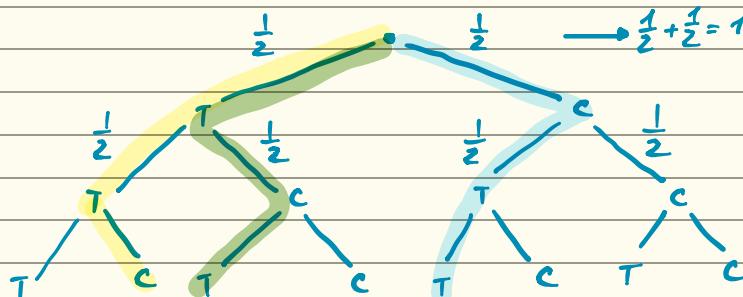
Rappresentazione ad albero

Diagramma in cui:

- al nodo (stadio) ci sono tanti rami quanti possibili esiti
- ad ogni ramo è associata una probabilità corrispondente
- la somma delle probabilità dei rami per colonna deve essere = 1

es. ottenere 2 teste in 3 lanci di monete consecutivi

$$\Omega = \{T, C\} \quad P(T) = P(C) = \frac{1}{2}$$



$$TTT = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

$$TTC = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

$$CCT = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

$$\frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$$

Regola della catena

$P(\text{ramo}) = \text{prodotto delle probabilità dei sottorami:}$

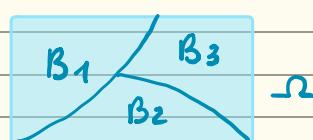
Regola della fattorizzazione

Dato un esperimento con spazio campionario Ω , è il dato di B_1, \dots, B_n eventi tali che:

- $B_i \cap B_j = \emptyset \quad \forall i \neq j$
- $\bigcup B_i = \Omega$
- ciascun B_i è non trascurabile ($P(B_i) \neq 0$)

Traduzione: quando effettua un esperimento avviene uno e uno solo dei B_i :

$$\text{es. } B_1 = \{<2\} \quad B_2 = \{2 < x < 4\} \quad B_3 = \{>4\}$$



Formula di fattorizzazione

Sia B_1, \dots, B_n un sistema di alternative e sia $A \subseteq \Omega$ un evento, allora:

$$P(A) = \sum_{i=1}^n P(A \cap B_i) = \sum P(A|B_i) \cdot P(B_i)$$

regola dei prodotti:

$$\text{es. Urna 1} = \begin{cases} 5 \text{ palline rosse} \\ 5 \text{ palline blu} \end{cases}$$

$$\text{Urna 2} = \begin{cases} 8 \text{ palline rosse} \\ 2 \text{ palline blu} \end{cases}$$

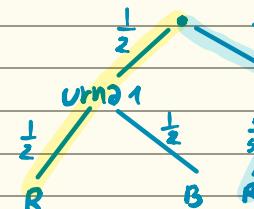
Scegliere a caso una pallina da una delle due urne, qual è la probabilità che la pallina scelta sia rossa?

$$\Omega = \{(\text{*urna, colore pallina}) | \text{urna} \in \{1, 2\}, \text{colore} \in \{R, B\}\}$$

$$P(\text{urna 1}) = P(\text{urna 2}) = \frac{1}{2}$$

$$P(\text{pallina rossa | urna 1}) = \frac{5}{10} = \frac{1}{2}$$

$$P(\text{pallina rossa | urna 2}) = \frac{8}{10} = \frac{4}{5}$$



$$P(U_1 \cup R) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$$P(U_2 \cup R) = \frac{1}{2} \cdot \frac{4}{5} = \frac{2}{5}$$

$$P(R) = \frac{1}{4} + \frac{2}{5} = \frac{13}{20}$$

Formula di Bayes

Siano A e B eventi non trascurabili, allora

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

In generale, se B_1, \dots, B_n è un sistema di alternative, allora:

$$P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{\sum_{j=1}^n P(A|B_j) \cdot P(B_j)} \quad \forall i = 1, \dots, n$$

Oss: utile per invertire il condizionamento: da $P(A|B)$ si ottiene $P(B|A)$

es. urna1 = $\begin{cases} 5R \\ 5B \end{cases}$ urna2 = $\begin{cases} 8R \\ 2B \end{cases}$

Suppongo di estrarre una pallina rossa, qual è la probabilità che l'urna scelta sia la urna1?

A = "estrarre pallina rossa"
B = "estrarre urna1"

$$P(B|A) = P(1|R) = \frac{P(A|R) \cdot P(R)}{P(R)} = \frac{\frac{1}{2} \cdot \frac{1}{2}}{\frac{13}{20}} = \frac{5}{13}$$

EVENTI INDEPENDENTI

Eventi
indipendenti

Due eventi A e B si dicono indipendenti se
 $P(A \cap B) = P(A) \cdot P(B)$

Oss: Se due eventi A e B sono indipendenti, allora $P(A|B) = P(A)$ e $P(B|A) = P(B)$

Se A e B sono indipendenti, allora sono indipendenti anche le coppie A^c e B , A e B^c , A^c e B^c

1) Se $P(A) = 0$ oppure $P(A) = 1 \Rightarrow A$ è indipendente da qualsiasi altro evento

dimostrazione:

Supponiamo che $P(A) = 0$ e che B sia un altro evento qualsiasi.

Allora $P(A \cap B) = 0 = 0 \cdot P(B) = P(A) \cdot P(B)$ $\forall B$ per definizione è indipendente

Se $P(A) = 1 \Rightarrow P(A^c) = 0 \Rightarrow A^c$ è indipendente da qualsiasi altro evento

Lo stesso per $(A^c)^c$ ovvero per A

2) Due eventi disgiunti non possono essere indipendenti, a meno che uno dei due non sia trascurabile

dimostrazione

Se A e B sono disgiunti: $A \cap B = \emptyset \Rightarrow P(A \cap B) = 0$

es. Mazzo da 40 carte con numeri $1, \dots, 10$ e semi: $\{C, D, P, F\}$

A = "estraggo un asso (1)" B = "estraggo una carta D (denari)"

A e B sono indipendenti?

$$P(A) = \frac{4}{40} = \frac{1}{10} \quad P(B) = \frac{10}{40} = \frac{1}{4}$$

$$P(A \cap B) = \text{"esce l'asso di denari"} = \frac{1}{40} = \frac{1}{10} \cdot \frac{1}{4} \Rightarrow \text{Sì, sono indipendenti!}$$

Indipendenza
di 3 eventi

Supponiamo di avere 3 eventi A, B e C . Si dicono indipendenti se:

- sono indipendenti 2 a 2 (i.e. $P(A \cap B) = P(A) \cdot P(B)$, $P(A \cap C) = P(A) \cdot P(C)$ e $P(B \cap C) = P(B) \cdot P(C)$)
- $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$

Oss: è necessario che siano vere entrambe le condizioni!

Indipendenza
di n eventi

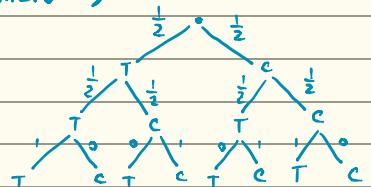
Dati A_1, \dots, A_n eventi, si dicono indipendenti se \forall intero k con $2 \leq k \leq n$ e \forall scelta di interi $1 \leq i_1 \leq i_2 \leq \dots i_k \leq n$ si dice che:

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdot \dots \cdot P(A_{i_k})$$

Traduzione: sono tutti indipendenti: 2 a 2, 3 a 3, 4 a 4, ...

es. 3 lanci di monete, il terzo è truccato in modo che sia T se i primi 2 sono concordi
C altrimenti

$A: \text{"T al 1° lancio"}$ } Verificare che A, B, C sono 2 a 2 indipendenti
 $B: \text{"T al 2° lancio"}$ } ma non sono globalmente indipendenti
 $C: \text{"T al 3° lancio"}$ }



$$P(A \cap B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = P(A) \cdot P(B) \text{ ok! } \text{IND}$$

$$P(A \cap C) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = P(A) \cdot P(C) \text{ ok! } \text{2 a 2}$$

$$P(B \cap C) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = P(B) \cdot P(C) \text{ ok!}$$

$$P(A \cap B \cap C) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8} \neq \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \text{ No! non globale! } 12$$

Caso importante
su eventi indipen.

Prove ripetute di un esperimento svolte nelle stesse condizioni
es. n estrazioni di biglie, che vengono re-inserite una volta estratte (eventi indipendenti)

Schema di:
Bernoulli

Consideriamo n prove ripetute di un esperimento, in cui gli esiti sono successo(1)
e insuccesso(0). Sia p la probabilità di successo nella singola prova.

$$\Rightarrow \Omega = \{(a_1, \dots, a_n) \mid a_i \in \{0, 1\}\} = \{0, 1\}^n$$

Allora, visto che gli eventi sono indipendenti, la probabilità è:

$$P\{(a_1, \dots, a_n)\} = P(a_1) \cdot \dots \cdot P(a_n) = p^{\text{#successi}} \cdot (1-p)^{\text{#insuccessi}} = p^{\sum a_i} \cdot (1-p)^{n - \sum a_i}$$

es. Lancio 10 volte una moneta. È più probabile che ottenga 10 volte croce oppure 4 volte testa e 6 volte croce?

P(T) = P(C) = 1/2 successo(1) = croce insuccesso(0) = testa

$$P(C, \dots, C) = P(\underbrace{1, \dots, 1}_{10 \text{ volte}}) = \left(\frac{1}{2}\right)^{10}$$

$$P(1111110000) = \left(\frac{1}{2}\right)^6 \cdot \left(\frac{1}{2}\right)^4 = \left(\frac{1}{2}\right)^{10} \text{ No! L'evento che ci interessa non è } 1111110000 \text{ ma è l'unione di tutti gli eventi con quel risultato}$$

$$P(6 \text{ croci e } 4 \text{ testa}) = \sum_{i=1}^{210} \left(\frac{1}{2}\right)^{10} = 210 \cdot \left(\frac{1}{2}\right)^{10}$$

$$\text{tutte le combinazioni di } 6C+4T \text{ cioè } \binom{10}{4} = \frac{10!}{4!6!} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6!}{4! \cdot 6!} = 210 = \binom{10}{6}$$

Quindi la probabilità di 6C+4T è molto più alta rispetto a 10T

ESERCIZI NELLA LEZIONE 7

VARIABILI ALEATORIE

Aleatorio \Rightarrow casuale

Variabile aleatoria

Funzione X che associa ad ogni elemento dello spazio campionario un numero reale $X: \Omega \rightarrow \mathbb{R}$

es. esperimento = n lanci di monete $\Omega = \{\text{T, C}\}^n = \{(w_1, \dots, w_n) | w_i \in \{0, 1\}\}$

$X: "N \text{ teste in } n \text{ lanci}" X: \Omega \rightarrow \mathbb{R} \Rightarrow X(w_1, \dots, w_n) = \sum_{i=1}^n w_i$

Aggiungo un'altra variabile aleatoria per il numero delle croci $y = "N \text{ croci in } n \text{ lanci}"$
 $y: \Omega \rightarrow \mathbb{R} \Rightarrow (w_1, \dots, w_n) = n - \sum w_i \Rightarrow y = n - X$

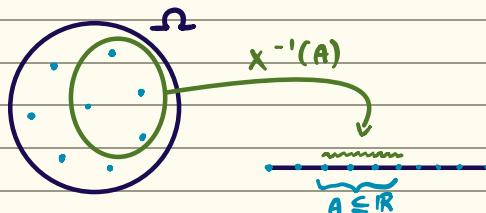
$Z = "N \text{ coppie teste consecutive}" \text{ idea } z((1, 1, 0, 0, 1)) = 1$

$Z: \Omega \rightarrow \mathbb{R} \quad (w_1, \dots, w_n) = w_1 \cdot w_2 + w_2 \cdot w_3 + \dots + w_{n-1} \cdot w_n = \sum_{i=1}^{n-1} w_i \cdot w_{i+1}$

Legge di probab.
di una variabile
aleatoria

Sia $X: \Omega \rightarrow \mathbb{R}$ una variabile aleatoria, definiamo legge (o distribuzione) di X la funzione $P_X: \mathcal{P}(\mathbb{R}) \rightarrow \mathbb{R}$ data da:

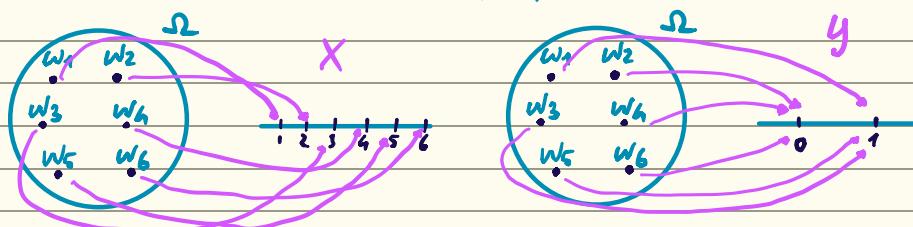
$P_X(A) = P(X \in A) \quad \forall A \subseteq \mathbb{R}$ dove $\{X \in A\} \doteq X^{-1}(A) = \{w \in \Omega | X(w) \in A\}$



es. esperimento = lancio di un dado $\Omega = \{w_1, w_2, \dots, w_6\}$ dove $w_i = "è uscito il numero i"$

$X: \Omega \rightarrow \mathbb{R}$ dove X = variabile che esprime N puntini associati al lancio del dado

$y: \Omega \rightarrow \mathbb{R}$ dove $w_i = 1$ se i è dispari, $w_i = 0$ se i è pari



Dall'esempio voglio trovare la legge di X e y
 Per esempio, $P_X(3) = (X^{-1}(3)) = P(w_3) = 1/6$

es. $A = \{3, 4\} \subseteq \mathbb{R} \quad P(X^{-1}(\{3, 4\})) = P(w_3 \cup w_4) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$

es. $P_y(\{0\}) \doteq P(y^{-1}(0)) = P(w_2 \cup w_4 \cup w_6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$

Proposizione

X^{-1} commuta con le operazioni insiemistiche

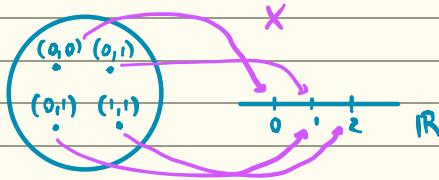
es. $X^{-1}(A \cup B) = X^{-1}(A) \cup X^{-1}(B)$ $X^{-1}(A^c) = (X^{-1}(A))^c$

$P_X: \mathcal{P}(\mathbb{R}) \rightarrow \mathbb{R}$ è una probabilità su \mathbb{R}



es. 2 lanci di moneta equilibrata

$$\Omega = \{C, T\}^2 = \{(w_1, w_2) \mid w_1, w_2 \in \{C, T\}\} = \{0, 1\}^2 \rightarrow 0 = \text{esce croce}, 1 = \text{esce testa}$$



$X = \text{"* di teste in 2 lanci"}$

$$P(0) = P(1) = \frac{1}{2}$$

$$P_X(0) = P(X^{-1}(0)) = P(\{(0,0)\}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$$P_X(1) = P(\{(0,1)\} \cup \{(1,0)\}) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$$

$$P_X(2) = P(\{(1,1)\}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

V.A.

equidistribuite

X e Y si dicono equidistribuite se hanno la stessa legge (distribuzione) di probabilità ($P_X = P_Y$)

es. 2 lanci di moneta equilibrata $\Omega = \{C, T\}^2 = \{0, 1\}^2$

$$X = \text{"*+teste"} = \sum_{i=1}^2 w_i = w_1 + w_2$$

$$P_X = P(\text{IR}) \rightarrow \text{IR} \quad A \mapsto P(X^{-1}(A))$$

$$Y = \text{"*croci"} = n - \sum_{i=1}^2 w_i = 2 - X$$

$$P_Y = P(\text{IR}) \rightarrow \text{IR} \quad A \mapsto P(Y^{-1}(A))$$

$P_X = P_Y$ quindi le v.a. sono equidistribuite (ma non uguali)

Notazione

Lettere maiuscole (X, Y, T, \dots) per le variabili aleatorie, lettere minuscole (a, b, x, z, \dots) per i valori assunti dalle variabili aleatorie

V.A. discrete

Una variabile aleatoria X si dice discreta se assume un numero finito oppure numerabile di valori (modalità) a_1, \dots, a_n, \dots

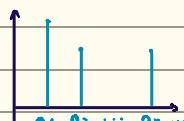
es.

Funzione di massa

Supponiamo di avere una variabile aleatoria discreta $X: \Omega \rightarrow \text{IR}$, la funzione di massa (o densità discreta) di X è $P_X: \{a_1, \dots, a_n, \dots\} \rightarrow \text{IR}$ definita come $P_X(a_i) = P(X=a_i) = P\{X^{-1}(a_i)\} = P(X^{-1}(a_i))$

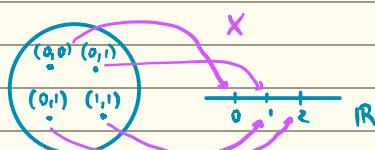
È poi possibile estendere P_X per il numero definito solo sull'insieme discreto dei valori di X a tutto IR , ponendo $P_X(a) = 0 \forall a \notin \{a_1, \dots\}$

La funzione di massa di una v.a. discreta si rappresenta bene con un diagramma a barre:



es. esperimento = 2 lanci di moneta equilibrata

$$\Omega = \{C, T\}^2 = \{0, 1\}^2 \quad X = \text{* teste in 2 lanci}$$



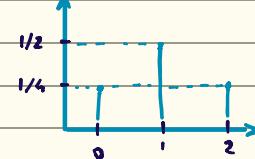
$X: \Omega \rightarrow \text{IR}$ prende solo i valori: 0, 1, 2

La sua funzione di massa è: $P_X: \{0, 1, 2\} \rightarrow \text{IR}$

$$0 \mapsto P(X^{-1}(0)) = 1/4$$

$$1 \mapsto P(X^{-1}(1)) = P(\{(0,1)\} \cup \{(1,0)\}) = 1/2$$

$$2 \mapsto P(X^{-1}(2)) = 1/4$$



Calcolo delle probabilità di X usando P_X

Preso X v.a. discreta con funzione di massa P_X , allora

$$P(X \in A) [= P(X^{-1}(A))] = P(X = a_1 \cup X = a_2 \cup \dots) = \sum_{a_i \in A} P(X = a_i) = \sum_{a_i \in A} P_X(a_i)$$

es. Qual è la probabilità di avere almeno una testa in 2 lanci?

$$A = \{1, 2\} \Rightarrow P(X \in \{1, 2\}) = P(X=1) + P(X=2) = 1/4 + 1/2 = 3/4$$

Qual è la probabilità di avere almeno 2 teste in 4 lanci?

$$\Omega = \{0, 1, 2, 3, 4\} \quad X = \# \text{ teste} \quad \Omega \rightarrow \begin{array}{ccccccccc} & & & & & & & & \\ & & & & & & & & \\ 0 & 1 & 2 & 3 & 4 & & & & \end{array}$$

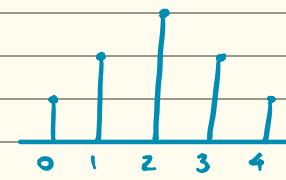
$$P_X(0) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{16}$$

$$P_X(1) = \text{TCCT CCTC CTCT CCTC} = \frac{4}{16} = \frac{1}{4}$$

$$P_X(2) = \text{TTCC TCTC TCTT CTCT CCTC CCTT} = \frac{6}{16} = \frac{3}{8}$$

$$P_X(3) = \text{TTTC TTCT TCTT CCTT} = \frac{4}{16} = \frac{1}{4}$$

$$P_X(4) = \frac{1}{16}$$



$$P(\geq 2 \text{ teste in 4 lanci}) = P_X(2) + P_X(3) + P_X(4) = \frac{3}{8} + \frac{1}{4} + \frac{1}{16} = \frac{11}{16}$$

Proprietà

Sia X una v.a. discreta con funzione di massa P_X , allora:

$$1) P_X(a_i) \geq 0 \quad \forall i$$

$$2) \sum_i P_X(a_i) = 1$$

dimostrazione:

$$1) P_X(a_i) = P(X = a_i) = P(X^{-1}(a_i)) \geq 0 \quad P \text{ è una probabilità}$$

$$2) \sum_i P_X(a_i) = \sum_i P(X = a_i) = P(\bigcup X^{-1}(a_i)) = P(\Omega) = 1$$

Sopra

Proposizione

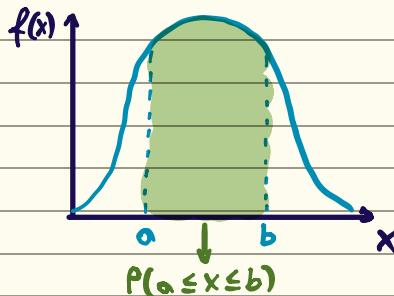
Se $q: \{a_1, a_2, \dots\} \rightarrow \mathbb{R}$ è una funzione che soddisfa le proprietà 1 e 2 allora \exists una v.a. X di cui q è la funzione di massa.
es. ($q = P_X$ per qualche X v.a. discreta)

V.A. assolutamente continua

Sia X una variabile casuale che assume tutti i valori di un intervallo (limitato o illimitato), allora diciamo che X è una v.a. assolutamente continua con densità f se ad X è associata una funzione $f: \mathbb{R} \rightarrow \mathbb{R}$ tale che $P\{X \in A\} = P_X(A) = \int_A f(x) dx \quad \forall A \in \mathbb{R}$

Ovvero: $A \subseteq \mathbb{R}$ sia un intervallo $A = [a, b]$, allora

$$P(X \in A) = P(a \leq X \leq b) = \int_a^b f(x) dx \quad (\text{area sottesa del grafico di } f \text{ sull'intervallo } [a, b])$$



Proprietà della funzione densità

Sia X una v.a. con funzione di densità f , allora:

$$1) f(x) \geq 0 \quad \forall x$$

$$2) \text{la misura dell'area sotto al grafico di tutte } f \text{ è } = 1$$

Calcoli al
variare di:
 $A \subseteq \mathbb{R}$

- $A = [a, b] \circ (a, b) \Rightarrow P(a \leq x \leq b) = P(a < x < b) = \int_a^b f(x) dx$
- $A = (-\infty, b] \circ (-\infty, b) \Rightarrow P(x \leq b) = P(x < b) = \int_{-\infty}^b f(x) dx = \lim_{a \rightarrow -\infty} \int_a^b f(x) dx$
- $A = [a, +\infty) \circ (a, +\infty) \Rightarrow P(x \geq a) = P(x > a) = \int_a^{+\infty} f(x) dx = \lim_{b \rightarrow \infty} \int_a^b f(x) dx$
- $A = \bigcup_{i=1}^n (a_i, b_i)$ unione di intervalli disgiunti $\Rightarrow P(\bigcup_{i=1}^n (a_i, b_i)) = \sum_{i=1}^n \int_{a_i}^{b_i} f(x) dx$

es. asta di un certo materiale che si puo' rompere in ogni punto con la stessa probabilita'. L'asta ha lunghezza 1, cerco la probabilita' che l'asta si rompa in suo punto.

Definisco la v.a. X con densita' f



$$f(x) = \begin{cases} 1 & \text{se } x \in (0, 1) \\ 0 & \text{se } x \notin (0, 1) \end{cases} = 1_{(0,1)}$$

Se invece prendessi $a=0$ e $b=\frac{1}{4}$ $P(0 \leq x \leq \frac{1}{4}) = \int_0^{\frac{1}{4}} 1 dx = \frac{1}{4}$

Se invece prendessi $a=3/4$ e $b=5$ $P(\frac{3}{4} \leq x \leq 5) = \int_{\frac{3}{4}}^5 f(x) dx = \int_{\frac{3}{4}}^1 1 dx + \int_1^5 0 dx = 1 - \frac{3}{4} = \frac{1}{4}$
fuori: intervalli

Interpretazioni
delle variabili
aleatorie

Nel caso di estrazioni di popolazioni: $\Omega = \{\text{popolazione}\} = \{\text{cittadini italiani}\}$

$X = \text{carattere quantitativo discreto, considero } X(w) = \# \text{ figli di: } w$ "

$P_x(1) = P(X^{-1}(1)) = \#\{\omega \in \Omega \mid X(\omega) = 1\} = \#\text{ cittadini italiani con figlio unico}$

probabilita' che un cittadino italiano $\in \Omega$ abbia un figlio unico $\in \Omega$ $\#\text{ popolazione italiana}$

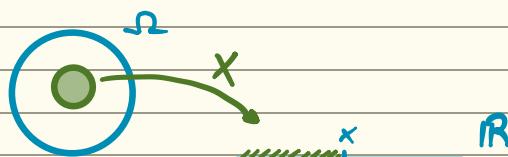
$X = \text{carattere quantitativo con densita', considero } X(w) = \text{altezza di: } w$

$$P(1,65 \leq x \leq 1,75) = \int_{1,65}^{1,75} f(x) dx =$$


Funzione di
ripartizione

Prese X v.a. e $P_x: P(\mathbb{R}) \rightarrow \mathbb{R}$ legge di probabilita', la funzione di ripartizione di X è la funzione $F_x: \mathbb{R} \rightarrow [0, 1]$ definita come $F_x(x) = P(X \leq x) = P_x((-\infty, x]) \quad \forall x \in \mathbb{R}$ assegna ad ogni $x \in \mathbb{R}$ la probabilita' che X assuma un valore al massimo pari a x

Oss: attenzione tra X grande e x piccolo.



Proprietà di F_x

Preso F_x funzione di ripartizione di X :

1) F_x è non decrescente ($x_1 \leq x_2 \Rightarrow F(x_1) \leq F(x_2)$)

2) $\lim_{x \rightarrow -\infty} F_x(x) = 0$; $\lim_{x \rightarrow +\infty} F_x(x) = 1$

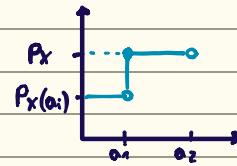
3) F_x è continua a destra ($x_n \rightarrow x \Rightarrow F(x_n) \rightarrow F(x)$)

F_x per X v.a. discreta

Se X è una v.a. discreta, allora $F_x(x) \doteq P_x\{X \leq x\} = \sum_{a_i \leq x} P_x(a_i)$

Ne deriva una funzione a gradini:

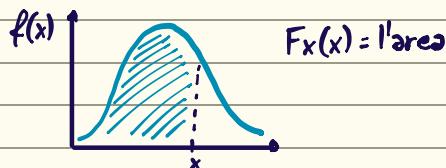
- ha un salto di ampiezza $P_x(a_i)$ in corrispondenza di ogni valore
- è costante tra due valori consecutivi
- Se X assume un valore finito di valori: $a_1, \dots, a_n \Rightarrow F(x) = 1 \quad \forall x \geq x_n$



F_x per X v.a. continue

Se X è una v.a. con densità f , allora $F_x(x) \doteq P\{X \leq x\} = \int_{-\infty}^x f(y) dy$

Ovvero è uguale all'area sottesa alla funzione densità nell'intervallo $(-\infty, x]$



Calcolo della probabilità di intervalli:

$$P(a < x < b) = F_x(b) - F_x(a)$$

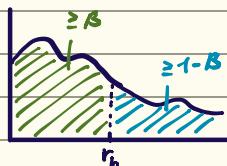
dimostrazione

$$F_x(b) - F_x(a) = P\{X \leq b\} - P\{X \leq a\} = P(\{X \leq b\} \setminus \{X \leq a\}) = P(a \leq X \leq b)$$

B-quantile

Data X v.a. e dato $\beta \in (0,1)$, definisco B-quantile (o 100 β percentile) un numero r_β tale che $P\{X \leq r_\beta\} \geq \beta$, $P\{X \geq r_\beta\} \geq 1 - \beta$

Ovvero, il β -quantile è un numero r tale che con probabilità (almeno) β cado a sinistra di r e con probabilità (almeno) $1 - \beta$ cado a destra di r



oss: se F_x è invertibile, allora $\forall \beta \in (0,1) \exists! \text{ B-quantile } r_\beta = F_x^{-1}(\beta)$

ESERCIZI FUNZIONE DI RIPARTIZIONE (F_x)

1) Un negoziante di abbigliamento ha acquistato un capo per 200€ e può:

- rivenderlo prima dei saldi: a 400€
- rivenderlo coi saldi: a 250€
- non rivenderlo

La probabilità che lo rivenda prima dei saldi è 0.5 e coi saldi: 0.3

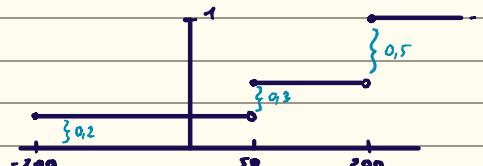
X v.a. "profilo del negoziante" $\begin{cases} 200 & P=0.5 \\ 50 & P=0.3 \\ -200 & P=0.2 \end{cases}$

$P_x(-200, 50, 200) \rightarrow \mathbb{R}$

$$P_x(-200) = 0.2 \quad P_x(50) = 0.3 \quad P_x(200) = 0.5$$

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

$$F_x(x) = P\{X \leq x\} = P_x(X^{-1}(-\infty, x])$$



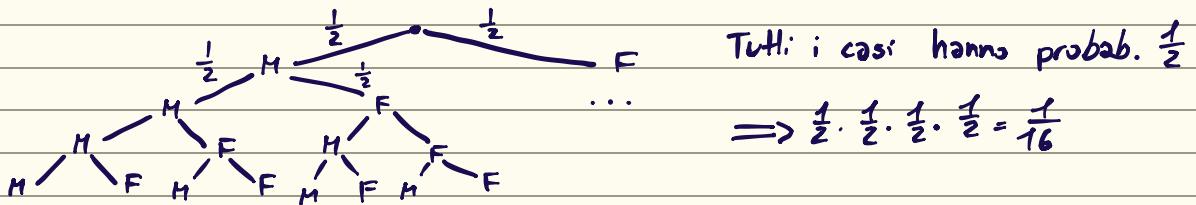
2) $X = \text{"figli: maschi: in famiglia con 4 figli:}"$

- costruire una tabella della distribuzione di prob. di X
- rappresentare graficamente la distribuzione

I possibili valori di X sono $\{0, 1, 2, 3, 4\} \Rightarrow$ v.a. discreta

$P_X : \{0, 1, 2, 3, 4\} \rightarrow \mathbb{R}$ $\omega_i \rightarrow P(X=\omega_i) = P(X^{-1}(\omega_i))$

Consideriamo M e F equiprobabili $\Rightarrow P(M) = P(F) = \frac{1}{2}$



Tutti i casi hanno probab. $\frac{1}{2}$
 $\Rightarrow \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{16}$

V.A. notevoli

Vari tipi di variabili aleatorie che sono modelli per vari casi

Variabile Casuale di Bernoulli

Ogni volta che ho un esperimento con un risultato dicotomico (successo, insuccesso) definisco la v.a. X con lo spazio $\Omega = \{S, F\}$, $X : \Omega \rightarrow \mathbb{R}$

Probabilità di successo $p \in (0, 1) = P_X(S)$



$$X(S) = 1, X(F) = 0$$

$$P_X(0) = P(X=0) = P(X^{-1}(0)) = P(F) = 1-p$$

$$P_X(1) = P(X=1) = P(X^{-1}(1)) = P(S) = p$$

Quindi se identifico una v.a. come in questo caso X , con questo spazio Ω ho le due proprietà $P_X(0)$ e $P_X(1)$ e chiamo X come $B(0, p)$

$$\text{es. } \Omega = \{T, C\} \quad X : \Omega \rightarrow \mathbb{R}$$

$$T \rightarrow 1 \quad C \rightarrow 0 \quad p = 0.5$$

$$P_X(1) = 0.5$$

$$P_X(0) = 1 - 0.5 = 0.5$$

Quindi X è $B(1, 0.5)$

Moneta truccata con $p = 0.8$

$$P_X(1) = 0.8$$

$$P_X(0) = 1 - p = 0.2$$

$$B(1, 0.8)$$

Variabile casuale binomiale

Esperimento con risultato dicotomico $\{S, F\}$ ripetuto n volte nelle condizioni iniziali (indipendenti).

Preso quindi: $\Omega = \{0, 1\}^n = \{(w_1, \dots, w_n) | w_i \in \{0, 1\}\}$ e p probabilità di successo nella singola prova. Definisco $X = \# \text{successi in } n \text{ prove ripetute}$. La probabilità di avere una sequenza di k successi è $P_X(k) = P\{X=k\} = P(X^{-1}(k))$

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

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

$$\left\{ \begin{array}{l} \text{Il caso generale è} \\ \sum_{\text{sequenze con } k \text{ successi}} p^k \cdot (1-p)^{n-k} = P_X(k) \\ \binom{n}{k} \end{array} \right.$$

Quindi questa v.a. è $B(n, p)$ = v.a. discreta con distribuzione di probabilità data da $P(k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$.

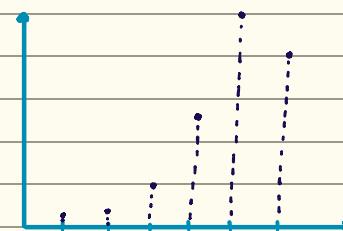
es. esperimento = lancio moneta truccata con $p = 0.8$

$X = \# \text{teste in 5 lanci}$

$\Rightarrow B(5, 0.8)$ visto che X è v.a. discreta

$$P_X(k) = P(X^{-1}(k)) = \binom{5}{k} (0.8)^k \cdot (1-0.8)^{5-k}$$

num	P_X
0	$(0.2)^5 = 0.00032$
1	$\binom{5}{1} \cdot (0.8)^1 \cdot (0.2)^4 = 0.00064$
2	$\binom{5}{2} (0.8)^2 (0.2)^3 = 0.016$
3	$\binom{5}{3} (0.8)^3 (0.2)^2 = 0.2$
4	$\binom{5}{4} (0.8)^4 (0.2)^1 = 0.4$
5	$\binom{5}{5} (0.8)^5 = 0.32$



Esempi di v.a. binomiali:

Sono esempi di v.a. binomiali:

. * teste in n lanci di monete $\rightarrow B(n, \frac{1}{2})$

. * di "6" in n lanci di un dado $\rightarrow B(n, \frac{1}{6})$

. * biglie rosse estratte, in n estrazioni con rimpiazzo, da un'urna con biglie rosse e blu

Variabile casuale geometrica

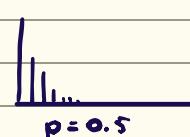
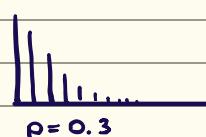
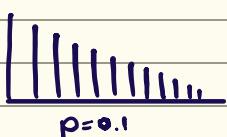
Esperimento con esito dicotomico {S, F} ripetuto infinite volte, con probabilità p .

Allora $\Omega = \{0, 1\}^{\mathbb{N}^+} = \{(a_1, a_2, \dots) | a_i \in \{0, 1\} \forall i \in \mathbb{N}^+\}$ e $T = \min\{n \in \mathbb{N}^+ | w_n = 1\}$

ovvero l'istante (n° della prova) in cui si verifica il primo successo. $0, 0, 0, 0, 1, 1, 0, \dots \Rightarrow T$ è una v.a. discreta a valori in \mathbb{N}^+

Quindi la distribuzione di T è: $P_T(k) = P(T^{-1}(k)) = P(\underbrace{\{0, \dots, 0, 1\}}_{k-1}) = p \cdot (1-p)^{k-1} \quad k \in \mathbb{N}^+$

Ovvero la v.c. geometrica è $G(p)$ = v.a. discreta con valori in \mathbb{N}^+ con distribuzione di probabilità $P_T(k) = p \cdot (1-p)^{k-1}$



All'aumentare di p la funzione probabilità decresce più rapidamente

es. probabilità che un docente che ha ricevuto la copia omaggio abbia il libro è 0,11
determinare la distribuzione del numero di prof a cui dare la copia omaggio, affinché l'ultimo abbia il libro.

$X = *$ di prof. a cui dare il libro affinché l'ultimo lo abbia: $\sim G(0,11)$

$$P_X \left\{ \underbrace{\{0, 0, 0, 1\}}_k \right\} = p \cdot (1-p)^{k-1} \Rightarrow P\{X=1\} = p(1-p)^0 = 0,11 \\ \Rightarrow P\{X=2\} = p(1-p)^1 = 0,11 \cdot 0,89 = 0,098 \\ \Rightarrow P\{X=3\} = p(1-p)^2 = 0,11 \cdot (0,89)^2 = 0,087$$

dove $\{X=1\}$ = primo prof accetta, $\{X=2\}$ = primo rifiuta, secondo accetta, ...

La probabilità va gradualmente a diminuire

Proprietà assenza
di memoria

Se $T \sim G(p)$, allora $P(T=n+k | T > n) = P(T=k) \quad \forall n, k \in \mathbb{N}^+$

Traduzione: la distribuzione del 1° successo dalla $(n+1)$ -esima prova non cambia se non si è verificato nelle prime n prove

Var. di Poisson
di parametro $\lambda > 0$

$Poisson(\lambda)$ con $\lambda > 0$ è una v.a. discreta che assume i valori in \mathbb{N} .

La funzione di massa è data da $P_X(k) = P\{X=k\} = \frac{\lambda^k \cdot e^{-\lambda}}{k!} \quad k \in \mathbb{N}$

. λ rappresenta il *medio di eventi nell'intervalle $Poisson(\lambda) \sim B(n, p)$

. $Poisson(\lambda)$ conta *successi in prove ripetute

. Il fattore $e^{-\lambda}$ serve per far sì che P_X sia una funzione di massa

infatti: $\sum_{k=0}^{\infty} P_X(k) = \sum_{k=0}^{\infty} \frac{\lambda^k \cdot e^{-\lambda}}{k!} = e^{-\lambda} \cdot \underbrace{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}_{\sum_{k=0}^{\infty} \frac{1}{k!} \div e} = e^{-\lambda} \cdot e^{\lambda} = 1$

es. Alle casse di un supermercato si presentano in media 3 clienti/minuto

$X = *$ clienti che arrivano alla cassa in un minuto $\sim Poiss(3)$

$$P_X(k) = \frac{3^k}{k!} \cdot e^{-3}$$

$$P_X\{X > 4\} = 1 - P\{X \leq 4\} = 1 - [P\{X=0\} + P\{X=1\} + P\{X=2\} + P\{X=3\} + P\{X=4\}]$$

$$P_X(0) = \frac{3^0}{0!} \cdot e^{-3} \approx 0,0498$$

$$P_X(1) = \frac{3^1}{1!} \cdot e^{-3} \approx 0,1494$$

V.A.

Ipergeometriche

Supponiamo di avere N unità di cui:

- F "favorevoli" \leftrightarrow SUCCESSO
- $N-F$ "non favorevoli" \leftrightarrow INSUCCESSO

Dell'insieme delle N unità ne estraggo n a caso (senza rimpiazzo)

$X =$ "nunità favorevoli nelle n prove"

$X =$ v.a. ipergeometrica di parametri $IP(N, F, n)$

$$P_X(k) = P(X=k) = \frac{\text{caso favorevole}}{\text{caso possibile}} = \frac{\text{n-uple in cui ho } k \text{ elementi di } F}{\text{n-uple che posso estrarre}}$$

In particolare:

• $\text{caso favorevole} = \text{n-uple in cui ci sono } k \text{ elementi di } F \text{ e } n-k \text{ di } N-F = \binom{F}{k} = \binom{N-F}{n-k}$

• $\text{caso possibile} = \text{n-uple di } N \text{ elementi} = \binom{N}{n} = \frac{N!}{n!(N-n)!}$

Ne deriva $P\{X=k\} = \frac{\binom{F}{k} \binom{N-F}{n-k}}{\binom{N}{n}}$

es. Un'azienda riceve da un fornitore un pacco con 20 pezzi.

L'azienda ne prova 5 estratti dal pacco, se nessuno è difettoso la fornitura è accettata, altrimenti è respinta. Qual è la probabilità di accettare la fornitura con 6 pezzi difettosi?

$X =$ "pezzi difettosi nelle 5 estrazioni"

$$\left. \begin{array}{l} N=20 \\ F=6 \text{ difettosi} \\ n=5 \text{ estratti} \end{array} \right\} \Rightarrow X \sim IP(20, 6, 5)$$

L'idea è voler calcolare la probabilità che nessuno di quelli estratti sia difettoso ovvero: $P\{X=0\}$ dove $X \sim IP(20, 6, 5)$

$$P\{X=0\} = \frac{\binom{6}{0} \binom{20-6}{5-0}}{\binom{20}{5}} = \frac{1 \cdot \binom{14}{5}}{\binom{20}{5}} = \frac{14!}{5! 9!} \cdot \frac{5! 15!}{20!} = \frac{14 \cdot 13 \cdot 12 \cdot 11 \cdot 10}{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16} = 0.125$$

Osservazione

Le v.a. viste finora sono discrete, le prossime sono v.a. con densità, in generale:

Sia X una v.a. con densità f , se $f=0$ al di fuori di un certo intervallo

[ovvero su $(c, d)^c$] \Rightarrow

$$P\{X \in (c, d)^c\} = P\{X \in (a, b) \cap (c, d)\} = P\{X^{-1}(c, d)^c\} = \int_c^d 0 dx = 0$$

V.A. Uniformi

Assumono valori in un intervallo e sono costanti in esso.

Definiamo X v.a. uniforme sull'intervallo (a, b) (notazione $U(a, b)$)

la cui f di densità è $f(x) = \begin{cases} \frac{1}{b-a} & \text{per } x \in (a, b) \\ 0 & \text{per } x \notin (a, b) \end{cases}$

Si indica con $U_{(a, b)}$

$$P\{a \leq X \leq b\} = \int_a^b \frac{1}{b-a} \cdot U_{(a, b)}(x) dx = \frac{1}{b-a} \int_{(a, b)} 1 dx = \frac{\text{Lung}((a, b))}{b-a} = \frac{\text{Lung}((a, b))}{b-a} \rightarrow (= \text{lung} (a, b))$$



es. Nel gioco roulette un ago viene fatto ruotare in modo tale da non privilegiare nessun punto della circonferenza quando si arresta, quindi scelta di un punto $(0, 2\pi)$ senza preferenze.

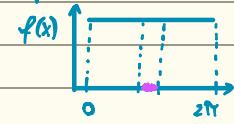
$X =$ "angolo ottenuto quando l'ago si ferma" \Rightarrow v.a. che prende valori in $(0, 2\pi)$ senza preferenze

$$\Rightarrow X \sim U(0, 2\pi) \text{ quindi: } f(x) = \begin{cases} \frac{1}{2\pi} & \text{per } x \in (0, 2\pi) \\ 0 & \text{altrimenti} \end{cases}$$

La ruota è divisa in spicchi. Tra questi c'è uno spicchio di ampiezza $\pi/6$ tale che, quando l'ago ci si ferma, il giocatore VINCÉ il turno. Determinare la probabilità che il giocatore PERDA il turno (non cade in quello spicchio).

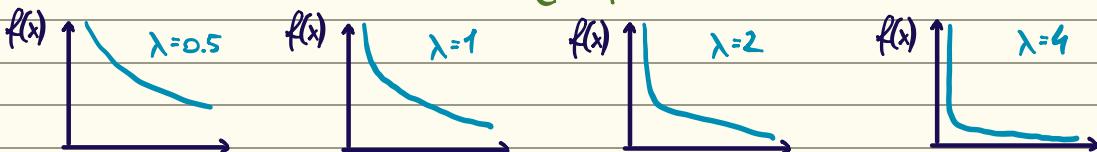
$$P\{X \in (\text{quello spicchio})^c\} = \text{area al di sotto del grafico del complementare dello spicchio}$$

$$= 1 - \frac{\pi}{6} \cdot \frac{1}{2\pi} = 1 - \frac{1}{12} = \frac{11}{12}$$



V.A.
esponenziali
di parametro
 $\lambda > 0$

$$\text{Exp}(\lambda): \text{v.a. con densità } f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{per } x > 0 \\ 0 & \text{per } x \leq 0 \end{cases} = \lambda e^{-\lambda x} \cdot \mathbb{1}_{(0,+\infty)}(x)$$



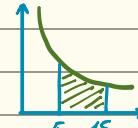
Si usa per descrivere i tempi di attesa affinché si verifichi un evento

es. Il tempo impiegato dal commesso di un negozio per assistere i clienti si distribuisce come una v.a. esponenziale negativa

$X = \text{"tempo impiegato per assistere i clienti"} \sim \text{EXP}(\lambda)$

Con $\lambda = \frac{1}{10}$ minuti, calcolare la probabilità che per assistere un cliente il commesso impieghi tra 5 e 15 minuti.

$$f(x) = \begin{cases} \frac{1}{10} e^{-\frac{1}{10}x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$



$$P(5 \leq x \leq 15) = \int_5^{15} \frac{1}{10} \cdot e^{-\frac{1}{10}x} dx = -e^{-\frac{1}{10} \cdot 15} + e^{-\frac{1}{10} \cdot 5} = e^{-\frac{1}{2}} - e^{-\frac{3}{2}} = 0.3834$$

$$\boxed{\int_a^b \frac{1}{\lambda} e^{-\lambda x} dx = -e^{-\lambda x} \Big|_a^b = -e^{-\lambda b} + e^{-\lambda a}}$$

es.2 Il tempo di vita in giorni di un certo macchinario è descritto da una v.a. espon. di parametro $\frac{1}{8}$. Qual è la probabilità che il primo guasto si verifichi dopo 6 giorni?

$$X = \text{"giorni prima del primo guasto"} \sim \text{EXP}\left(\frac{1}{8}\right) = \begin{cases} \frac{1}{8} e^{-\frac{1}{8}x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

$$P\{X \leq 6\} = \int_0^6 \frac{1}{8} \cdot e^{-\frac{1}{8}x} dx = [-e^{-\frac{1}{8}x}]_0^6 = e^{-\frac{1}{8} \cdot 0} - e^{-\frac{1}{8} \cdot 6} = e^0 - e^{-\frac{3}{4}} = 1 - e^{-\frac{3}{4}}$$

Se il macchinario non si è rotto per i primi 2 gg, qual è probabilità che non si rompa prima di altri 6 giorni.

$$P\{X > 8 | X > 2\} = P\{X > 6\} = 1 - P\{X \leq 6\} = 1 - (1 - e^{-\frac{3}{4}}) = e^{-\frac{3}{4}}$$

└ assenza di memoria

Trasformazione
della v.a.
con densità

Sia $X: \Omega \rightarrow \mathbb{R}$ una v.a. con densità $f_X: \mathbb{R} \rightarrow \mathbb{R}$ ($P\{X \in A\} = \int_A f_X(x) dx$)

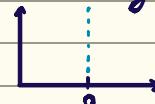
Sia poi $h: \mathbb{R} \rightarrow \mathbb{R}$ una funzione (es. $h(x) = x^2$)

Allora $Y = h \circ X: \Omega \rightarrow \mathbb{R}$ è anche lei una v.a.

Non è detto che Y ammetta densità, e anche quando la ammette, c'è una regola generale per calcolarla.

$$\cdot \equiv 0 \text{ allora } Y = h \circ X -$$

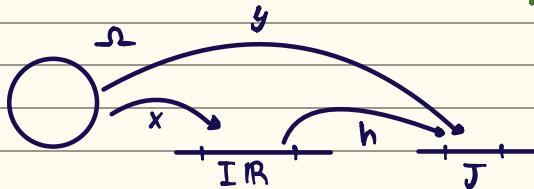
Se, per esempio, h è la funzione $, e a sua volta \equiv 0$ e non c'è modo di trovare una funzione tale che $P(Y \in A)$ sia uguale all'area al di sotto del grafico



- Lo stesso è vero ~~f~~ funzione $h: \mathbb{R} \rightarrow \mathbb{R}$ a valori in un intervallo finito o numerabile, per lo stesso motivo
- La funzione $Y = h \circ X$ ammette invece densità se h è C^1 e bivoca, più precisamente vale la proposizione seguente

Proposizione cambio di variabile

Sia X una v.a. con densità f_X supportata in un intervallo aperto I (cioè t.c. $f(x)=0$ in I^c)
Sia $h: I \rightarrow J$ un **DIFFOMORFISMO** (C^1 invertibile con inversa C^1) allora la v.a. $Y = h \circ X$ ammette densità data da $f_Y(y) = \begin{cases} f_X(h^{-1}(y)) \left| \frac{dh^{-1}}{dy}(y) \right| & \text{con } y \in J \\ 0 & \text{con } y \notin J \end{cases}$
 $= f_X(h^{-1}(y)) \cdot \left| \frac{dh^{-1}}{dy}(y) \right| \cdot u_J(y)$



V.A. gaussiane (o normali)

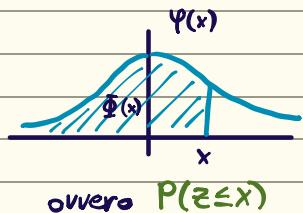
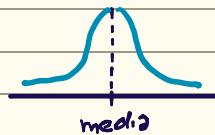
Modello adatto a problemi in cui si può ipotizzare che la v.a. assuma con probabilità più alta valori attorno alla media. Deve avere una funzione di densità che ha il massimo in corrispondenza della media e decresce in modo simmetrico.

È v.a. normale standard $\hat{=}$ la v.a. con densità $\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$ $x \in \mathbb{R}$

N(0,1)

Notazione **N(0,1)**

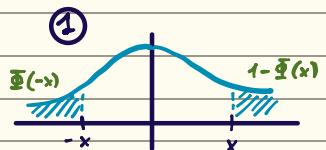
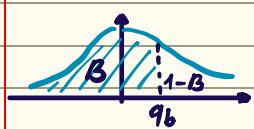
Si può dire che $\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$
 \Rightarrow L'area sotto al grafico è $\frac{\sqrt{2\pi}}{\sqrt{2\pi}} = 1$



Funzione di ripartizione di N(0,1)

Sì dimostra che è $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$

Si indica con q_β il β -quantile di $N(0,1)$, ovvero il numero tale che $\Phi(q_\beta) = \beta$ con $\beta \in (0,1)$



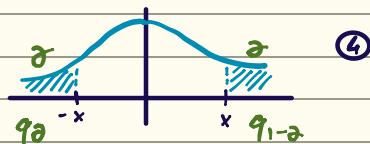
Proprietà di: N(0,1)

$$1) \Phi(-x) = 1 - \Phi(x)$$

$$2) P\{-x \leq Z \leq x\} = \Phi(x) - \Phi(-x) = \Phi(x) - [1 - \Phi(x)] = 2\Phi(x) - 1$$

$$3) \Phi(0) = P\{Z \leq 0\} = \frac{1}{2}$$

$$4) q_{1-\alpha} = -q_\alpha \Rightarrow \text{l'area fino a } q_\alpha \text{ è } \alpha, \text{ fino a } q_{1-\alpha} \text{ l'area è } 1-\alpha, \text{ quella dopo è } \alpha$$



Se $Z = N(0,1)$ non è possibile rappresentare l'integrale $P\{a < Z < b\} = \int_a^b \varphi(x) dx$ tramite funzioni più semplici.

In altri termini: \nexists formule esplicite per quell'integrale. Si usa una funzione di ripartizione Φ : $P\{a < Z < b\} = \int_a^b \varphi(x) dx = \Phi(b) - \Phi(a)$

In particolare:

- Per $x \geq 4$, $\Phi(x) \approx 1$ (cioè Φ è talmente vicina a 1 da potercela approssimare)
- Per $x < 0$, ci si riporta al caso positivo, essendo $\Phi(x) = 1 - \Phi(-x)$
- \Rightarrow servono sostanzialmente solo i valori di $\Phi(x)$ con $0 < x < 4$

Usando i valori numerici ottenuti per Φ si possono ottenere i quantili:

- . per $a \in (\frac{1}{2}, 1)$ vale che q_a è tale che $\Phi(q_a) = a$
- . per $a \in (0, \frac{1}{2})$ vale che $q_a = -q_{1-a}$

$$\text{es. } P\{-1 \leq Z \leq 1\} = 2\Phi(1) - 1 \approx 0.68$$

$$P\{-2 \leq Z \leq 2\} \approx 0.99$$

$$P\{-3 \leq Z \leq 3\} \approx 0.997$$

Variabile Gaussiana Generale

Se $Z \sim N(0,1)$, posso ottenere un'altra v.a. con densità componendo con $h: \mathbb{R} \rightarrow \mathbb{R}$ $X \mapsto \sigma X + m$
La variabile Y ottenuta è $Y = \sigma Z + m$ ed è possibile calcolare la sua densità tramite la formula di cambio variabile

$$f_Y(y) = f_Z(h^{-1}(y)) \cdot \left| \frac{dh}{dy}(y) \right| = f_Z\left(\frac{y-m}{\sigma}\right) \cdot \frac{1}{\sigma} = \frac{1}{\sigma} \varphi\left(\frac{y-m}{\sigma}\right) \cdot \frac{1}{\sigma} = \frac{1}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} \cdot \exp^{-\frac{(y-m)^2}{2\sigma^2}}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{(y-m)^2}{2\sigma^2}\right] \Rightarrow \text{v.a. gaussiana di media "m" e varianza "\sigma^2"}$$

Calcoli e standardizzazione

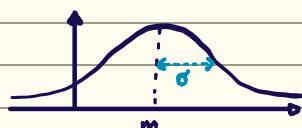
Dato $N(m, \sigma^2)$, per fare i calcoli ci si riconduce a $N(0,1)$ tramite il cambiamento di variabile inverso rispetto a quello visto nella simmetria (STANDARDIZZAZIONE)

$$Z \sim N(0,1) \iff Y = \sigma Z + m \sim N(m, \sigma^2), \text{ quindi:}$$

$$Z = \frac{Y-m}{\sigma} \Rightarrow P\{a \leq Y \leq b\} = P\{a \leq \sigma Z + m \leq b\} = P\left\{\frac{a-m}{\sigma} \leq Z \leq \frac{b-m}{\sigma}\right\} = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right)$$

In particolare:

- . $P\{m-\sigma \leq Y \leq m+\sigma\} = P\left\{\frac{m-\sigma-m}{\sigma} \leq Z \leq \frac{m+\sigma-m}{\sigma}\right\} = P\{-1 \leq Z \leq 1\} \approx 0.68$
- . $P\{m-2\sigma \leq Y \leq m+2\sigma\} = P\{-2 \leq Z \leq 2\} \approx 0.99$
- . $P\{m-3\sigma \leq Y \leq m+3\sigma\} = P\{-3 \leq Z \leq 3\} \approx 0.997$



Proprietà di riproducibilità

Sia Y una v.a. gaussiana $\sim N(m, \sigma^2)$
Sia $V = \alpha Y + \beta$ con $\alpha, \beta \in \mathbb{R}, \alpha \neq 0$

dimostrazione:

$$Y \sim N(m, \sigma^2) \rightarrow f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{(y-m)^2}{2\sigma^2}\right] = Y \sim N(m, \sigma^2) \Rightarrow Z = \frac{Y-m}{\sigma} \sim N(0,1)$$

$$V = \alpha Y + \beta \iff Y = \frac{V-\beta}{\alpha} \Rightarrow Z = \frac{V-\beta-m}{\sigma} = \frac{V-\beta-\alpha m}{\sigma} = \frac{V-(\alpha m + \beta)}{\sigma}$$

$$\Rightarrow V \sim N(\alpha m + \beta, \sigma^2 \alpha^2)$$

es. Una fabbrica produce delle viti, la cui lunghezza in cm $\sim N(1.7, (0.6)^2)$

a) Calcola la probabilità che la lunghezza di una vite estratta a caso sia < 1.6

$$Y \sim N(1.7, (0.6)^2) \rightarrow Z \sim N(0,1) \text{ se } \frac{Y-1.7}{0.6} \text{ (Standardizzazione)}$$

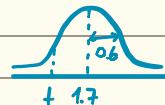
$$\Rightarrow P\{Y < 1.6\} = P\{0.62 + 1.7 < 1.6\} = P\{Z < \frac{-0.17}{0.6}\} = \Phi(-0.17) =$$

$$1 - \Phi(0.17) \approx 1 - 0.567 \approx 0.433$$

b) Determinare un valore $t \in \mathbb{R}$ t.c. il 95% delle viti abbia lunghezza almeno t

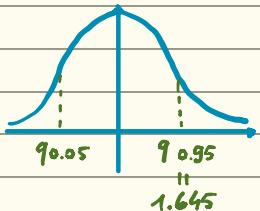
$$P\{Y \geq t\} = P\{0.6Z + 1.7 \geq t\} = P\left\{Z \geq \frac{t-1.7}{0.6}\right\} = 1 - P\left\{Z \leq \frac{t-1.7}{0.6}\right\} = 0.95$$

$$\Rightarrow 1 - \Phi\left(\frac{t-1.7}{0.6}\right) = 0.95$$



Cerco quindi il quantile $q_{0.05}$

$$q_{-0.05} = -q_0 \Rightarrow q_{0.05} = q_{1-0.95} = -q_{0.95} \approx 1.645 \Rightarrow \frac{t-1.7}{0.6} \approx -1.645 \Rightarrow t \approx 0.713$$



STATISTICA DESCRIPTIVA NELLA PROBABILITÀ

Portiamo i concetti della statistica descrittiva nel contesto della probabilità.
Servono degli indici che caratterizzano le v.a., ad esempio posizione e variabilità.

Valore atteso $E[X]$
o speranza
o momento primo

Indice che fornisce informazioni sulla posizione di una v.a. X , si indica con $E[X]$ oppure μ ed è definito come segue:

- . X v.a. discreta $\Rightarrow E[X] \doteq \sum_i a_i \cdot p(a_i)$ [in caso gl: a_i siano infiniti]

$$\sum_{i=1}^{\infty} a_i \cdot p(a_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n a_i \cdot p(a_i)$$
- . X v.a. con densità $f \Rightarrow E[X] = \int_{-\infty}^{+\infty} x f(x) dx$ ($= \lim_{b \rightarrow +\infty} \int_{-\infty}^b x f(x) dx$)

es. v.a. discreta

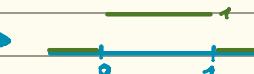
x = esito di un lancio di un dado equilibrato

$$\{a_1, \dots, a_6\} = \{1, 2, \dots, 6\} \quad P\{X=k\} = \frac{1}{6} \quad \forall k \in \{1, \dots, 6\}$$

$$\text{Allora } E[X] = \sum_{i=1}^6 a_i \cdot p(a_i) = \sum_{i=1}^6 i \cdot \frac{1}{6} = \frac{1}{6} \cdot \sum_{i=1}^6 i = \frac{1}{6} \cdot \frac{6 \cdot 7}{2} = \frac{7}{2} = 3.5$$

$\hookrightarrow S(n) = \frac{n(n+1)}{2}$

es. v.a. con densità

Sia $X \sim U(0,1)$ \rightarrow  ($\Rightarrow f(x) = \begin{cases} 1 & \text{per } x \in (0,1) \\ 0 & \text{altrimenti} \end{cases}$)

$$\text{Allora } E[X] = \int_{-\infty}^{+\infty} x \cdot f(x) dx = \int_{-\infty}^{+\infty} x \cdot \mathbb{1}_{(0,1)}(x) dx = \int_0^1 x dx = \left[\frac{1}{2} x^2 \right]_0^1 = \frac{1}{2}$$

Significato
di $E[X]$

- . Indica intorno a che valore è "centrata" la distribuzione, una sorta di baricentro
- . Se X è discreto, l'idea è considerare come valore atteso una media dei valori assunti da X , pesati in modo che quelli più probabili contino di più
- . Se X è una funzione con densità, il concetto è lo stesso, si sostituisce la somma pesata delle f di massa con l'integrale pesato delle densità.
Lo valore atteso non è sempre finito e non sempre definito.

es. v.a. X con densità $f_X(x) = \frac{1}{x^2} \cdot \mathbb{1}_{(1, \infty)}(x)$

$$\text{allora } E[X] = \int_{-\infty}^{+\infty} x f(x) dx = \int_{-\infty}^{+\infty} x \cdot \frac{1}{x^2} \cdot \mathbb{1}_{(1, \infty)}(x) dx = \int_1^{+\infty} \frac{1}{x} dx = [\log(x)]_1^{+\infty} = \infty$$

Quando $E[X]$
è definito?

- . Se X v.a. (discreta o con densità) è non negativa
 \Rightarrow Ha sempre senso scrivere $E[X]$, ammettendo anche che possa assumere ∞

infatti: $E[X] = \begin{cases} \sum_k a_k \cdot p(a_k) & \text{caso discreto} \rightarrow \text{Se i valori sono tutti positivi ha sempre senso} \\ \int_{\mathbb{R}} x f(x) dx & \text{caso con densità} \rightarrow \int_{\mathbb{R}} x f(x) dx \in [0, \infty] \text{ l'integrale indefinito converge o tende a } +\infty \end{cases}$

. Se X v.a. generica ammette $E[X]$ se:

- caso discreto: $\sum_k |a_k| p_x(a_k) < \infty \rightarrow$ Se X assume un *finito di valori allora ammette valore atteso (somma finita)

- caso con densità: $\int_{-\infty}^{+\infty} |x| f(x) dx < \infty \rightarrow$ Se f è zero fuori da un intervallo limitato $[a, b]$ (es. $P\{X \in [a, b]\} = 1$), ammette valore atteso perché si riduce a \int_a^b

In tal caso, $\exists E[X] \in \mathbb{R}$

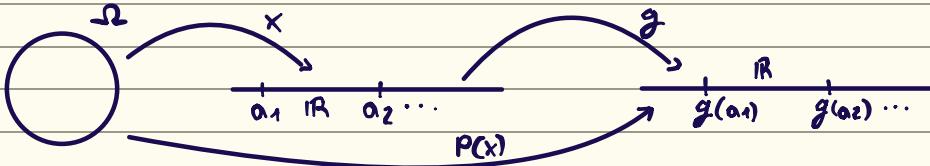
Proposizione

X v.a. discreta, oppure con densità f , $g: \mathbb{R} \rightarrow \mathbb{R}$ t.c. $g(x)$ sia v.a.

Allora:

1) Se X è discreta con funzione di massa $p \Rightarrow E[g(x)] = \sum g(a_k) \cdot p(a_k)$

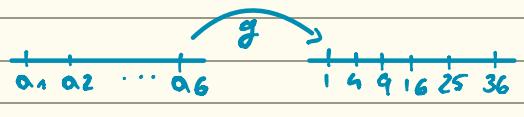
2) Se X ha densità $f \Rightarrow E[g(x)] = \int_{-\infty}^{+\infty} g(x) f(x) dx$



es. X = lancio di un dado equilibrato $\{1, 2, 3, 4, 5, 6\}$, $g: \mathbb{R} \rightarrow \mathbb{R}$ $x \mapsto x^2$

$$\Rightarrow E[g(x)] = E[x^2] = \sum g(a_k) \cdot p(a_k)$$

$$= \sum_k k^2 \frac{1}{6} = \frac{1}{6} \cdot [1+4+9+16+25+36] = \frac{91}{6}$$



Proprietà del valore atteso

X e Y v.a. con valore atteso e $a, b \in \mathbb{R}$, allora:

1) $X+Y$ ha valore atteso e $E[X+Y] = E[X] + E[Y]$

2) $aX+b$ ha valore atteso e $E[aX+b] = aE[X] + b$ (quindi $E[b] = b$)

3) Se $X \geq 0$ allora $E[X] \geq 0$

Se $X \geq Y$ allora $E[X] \geq E[Y]$

Traduzione: la funzione $E: \{\text{v.a. con val. atteso}\} \rightarrow \mathbb{R}$ è lineare ① ② e monotona ③

es. Consideriamo due lanci di un dado equilibrato, qual è il valore atteso della somma?

X = "esito 1° lancio" Y = "esito 2° lancio"

$$E[X] = E[Y] = \sum_{k=1}^6 k \cdot \frac{1}{6} = \frac{1+2+3+4+5+6}{6} = \frac{21}{6}$$

$$E[X+Y] = E[X] + E[Y] = 7/2 + 7/2 = 7$$

Momento di ordine n di X $E[X^n]$

Data X v.a., chiamiamo momento di ordine n di X il valore atteso di X^n , ovvero $E[X^n]$ (ad esempio, $E[X] = \text{momento di ordine } 1$)

. caso discreto: $E[X^n] = \sum_k a_k^n \cdot p(x=a_k)$

. caso con densità: $E[X^n] = \int_{-\infty}^{+\infty} x^n f(x) dx$

Diseguaglianza di Markov

$X \geq 0$ v.a., $a > 0$, allora $P\{X \geq a\} \leq \frac{1}{a} E[X]$

Varianza di X $\text{var}(x)$

X v.a. con momento secondo (es. t.c. $E[X^2] < \infty$), allora

$$\begin{aligned} \text{var}(x) &= E[(X - E[X])^2] = E[X^2 - 2X E[X] + E[X]^2] = E[X^2] - 2E[X] \cdot E[X] \cdot E[E[X]^2] \\ &\quad \Rightarrow E[X^2] - E[X]^2 \end{aligned}$$

Deviazione stand. di X $d(x)$

$$\sigma(x) = \sqrt{\text{var}(x)}$$

Significato di: $\text{var}(x)$

Come per $E[X]$, $\text{var}(x)$ dipende dalla distribuzione di X .

Come nella statistica descrittiva, la varianza fornisce una misura della dispersione di una v.a. rispetto al suo valore atteso

es. X = "lancio di un dado" v.a. discreta con valori $\{1, 2, \dots, 6\}$, calcolare $\text{Var}(X)$
 $\text{Var}(X) = E[X^2] - E[X]^2 = \underbrace{9/6}_{\hookrightarrow \text{dall'esempio precedente}} - (7/2)^2 = 35/12$

Proprietà di Scaling

- $\text{Var}(\alpha X + b) = \alpha^2 \cdot \text{Var}(X)$
- In particolare, $\text{Var}(b) = 0$

dimostrazione

$$\text{Var}(\alpha X + b) = E[(\alpha X + b)^2] - (E[\alpha X + b])^2 = \alpha^2 [E[X^2] - E[X]^2] = \alpha^2 \text{Var}(X)$$

Diseguaglianza di Chebyshev

X v.a. con momento secondo, $d > 0$. Allora:

$$P\{|X - E[X]| \geq d\} \leq \frac{1}{d^2} \cdot \text{Var}(X)$$

Traduzione:

La probabilità che X si allontani dal suo valore atteso più di d è $\frac{1}{d^2} \text{Var}(X)$
 quindi se d è piccolo questo è grande, altrimenti è piccolo

Corollario

$\text{Var}(x) = 0 \iff X$ costante

dimostrazione

$$(\Rightarrow) \text{Var}(x) = 0 \Rightarrow P\{|X - E[X]| \geq d\} \leq \frac{1}{d^2} \cdot 0 = 0$$

$$(\Leftarrow) X \text{ è costante} \Rightarrow E[X] = \text{cost} \Rightarrow \text{Var}(x) = E[X^2] - E[X]^2 = \text{cost}^2 - \text{cost}^2 = 0$$

VARIANZA E VALORE ATTESO DELLE V.A. NOTEVOLI

B(n, p)

$$n=1 \quad X = \begin{cases} 1 & \text{con prob. } p \\ 0 & \text{con prob. } 1-p \end{cases} \quad (\text{successo/insuccesso})$$

$$- E[X] = \sum a_k \cdot p(X=a_k) = 1 \cdot p + 0 \cdot (1-p) = p$$

$$- E[X^2] = \sum a_k^n \cdot p(X=a_k) = 1^2 \cdot p + 0^2 \cdot (1-p) = p$$

$$- \text{Var}(x) = E[X^2] - E[X]^2 = p - p^2 = p(1-p)$$

$$\text{caso generale} \quad X_i = \begin{cases} 1 & \text{se successo alla } i\text{-esima} \\ 0 & \text{altrimenti} \end{cases} \quad (X = X_1 + X_2 + \dots + X_n)$$

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

$$- E[X] = E[X_1 + X_2 + \dots + X_n] = \sum_i E[X_i] = n \cdot p$$

$$- \text{Var}(x) = E[X^2] - E[X]^2 = np(1-p)$$

$T \sim \text{Geom}(p)$

T = "prova in cui si verifica il 1° successo"

$$\bullet P_T(\underbrace{\{0, \dots, 0\}}_{k-1}, k) = p \cdot (1-p)^{k-1} \quad \forall k \in \mathbb{N}^+$$

$$\bullet E[T] = \frac{1}{p}$$

$$\bullet \text{Var}[T] = \frac{1-p}{p^2}$$

$$\bullet T \geq 0 \Rightarrow \exists E[T]$$

Significato:

Il valore atteso è inversamente proporzionale a p (probabilità di successo). Infatti, se p è grande ci vorranno poche prove per avere un successo, al contrario più è piccolo p più prove ci vorranno.

$X \sim \text{Poiss}(\lambda)$ "utenti che richiedono il servizio in un certo intervallo di tempo"

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

$$\cdot E[X] = \sum_k a_k p(x=a_k) = \sum_k k \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} \sum_k k \cdot \frac{\lambda^k}{k!} = e^{-\lambda} \sum_k \frac{\lambda^k}{(k-1)!} = e^{-\lambda} \cdot \lambda \cdot \boxed{\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}} = \lambda$$

$$\Rightarrow \text{Var}(X) = \lambda$$

$X \sim \text{IP}(N, F, n)$ "successi in n estrazioni: in blocco"

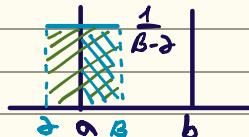
N unità totali, F favorevoli, n unità estratte

$$\cdot P\{X=k\} = \frac{\binom{F}{k} \binom{N-F}{n-k}}{\binom{N}{n}}$$

$$\cdot E[X] = n \cdot \frac{F}{N}$$

$$\cdot \text{Var}(X) = n \cdot \frac{F}{N} \left(1 - \frac{F}{N}\right) \frac{N-n}{N-1}$$

$X \sim U(a, b)$ Si passa alle v.a. con densità



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

N.B. Che gli intervalli siano aperti o chiusi è indifferente perché nel singolo punto è sempre nullo

$$\cdot P\{a \leq X \leq b\} = \frac{\text{lunghezza}(a, b)}{\text{lunghezza}(a, B)}$$

(quindi $\exists E[X], \exists E[X^n] \forall n$)

Essendo f supportata su un intervallo $(a, B) \Rightarrow$ ha momenti di ogni ordine

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

$$\cdot E[X^n] = \int_{-\infty}^{+\infty} x^n f(x) dx = \int_a^B x^n \cdot \frac{1}{B-a} dx = \frac{1}{B-a} \cdot \frac{1}{n+1} \left[x^{n+1} \right]_a^B$$

$$\Rightarrow E[X^2] = \frac{1}{B-a} \cdot \frac{1}{3} \cdot (B^3 - a^3) = \frac{1}{3} (a^2 + aB + B^2)$$

$$\cdot \text{Var}(X) = E[X^2] - E[X]^2 = \frac{1}{3} (a^2 + aB + B^2) - \left(\frac{B+a}{2} \right)^2 = \frac{1}{3} a^3 + \frac{1}{3} aB + \frac{1}{3} B^2 - \frac{B^2}{4} - \frac{a^2}{4} - \frac{Ba}{2} = \frac{(B-a)^2}{12}$$

$X \sim \text{Exp}(\lambda)$

$$f(x) = \lambda e^{-\lambda x} \cdot \mathbb{1}_{(0, \infty)}(x)$$

Essendo $X \geq 0, \exists E[X^n] \in [0, \infty] \rightarrow$ esistono momenti di ogni ordine

$$\cdot E[X] = \int_{-\infty}^{+\infty} x \cdot \lambda e^{-\lambda x} \cdot \mathbb{1}_{(0, \infty)}(x) dx = \int_0^\infty x \cdot \lambda \cdot e^{-\lambda x} dx = -x e^{-\lambda x} \Big|_0^\infty + \int_0^\infty e^{-\lambda x} dx = -\frac{x}{\lambda} e^{-\lambda x} \Big|_0^\infty + \left(-\frac{1}{\lambda} e^{-\lambda x} \right) \Big|_0^\infty = \frac{1}{\lambda}$$

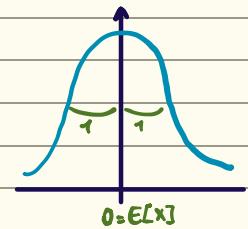
$$\cdot E[X^2] = \frac{2}{\lambda^2}$$

$$\cdot \text{Var}(X) = E[X^2] - E[X]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

$X \sim N(m, \sigma^2)$. Caso gaussiana standard $\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$, $x \in \mathbb{R}$

$$- E[X] = \int_{-\infty}^{+\infty} x \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \left[(-e^{-\frac{x^2}{2}}) \right]_{-\infty}^{\infty} = \frac{1}{\sqrt{2\pi}} \cdot [-0+0] = 0$$

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



\Rightarrow Se n è dispari è una funzione dispari

Se n è pari:

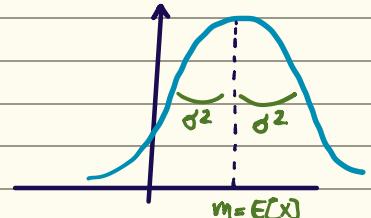
$$\text{Lo } E[X^2] = \int_{-\infty}^{+\infty} x^2 \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \cdot x e^{-\frac{x^2}{2}} dx = \frac{1}{2\pi} \left[\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx - x e^{-\frac{x^2}{2}} \right]_{-\infty}^{\infty} = 1 + 0 = 1$$

$$- \text{Var}(X) = E[X^2] - E[X]^2 = 1$$

. Caso generale $X = \sigma z + m$ dove $Z \sim N(0,1)$

$$- E[X] = E[\sigma z + m] = \sigma E[z] + E[m] = E[m] = m$$

$$- \text{Var}(X) = \sigma^2 \text{Var}(z) = \sigma^2$$

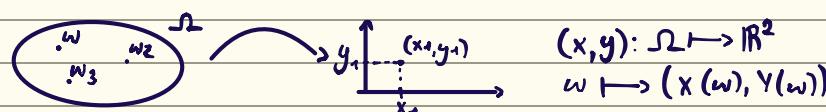


$\Rightarrow m$ e σ^2 sono rispettivamente valore atteso e varianza di $N(m, \sigma^2)$

Quindi si vede che il valore atteso rappresenta il picco, mentre la varianza come si concatenano gli elementi

Variabile casuale doppia (BIVARIATA)

Dato un esperimento casuale con spazio di probabilità (Ω, p) , si tratta di una coppia (X, Y) di variabili casuali, che ad ogni $\omega \in \Omega$ associa una coppia $(x, y) \in \mathbb{R}^2$



es. $\Omega = \text{laureati triennali nel 2023}$

$X(\omega) = \text{"voto di laurea"} \quad Y(\omega) = \text{"mesi intercorsi tra laurea e primo impiego"}$

$$\Rightarrow (x, y)(\omega) = (\text{voto di laurea, mesi tra laurea e 1° impiego})(\omega)$$

Legge congiunta di X e Y

Preso (Ω, p) , $(X, Y) : \Omega \rightarrow \mathbb{R}^2$ v.a. doppia, la legge congiunta di X e Y è la funzione $P_{(X,Y)} : \mathcal{P}(\mathbb{R}^2) \rightarrow \mathbb{R}$ data da $P_{(X,Y)}(A) = P\{(X, Y) \in A\} \doteq P\{(X, Y)^{-1}(A)\} = P\{\omega \in \Omega \mid (X, Y)(\omega) \in A\}$



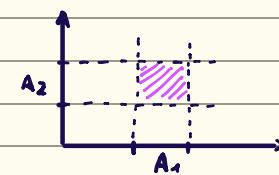
Significato: esprime la probabilità con la quale contemporaneamente entrambe le variabili assumono specifici valori. È quindi indispensabile per l'analisi dei legami e delle forme di dipendenza.

Distribuzioni marginali di una v.a. doppia

Sono definite leggi marginali di (X, Y) le due distribuzioni: P_X e P_Y dove: $P_X(B) = P\{X \in B\} \doteq p(X^{-1}(B))$ e $P_Y(B) = P\{Y \in B\} \doteq p(Y^{-1}(B))$

Nel caso particolare in cui $A \subset \mathbb{R}^2$ sia un prodotto $A = A_1 \times A_2$ si ha che:

$$\begin{aligned} P\{(X, Y) \in A_1 \times A_2\} &= P\{X \in A_1, Y \in A_2\} \\ &= P\{(X \in A_1) \cap (Y \in A_2)\} \\ &= P\left(\text{"si verificano congiuntamente } X \in A_1 \text{ e } Y \in A_2"\right) \end{aligned}$$



es. $X = \text{voto}$ $Y = \text{mesi laurea - I impiego}$
 $P\{(X, Y) \in [105, 110] \times [0, 2]\} = P[\text{"laureato con voto } \in [105, 110] \text{ che trova lavoro tra 0 e 2 mesi dalla laurea}]$

V.A. doppia discreta

Se X e Y sono v.a. discrete (assume un * finito o numerabile di valori)
 \Rightarrow la variabile doppia (X, Y) si dice discreta e assume anch'essa un * finito o numerabile di valori

Funzione di massa congiunta di (X, Y)

Preso $(X, Y): \Omega \rightarrow \mathbb{R}^2$, se (x_i, y_j) si può calcolare $P(x_i, y_j) = P\{X=x_i, Y=y_j\}$.

L'insieme di tutte le coppie (x_i, y_j) al variare di i e j con le corrispondenti prob. costituisce la funzione di massa congiunta di (X, Y) : $P\{(X, Y) \in A\} = \sum_{(x_i, y_j) \in A} P(x_i, y_j)$

es. Il direttore di un magazzino vuole verificare se i clienti dotati di carta di fedeltà sono minori o maggiori rispetto a quelli che non ce l'hanno.

$$X = \text{v.a.} = \begin{cases} 1 & \text{se il cliente ha la carta} \\ 0 & \text{se il cliente non ha la carta} \end{cases}$$

$$Y = \text{v.a. che conta il * acquisti}$$



Il dirigente stima la fun. di prob. congiunta tra X e Y , si riporta di seguito:

X	1	2	3	
0	0.20	0.15	0.05	0.40
1	0.20	0.25	0.15	0.60
	0.40	0.40	0.20	

es. $(X, Y)(w) = (0, 2)$ significa che il cliente w ha effettuato 2 acquisti e non ha la carta fedeltà

La fun. di prob. marginale della X si ottiene dalla $P(x_i, y)$. Precisamente:

$$\begin{aligned} P_X\{X=x_i\} &= P(X^{-1}(x_i)) = P(\bigcup_{j=1}^3 (x_i, y_j)^{-1}) = P((x, y)^{-1}(x_i, 1) \cup (x, y)^{-1}(x_i, 2) \cup (x, y)^{-1}(x_i, 3)) \\ &= P((x, y)^{-1}(x_i, 1)) + P((x, y)^{-1}(x_i, 2)) + P((x, y)^{-1}(x_i, 3)) \\ \implies P_X\{X=0\} &= P((x, y)^{-1}(0, 1)) + P((x, y)^{-1}(0, 2)) + P((x, y)^{-1}(0, 3)) = 0.2 + 0.15 + 0.05 = 0.40 \\ \implies P_X\{X=1\} &= 0.2 + 0.25 + 0.15 = 0.60 \rightarrow \text{prob. che un cliente possieda la carta fedeltà} \end{aligned}$$

Analogamente per P_Y

Osservazioni

1) $P(x_i, y_j) \geq 0 \quad \forall i, j$ dim: $P(x_i, y_j) = P\{X=x_i, Y=y_j\} \geq 0 \rightarrow \text{probabilità}$
 $\sum_i \sum_j P(x_i, y_j) = 1 \quad \sum_i \sum_j P(x_i, y_j) = P\{\bigcup_{i,j} (x_i, y_j)^{-1}\} = P(\Omega) = 1$

2) La funzione di massa congiunta determina le fun. di massa marginali
Cioè, nota $P\{(X, Y) \in A\} \forall A \subset \mathbb{R}^2$, si può determinare $P_X: \mathcal{P}(\mathbb{R}) \rightarrow \mathbb{R}$ e $P_Y: \mathcal{P}(\mathbb{R}) \rightarrow \mathbb{R}$

dimostrazione:

$$\begin{aligned} P_X(x_i) &= P\{X=x_i\} = P(X^{-1}(x_i)) = P\{\bigcup_{j=1}^3 (x_i, y_j)^{-1}\} \stackrel{\text{disgiunt.}}{=} \sum_j P\{(x_i, y_j)^{-1}\} = \sum_j P(x_i, y_j) \\ P_Y(y_j) &= P\{Y=y_j\} = P(Y^{-1}(y_j)) = P\{\bigcup_{i=1}^3 (x_i, y_j)^{-1}\} \stackrel{\text{disgiunt.}}{=} \sum_i P(x_i, y_j) \end{aligned}$$

3) Le marginali, invece, non determinano univocamente la congiunta

dimostrazione:

Consideriamo 2 lanci di dadi, di cui il primo sia equilibrato e il secondo truccato in modo da dare sempre lo stesso risultato del primo

$$X_i = \begin{cases} 1 \leftrightarrow T \text{ dell'i-mo lancio, } i=1,2 \Rightarrow P_{X_i} \sim B(1, 1/2) & i=1,2 \\ 0 \leftrightarrow C \end{cases}$$

Tuttavia, $P_{(X_1, X_2)}$ non è uniforme su $\{0, 1\}^2$. Infatti:

$$P_{(X_1, X_2)}(1, 0) = 0 \quad (P_{(X_1, X_2)}(1, 1) = \frac{1}{2}) \quad P_{(X_1, X_2)}(0, 0) = \frac{1}{2} \quad P_{(X_1, X_2)}(0, 1) = 0$$

V.A. doppia con densità

La v.a. doppia (X, Y) si dice una v.a. doppia con densità $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ (congiunta) se:

$$P_{(X,Y)}(A) = P\{(X, Y) \in A\} = \iint_A f(x, y) dx dy \quad \forall A \subseteq \mathbb{R}^2 \quad (\text{es. } A = [a, b] \times [c, d])$$

V.A. indipendenti

(Ω, \mathcal{P}) spazio di probabilità e $X, Y: \Omega \rightarrow \mathbb{R}$ v.a., allora X e Y si dicono indipendenti se $\forall A \subseteq \mathbb{R}, B \subseteq \mathbb{R}$ gli eventi $\{X \in A\}$ e $\{Y \in B\}$ sono indipendenti; ovvero:

$$P\{X \in A, Y \in B\} = P\{X \in A\} \cdot P\{Y \in B\}$$

Generalizzando, $X_1, \dots, X_n: \Omega \rightarrow \mathbb{R}$ v.a. si dicono indipendenti se $\forall A_i \subseteq \mathbb{R}$ gli eventi $\{X_i \in A_i\}_{i=1, \dots, n}$ sono indipendenti; ovvero $P\{X_1 \in A_1, \dots, X_n \in A_n\} = P\{X_1 \in A_1\} \cdot \dots \cdot P\{X_n \in A_n\}$

es. n lanci di moneta $X_i = \begin{cases} 1 & \text{se + all'i-esimo lancio} \\ 0 & \text{altrimenti} \end{cases} \quad i=1, \dots, n$
 \Rightarrow le X_i sono indipendenti

Criterio di indipendenza per V.A. discrete

X, Y v.a. discrete, allora X, Y sono indipendenti $\Leftrightarrow P_{(X,Y)}(x_i, y_j) = P_x(x_i) \cdot P_y(y_j) \quad \forall (x_i, y_j)$

dimostrazione:

$$(\Rightarrow) X, Y \text{ indipendenti} \Rightarrow P\{X=x_i, Y=y_j\} = P\{X=x_i\} \cdot P\{Y=y_j\} = P_x(x_i) \cdot P_y(y_j)$$

es. 2 lanci di moneta equilibrata $X_i = \begin{cases} 1 & \text{se + all'i-esimo lancio} \\ 0 & \text{altrimenti} \end{cases}$
Allora $P_{(X_1, X_2)}(a, b) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P_{X_1}(a) \cdot P_{X_2}(b) \Rightarrow$ sono indipendenti

es. 2 lanci di moneta, la prima equilibrata, la seconda truccata per dare lo stesso risultato della prima, $X_1 = X_2$ ma X_1 e X_2 non sono indipendenti:
 $P_{(X_1, X_2)}(1, 0) = 0 \neq P_{X_1}(1) = \frac{1}{2} \cdot P_{X_2}(0) = \frac{1}{2}$ non sono indipendenti

Stabilità della indipendenza per composizione

X, Y v.a. indipendenti e $h, k: \mathbb{R} \rightarrow \mathbb{R} \Rightarrow h(x)$ e $k(y)$ sono indipendenti

es. X, Y indipendenti $\Rightarrow X^2, X+Y$ rimangono indipendenti

Proprietà di riproducibilità

Nei 3 casi considero la v.a. univariata $h: \mathbb{R}^2 \rightarrow \mathbb{R} \quad (X, Y) \rightarrow X+Y$

. $X \sim B(m, p)$ e $Y \sim B(n, p)$ indipendenti $\Rightarrow X+Y \sim B(m+n, p)$

. $X \sim P(\lambda)$ e $Y \sim P(\mu)$ indipendenti $\Rightarrow X+Y \sim P(\lambda+\mu)$

. $X \sim N(m_1, \sigma_1^2)$ e $Y \sim N(m_2, \sigma_2^2)$ indipendenti $\Rightarrow X+Y \sim N(m_1+m_2, \sigma_1^2 + \sigma_2^2)$

es. ~~errori~~ errori commessi da un server in un giorno $\sim P(2.5)$

~~errori~~ errori in un dato giorno indipendente dai ~~errori~~ errori in altri giorni:
Probabilità di avere almeno 3 errori in 2 giorni consecutivi?

$$\begin{aligned} X &= \text{\del{errori}} \text{ errori al giorno 1} \sim P(2.5) \\ Y &= \text{\del{errori}} \text{ errori al giorno 2} \sim P(2.5) \end{aligned} \quad \left. \begin{array}{l} \text{indipendenti} \end{array} \right\}$$

La v.a. $X+Y$ conta gli errori totali nei 2 giorni consecutivi, quindi $P\{X+Y \geq 3\}?$

X, Y indipendenti $\Rightarrow X+Y \sim P(2.5+2.5) = P(5)$

$$\text{Quindi } P\{X+Y \geq 3\} = 1 - P\{X+Y < 3\} = 1 - P\{X+Y=0\} - P\{X+Y=1\} - P\{X+Y=2\}$$

$$P(X) \text{ è la v.a. con funzione di probabilità } P\{X=x\} = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$$

$$\Rightarrow P\{X+Y \geq 3\} = 1 - \frac{e^{-5} \cdot 5^0}{0!} - \frac{e^{-5} \cdot 5^1}{1!} - \frac{e^{-5} \cdot 5^2}{2!} = 1 - e^{-5} - 5e^{-5} - \frac{25}{2}e^{-5} \approx 0.8755$$

N.B. è indifferente che siano 2 giorni consecutivi, con 2 giorni a caso era uguale

Dimostrazione (che non si applica in caso di eventi non indipendenti):
 $X, Y \sim B(1, p)$ non indipendenti, considero $X = Y = 2X$

Allora $P_{2X}(2) = P$, $P_{2X}(0) = 1 - P$

$$x = \begin{cases} 1 & \text{con prob. } p \\ 0 & \text{con prob. } 1-p \end{cases}$$

$$2x = \begin{cases} 2 & \text{con prob. } p \\ 0 & \text{con prob. } 1-p \end{cases}$$

Invece $B(2, p)$ avrebbe la seguente distribuzione:

$$P_{B(2,p)}(x) = \binom{2}{x} \cdot p^x (1-p)^{2-x} \Rightarrow P_{B(2,p)}(2) = \binom{2}{2} p^2 (1-p)^0 = p^2$$

$$\Rightarrow P_{B(2,p)}(0) = \binom{2}{0} p^0 (1-p)^2 = (1-p)^2 \neq P_{2X}$$

Valore atteso del prodotto di V.A. indipen.

Siano $X, Y: \Omega \rightarrow \mathbb{R}$ v.a. e sia $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ con $g(x, y) = g \circ (x, y): \Omega \rightarrow \mathbb{R}$. Allora $g(X, Y)$ è una v.a. univariata per cui è possibile provare a calcolare il valore atteso.
Se X, Y sono discrete, $E[g(x, y)] = \sum_i \sum_j g(x_i, y_j) \cdot P(X=x_i, Y=y_j)$

Proposizione: Siano X, Y v.a. con valore atteso e indipendenti.
Allora $X \cdot Y$ ammette valore atteso e $E[X \cdot Y] = E[X] \cdot E[Y]$

dimostrazione:

Considero X, Y discrete, $E[X \cdot Y] = \sum_{i,j} x_i y_j \cdot P\{X=x_i, Y=y_j\} = \sum_{i,j} x_i y_j \cdot P_X(x_i) \cdot P_Y(y_j)$
 $= \sum_i x_i \cdot P_X(x_i) \cdot \sum_j y_j \cdot P_Y(y_j) = E[X] \cdot E[Y]$

es. 2 lanci di dado equilibrato, $X = \text{"esito 1 lancio"}$, $Y = \text{"esito 2 lancio"}$
 $E[X] = E[Y] = 7/2$ $E[X \cdot Y] = E[X] \cdot E[Y] = (7/2)^2$

Reminder

Per campione di dati bivariati (x_i, y_j) avevamo definito:

$$\text{cov}(x, y) = \frac{1}{n-1} \cdot \sum_{i,j} (x_i - \bar{x})(y_j - \bar{y})$$

$$r(x, y) = \text{coefficiente di correlazione} = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x) \cdot \text{var}(y)}}$$

Covarianza di 2 V.A. con momento secondo: Date X, Y v.a. con momento secondo, si definisce covarianza di X e Y il numero $\text{cov}(x, y) = E[(x - E[x])(y - E[y])]$

dimostrazione:

$$E[XY - XE[Y] - YE[X] + EXE[Y]] = E[XY] - E[Y]E[X]$$

Coefficiente di correlazione: Date X, Y v.a. con momento secondo e con $\text{var}(x) \neq 0$ e $\text{var}(y) \neq 0$ è il numero $r = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x)\text{var}(y)}}$

- Variabili Scorrrelate**
- Se $\text{cov}(x, y) = 0$ (equivalente a $\text{cov}(x, y) = 0$) $\Rightarrow X, Y$ sono scorrelate
 - Inoltre, se X, Y con momento secondo e indipendenti $\Rightarrow X, Y$ scorrelate

dimostrazione: X, Y indipendenti $\Rightarrow \forall A, B \subset \mathbb{R} \quad P\{X \in A, Y \in B\} = P\{X \in A\} \cdot P\{Y \in B\}$
 $\Rightarrow \text{cov}(x, y) = E[XY] - E[X] \cdot E[Y] = 0$

Scorrelazione VS. Indipendenza

Nonostante indipendenti \Rightarrow scorrelate, il viceversa non è valido.

Quindi: scorrelate $\not\Rightarrow$ indipendenti

Proprietà della covarianza bilineare e simmetrica

X, Y e v.a. con momento secondo e $a, b \in \mathbb{R}$. Allora:

- 1) $\text{cov}(aX + bY + c, Z) = a \cdot \text{cov}(X, Z) + b \cdot \text{cov}(Y, Z)$ bilineare
- 2) $\text{cov}(X, Y) = \text{cov}(Y, X)$ e $\text{cov}(X, X) = \text{var}(X)$ simmetrica

Varianza della somma

X, Y v.a. Allora $\text{var}(X+Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y)$

In particolare, se X e Y sono scorrelate $\Leftrightarrow \text{var}(X+Y) = \text{var}(X) + \text{var}(Y)$

dimostrazione: $\text{var}(X+Y) = \text{cov}(X+Y, X+Y) = \text{cov}(X, X) + \text{cov}(X, Y) + \text{cov}(Y, X) + \text{cov}(Y, Y)$
 $= \text{var}(X) + \text{cov}(X, Y) + \text{cov}(Y, X) + \text{var}(Y)$ (i numeri sono le proprietà di cui sopra)

es. Siano $X_i = \begin{cases} 1 & \text{con prob. } 1/2 \\ -1 & \text{con prob. } 1/2 \end{cases}$ v.a. indipendenti (\Rightarrow non correlate)

e definiamo $W_n = \sum_{i=1}^n X_i$

$$\text{Allora: } E[W_n] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n (1 \cdot \frac{1}{2} - 1 \cdot \frac{1}{2}) = 0 \quad \mu = 0$$

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

Teorema

X e Y con momento secondo e $\text{Var}(X) \neq 0, \text{Var}(Y) \neq 0$. Allora:

a) $|g(x, Y)| \leq 1$

b) $\min_{a, b \in \mathbb{R}} E[(Y - a - bX)^2] = \text{Var}(Y) \cdot (1 - g(x, Y)^2)$

Traduzione:

- Presi a, b che realizzano il minimo in b), la retta $Y = a + bX$ è la migliore approssimazione lineare tra X e Y
- Il valore del minimo è proporzionale a $1 - P(X, Y)^2$
 \Rightarrow la relazione tra X e Y è tanto meglio approssimata da una retta quanto più $g(x, Y)$ è vicino a ± 1
 $\Rightarrow g$ è una misura della DIPENDENZA LINEARE tra X e Y
 - $g \approx \pm 1$ alta dipendenza lineare
 - $g \approx 0$ bassa dipendenza lineare

Se X, Y sono due caratteri di n individui di una popolazione, $g(x, Y)$ è il coefficiente di correlazione lineare campionario tra i caratteri di TUTTA la popolazione (e non solo di un campione)

V.A. Indipend. ed egualmente distribuite

X_1, \dots, X_n, \dots (anche in $\# \infty$) di v.a. si dicono indipendenti e identicamente distribuite (i.i.d.) se:

- sono indipendenti
- hanno la stessa distribuzione (P_{X_i} è lo stesso P_i)

es. Consideriamo una popolazione, e sia X la v.a. che rappresenta un certo carattere della popolazione.

Estraggo n individui in modo indipendente e chiamo X_i il carattere dell'i-esimo $\Rightarrow X_i$ sono i.i.d.

Media aritmetica X_1, \dots, X_n v.a. la media è $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$

Convergenza

$\{X_i\}_{i \in \mathbb{N}}$ converge in probabilità ad X v.a. se $\forall \epsilon > 0 \lim_{i \rightarrow \infty} P\{|X_i - X| > \epsilon\} = 0 \quad \forall \epsilon > 0$

Trad se n è grande X_n con grande probabilità è vicino a X

Legge dei grandi numeri (debole)

X_1, \dots, X_n successione di v.a. i.i.d. con momento secondo.

Allora \bar{X}_n converge in probabilità a $E[X_i] = \mu$.

$$\text{Dove} \quad \forall \varepsilon > 0 \lim_{n \rightarrow \infty} P\{\|\bar{X}_n - \mu\| > \varepsilon\} = 0$$

Tra la media campionaria, per n grande, è vicina al valore atteso

es. $n = 1000$ lanci di un dado, X_i = esito dell'i-esimo lancio
 $\bar{X}_{1000} \approx \mu = 3.5 \quad E[X_i] = 3.5$

es 2 $X_i = \begin{cases} 1 & \text{successo all'i-me prova } p \\ 0 & \text{insuccesso all'i-me prova } (1-p) \end{cases}$ i.i.d. $\sim B(1, p)$

Allora $\bar{X}_n = \frac{X_1 + \dots + X_n}{n} = \frac{\# \text{successi}}{\# \text{prove}} = \text{frequenza relativa dell'evento successo in } n \text{ prove}$

$\Rightarrow \bar{X}_n$, secondo LGSN, converge in probabilità $n \rightarrow \infty$ a $E[X_i] = p$

$\Rightarrow \bar{X}_{1000} \approx 1/2$ lancio moneta $\bar{X}_{1000} \approx 1/6$ lancio dado

$$\downarrow S^2$$

X_1, \dots, X_n v.a. $S^2 = S_n^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (X_i - \bar{X}_n)^2$

Proprietà

X_1, \dots, X_n, \dots successione di v.a. i.i.d. con momento secondo, $\sigma^2 = \text{Var}(X_i)$.
Allora $S_n^2 \xrightarrow[\text{probabilità}]{\text{converge in}} \sigma^2$

Convergenza in legge

Sia $\{Y_n\}_n$ una successione di v.a. e Y una v.a.. Supponiamo che F_n, F siano le funzioni di ripartizione rispettivamente di Y_n, Y .
Supponiamo che F sia continua.

Allora diciamo che Y_n converge in legge (o distribuzione) a $Y \quad [Y_n \xrightarrow{n \rightarrow \infty} Y]$
Se $\lim_{n \rightarrow \infty} F_n(t) = F(t) \quad \forall t \in \mathbb{R}$

Teorema centrale del limite (TCL/Tc)

Sia X_1, X_2, \dots una successione di v.a. i.i.d. con momento secondo finito e siano $\mu = E[X_i]$ e $\sigma^2 = \text{Var}(X_i)$, con $\sigma > 0$ (cioè X_i non costanti). Sia inoltre $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

Allora $\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{n \rightarrow \infty} Z \sim N(0, 1)$

Cioè $\forall -\infty < a < b < +\infty \quad \lim_{n \rightarrow \infty} P\{\alpha < \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} < b\} = P\{a < Z < b\} = \Phi(b) - \Phi(a)$

Dimostrazione:

X_1, \dots, X_n v.a. i.i.d., la media è $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

So anche che $E[\bar{X}_n] = E[\frac{X_1 + \dots + X_n}{n}] = \frac{1}{n} \cdot E[X_1 + \dots + X_n] = \frac{1}{n} (E[X_1] + \dots + E[X_n]) = \frac{n\mu}{n} = \mu$

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

Standardizzazione della v.a. $\bar{X}_n \xrightarrow{\text{standardizzazione}} \bar{Z}_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}$

che secondo TCL la standardizzazione tende a $N(0, 1)$

Caso particolare \downarrow

Esperimento dicotomico: $X = \begin{cases} 1 & \text{successo all'i-esima prova} \\ 0 & \text{altrimenti} \end{cases} \sim B(1, p)$

Approssimazione normale della binomiale

X_1, \dots, X_n v.a. i.i.d. $\sum_{i=1}^n X_i \sim B(n, p)$

$\bar{X}_n = \frac{\# \text{successi}}{\# \text{prove}}$ $n \bar{X}_n = Y_n = \# \text{successi su } n \text{ ripetizioni} \sim B(n, p)$

$$\bar{E}[Y_n] = n E[\bar{X}_n] = n \cdot p = n \cdot \mu$$
$$\text{Var}(Y_n) = n^2 \cdot p(1-p) = n \cdot np(1-p) = n \sigma^2$$

Il TCL mi dice che $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1)$, dai seguenti calcoli:

$$\frac{n \bar{X}_n - n \mu}{n \sigma / \sqrt{n}} = \frac{Y_n - n \mu}{\sqrt{n} \sigma} = \frac{Y_n - n p}{\sqrt{n} \sqrt{np(1-p)}} = \frac{Y_n - np}{\sqrt{np(1-p)}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1)$$

Regole empirica $np(1-p) \geq 10 \rightarrow n$ abbastanza grande *

es. 1'000 lanci moneta equilibrata,

a) probabilità di avere almeno 480 teste?

$Y = \# \text{teste in 1'000 lanci} \sim B(1'000, 1/2)$

$$\frac{Y - np}{\sqrt{np(1-p)}} = \frac{Y - 1'000 \cdot 1/2}{\sqrt{1'000 \cdot 1/2 \cdot 1/2}} = \frac{Y - 500}{\sqrt{250}} \xrightarrow{TCL} Z = N(0, 1)$$

$$P\{Y \geq 480\} = P\left\{\frac{Y - 500}{\sqrt{250}} \geq \frac{480 - 500}{\sqrt{250}}\right\} = P\left(\frac{Y - 500}{\sqrt{250}} \geq -1,26\right) \xrightarrow{\uparrow} P(Z \geq -1,26) =$$
$$= 1 - \Phi(-1,26) = \Phi(1,26) \approx 0,896 \text{ ALTA}$$

b) qual è il k tale che la prob che ci siano almeno k teste in 1'000 lanci $\approx 0,95$?

So già che $k < 480$ perché $0,95 > 0,896$ (risultato punto a)

$$\text{Cerco } k: P(Y \geq k) = 0,95 \Rightarrow P\left(Z \geq \frac{k - 500}{\sqrt{250}}\right) = 1 - \Phi\left(\frac{k - 500}{\sqrt{250}}\right)$$

Cerchiamo quindi k t.c. $\Phi\left(\frac{k - 500}{\sqrt{250}}\right) = 0,05$

$$\text{Quindi } \frac{k - 500}{\sqrt{250}} = 0,05 = -0,95 = -1,64$$

$$\Rightarrow \frac{k - 500}{\sqrt{250}} = -1,64 \Rightarrow k = 500 - 1,64 \cdot \sqrt{250} \Rightarrow k = 473,91 \Rightarrow k = 473$$

EQUAZIONE X TROVARE K

Si arrotonda a difetto x garantire che prob sia almeno 0,95

STATISTICA INFERENZIALE

Scopo

Ottenere info sulla distribuzione di una popolazione studiando dati rilevati su un campione.

Statistica inferenziale

La popolazione non è nota, prendo un campione e tramite l'analisi dico qualcosa sulla popolazione in modo INDUTTIVO.

Ipotesi chiave

- a) ogni nuova estrazione è indipendente dalle precedenti \rightarrow indipendenza
- b) l'estrazione di ogni individuo che andrà a costituire il campione è fatto in modo identico al precedente \rightarrow equidistribuzione

Situazione

(Ω, P) spazio di probabilità su cui sono definite tutte le v.a. che considereremo.

- Il carattere che vogliamo studiare è rappresentato da una v.a. $X: \Omega \rightarrow \mathbb{R}$ con distribuzione P_X sconosciuta (in parte o totalmente);
- per determinare P_X , suppongo di avere da X_1, \dots, X_n v.a. i.i.d. che rappresentano il campione estratto.

Campione statistico

Si intende di togliere n per X il dato di X_1, \dots, X_n con X_i v.a. i.i.d. con legge P_X .

es. altezza $\sim N(m, \sigma^2)$ mi aspetto che sia distribuita così ma non sono conosciuti i parametri. Cerco delle stime sui parametri tramite la STATISTICA PARAMETRICA.

Statistica parametrica

Scopo:

- stima parametrica
- intervalli di fiducia
- test statistici

Notazione:

- β : indica un parametro non noto della popolazione;
- $B \subseteq H$: H indica l'insieme dei valori ammissibili;
- P_X^β : la β indica che la prob. dipende da un parametro.

Statistiche campionarie

Funzione $g(X_1, \dots, X_n)$ del campione (X_1, \dots, X_n)

Stimatori di un parametro β

Una statistica campionaria $P(X_1, \dots, X_n)$ atte a stimare β

es. Per stimare l'altezza di tutti gli iscritti alla triennale di informatica è $\bar{X}_{100} = \frac{X_1 + \dots + X_{100}}{100}$

Stimatore corretto (non distorto)



Proposizione

X_1, \dots, X_n campione statistico con momento secondo, allora:

- ① \bar{X}_n è uno stimatore corretto per $E[X_i] = E[X]$;
- ② S_n^2 è uno stimatore corretto per $\text{Var}(X_i) = \text{Var}(X)$

Dimostrazione:

① \bar{X}_n stimatore corretto per $E[X_i] \Leftrightarrow E[\bar{X}_n] = E[X]$
 $E\left[\frac{X_1 + \dots + X_n}{n}\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \frac{1}{n} \cdot n E[X] = E[X]$

② S_n^2 stimatore corretto per $\text{Var}(X) \Leftrightarrow E[S_n^2] = \text{Var}(X)$
DA FARRE DIH.

Stimatore Consistente

Parametro β della distribuzione, X_1, \dots, X_n v.a. i.i.d. di X
 $g_n(X_1, \dots, X_n)$ è uno stimatore consistente di β se per $n \rightarrow \infty$ vale $g_n(X_1, \dots, X_n) \xrightarrow{\text{probabilità}} \beta$

es. X_1, \dots, X_n campione di v.a. i.i.d. con momento secondo
Allora \bar{X}_n è uno stimatore consistente per $E[X_i]$ ($= E[X] = \mu$)
e S_n^2 è uno stimatore consistente per $\text{Var}(X_i)$

Dim. si dimostra con LGN

Efficienza di uno stimatore

Presi $g(X_1, \dots, X_n)$ e $h(X_1, \dots, X_n)$ stimatori corretti del parametro β dico che g è più efficiente di h se $\text{Var}(g(X_1, \dots, X_n)) \leq \text{Var}(h(X_1, \dots, X_n))$

es. X_1, \dots, X_n v.a. i.i.d. $\Rightarrow \bar{X}_n$ è tanto più efficiente quanto più n aumenta

Dim. $\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n^2} \cdot (\text{Var}(X_1) + \dots + \text{Var}(X_n)) = \frac{1}{n^2} \cdot n \text{Var}(X_i) = \frac{\text{Var}(X_i)}{n}$

quindi più cresce n e più diventa piccolo il risultato, si disperde meno

Metodo della massima verosimiglianza

Metodo per scegliere un buon stimatore. Ho (X_1, \dots, X_n) campione i.i.d di X , X con distribuzione P_x^β con β parametro.

Le v.a. possono essere:

- discrete con funzione di massa $P_\beta(x)$
- con densità $f_\beta(x)$

Funzione di verosimiglianza

$L_\beta : \mathbb{R}^n \rightarrow \mathbb{R}$ definita come:

- caso discreto: $P_\beta(X_1, \dots, X_n) = P\{X_1 = x_1\} \cdot \dots \cdot P\{X_n = x_n\} = P\{X_1 = x_1, \dots, X_n = x_n\}$
- caso con densità: $f_\beta(x_1) \cdot \dots \cdot f_\beta(x_n)$

Stima massima verosimiglianza (MLE)

Si chiama MLE se \exists una statistica campionaria $\hat{\beta} = \hat{\beta}(X_1, \dots, X_n)$ tale che valga l'uguaglianza $L_{\hat{\beta}}(X_1, \dots, X_n) = \max_{\beta \in \mathbb{R}^n} L_p(X_1, \dots, X_n) \quad \forall (X_1, \dots, X_n) \in \mathbb{R}^n$

Il massimo si trova tramite la derivata della funzione
Ho $y = L(\beta)$ e cerco $\frac{dL}{d\beta} = (\hat{\beta})$

Alle volte non basta una stima ma servono degli intervalli di fiducia

Intervalli di fiducia

(X_1, \dots, X_n) campione statistico di legge P_θ , $\theta \in H \subset \mathbb{R}$
 $\sigma \in (0, 1)$ e $a, b: \mathbb{R}^n \rightarrow \mathbb{R}$ funzioni t.c. $a(X_1, \dots, X_n), b(X_1, \dots, X_n)$ siano v.a.

Allora un intervallo aleatorio $I[a(X_1, \dots, X_n), b(X_1, \dots, X_n)]$ si dice intervallo di fiducia per θ di livello $1-\sigma$ se

$$\forall \theta \in H \quad P_\theta(\theta \in I) = P(a(X_1, \dots, X_n) \leq \theta, b(X_1, \dots, X_n) \geq \theta) \geq 1-\sigma$$

