STATISTICA

STATISTICA DESCRITTIVA

PROBABILITÀ E INDIPENDENZA

EVENTI INDIPENDENTI

VARIABILI ALEATORIE

STATISTICA DESCRITTIVA NELLA PROBABILITÀ

STATISTICA INFERENZIALE

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 stessa proced
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 rappresentato (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 dato (1, 2, 3, 4, 5, 6)

Continuo: Assume una grande quantità di valori
es. altezza degli italiani

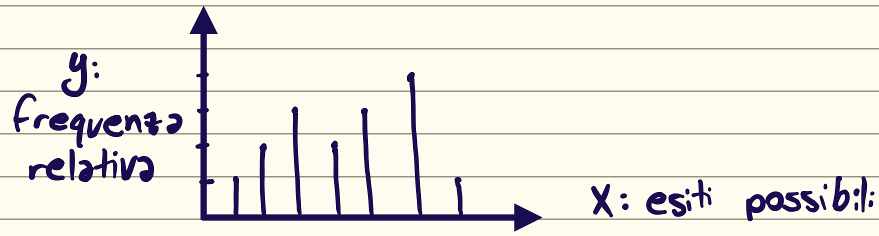
FREQUENZA

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
<u>Frequenza assoluta</u> di T = 6
<u>Frequenza relativa</u> di T = 6/8 = 75%

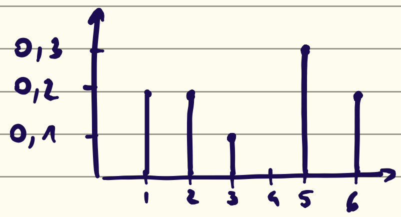
DIAGRAMMA A BARRE



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 può prendere il carattere (1, 2, ..., 6)



INDICI

Indice statistico Quantitativo numerico che riassume qualche proprietà significativa di una distribuzione di dati
$\Longrightarrow$ 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 $\Longrightarrow$ mediana $= x_{\left(\frac{n+1}{2}\right)}$
   - Se (n) è pari $\Longrightarrow$ mediana $= \dfrac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$

es. $x = (8, 3, 5, 12, 10) \in \mathbb{R}^5$
$\Longrightarrow$ media: $\bar{x} = (8+3+5+12+10)/5 = 38/5 = 7{,}6$
$\Longrightarrow$ mediana: ordino $x = (3, 5, 8, 10, 12) \Longrightarrow n = 5$ è dispari $\Longrightarrow \frac{5+1}{2} = 3$ $\Rightarrow x_{(3)} = 8$

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

Media Sfrondata

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, mediano, moda)

Varianza campionaria

Si calcola quando si ha un campione della popolazione

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

Varianza empirica

Si calcola quando si ha l'intera popolazione

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

Rapporto tra $var(x)$ e $var_e(x)$

$$var(x) = \frac{n}{n-1} \left(var_e(x) \right)$$

Esempio varianza campionaria ed empirica

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

$$var_e(x) = 1^2 \cdot \frac{2}{10} + 2^2 \cdot \frac{2}{10} + 3^2 \cdot \frac{1}{10} + 4^2 \cdot \frac{0}{10} + 5^2 \cdot \frac{3}{10} + 6^2 \cdot \frac{1}{10} - (3{,}6)^2$$

$$= 16{,}6 - 12{,}96 = 3{,}64$$

$$var(x) = \frac{10}{9} \cdot var_e(x) = \frac{10}{9} \cdot 3{,}64 \cong 4.04$$

Deviazione Standard

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

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

2) Empirico: $\sigma_e(x) = \sqrt{var_e(x)}$

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 $x_i \gg \bar{x}$
- Esalta con segno negativo gli $x_i \ll \bar{x}$

Funzione di ripartizione empirica (c.d.f)

Restituisce il numero di dati presenti sull'asse

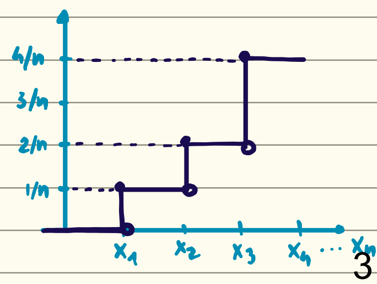
$F_e : \mathbb{R} \to \mathbb{R} \implies F_e(t) = \frac{\#\{i \mid x_i \leq t\}}{n} \longrightarrow$ f. relativa dei valori $\leq t$

- Si parte da $(x_1, \dots, x_n)$ e si ordinano i dati
- Per $t < x_{(1)}$, $F_e(t) = 0$

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

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

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



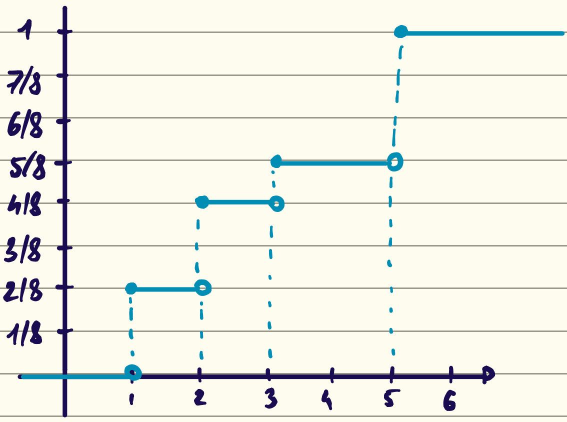
Esempio Funzione di ripartizione empirica

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 $t<1$ $\quad F_e(t)=0$, $\quad F_e(1)=2/8$ $\quad \}$ salto di $1/8 \cdot 2$
- Per $1 \le t < 2$ $\quad F_e(t)=2/8$, $\quad F_e(2)=4/8$ $\quad \}$ salto di $1/8 \cdot 2$
- Per $2 \le t < 3$ $\quad F_e(t)=4/8$, $\quad F_e(3)=5/8$ $\quad \}$ salto di $1/8$
- Per $3 \le t < 5$ $\quad F_e(t)=5/8$, $\quad F_e(5)=8/8=1$ $\quad \}$ salto di $1/8 \cdot 3$

salta il 4 perchè non c'è



- Per $5 \le t$ $\quad F_e(t)=1$

K-simo percentile o B-quantile

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

- almeno il $K\%$ dei dati è $\le x_i$
- almeno il $(1-K)\%$ dei dati è $\ge x_i$

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

Notazione Percentili e Quantili

$K=25 \Rightarrow 25°$ percentile $\longleftrightarrow B=25/100 \longrightarrow$ I Quartile
$K=50 \Rightarrow 50°$ percentile $\longleftrightarrow B=50/100 \longrightarrow$ mediana $\longrightarrow$ II Quartile
$K=75 \Rightarrow 75°$ percentile $\longleftrightarrow B=75/100 \longrightarrow$ III Quartile

Calcolare B-quantile

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

- <u>NON INTERO</u>: arrotondo al numero successivo $\Rightarrow X_{\lceil Bn \rceil}$
- <u>INTERO</u>: prendo la media tra X_{Bn} e X_{Bn+1}

Esempio B-quantili

Considero la tabella dei voti degli esami di 14 studenti ($n=14$)

A	25
B	23
C	25
D	25
E	28
F	28
G	26
H	25
I	26
L	22
M	2
N	22 6
O	20
P	27

Calcola I, II e III quartile

- ordino i dati:

20 22 22 **23** 25 25 **25 25** 26 26 **26** 27 28 28

I quartile (23), II quartile ($25\ 25$), III quartile (26)

I quartile: $B \cdot n = \frac{25}{100} \cdot 14 = \frac{350}{100} = [3,5] = 4 \Rightarrow X_4 = 23$

II quartile: $\frac{50}{100} \cdot 14 = 7 \Rightarrow \frac{X_7 + X_8}{2} = 25$

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

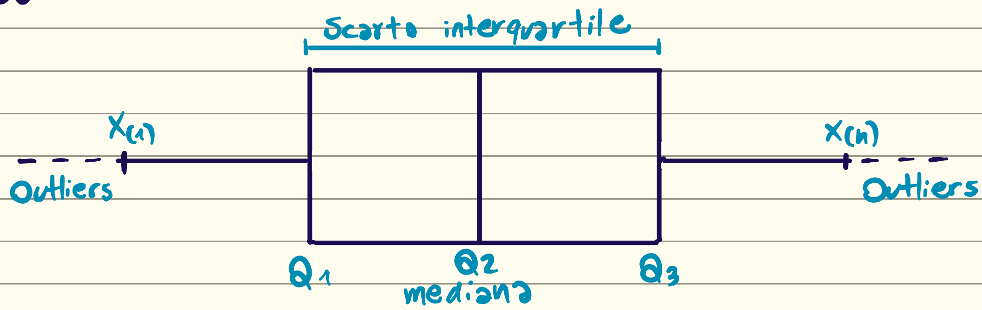
Box Plot

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

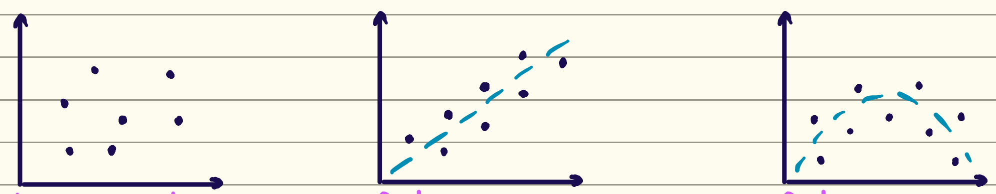
K-uple di vettori

Capita di avere più dati misurati per lo stesso campione. In questo caso i dati sono K-uple di vettori $(x, y, z, \dots) \in \mathbb{R}^{k \cdot n}$

es. (x_i, y_i) = altezza e peso iscritti informatica

Diagramma a dispersione

Rappresenta con un punto ciascuna coppia di dati



Nessuna relazione — Relazione Lineare $y_i \approx a x_i + b$ — Relazione non lineare

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

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

oss: $cov(x,y) = cov(y,x)$

Covarianza empirica

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

Relazione tra $cov(x,y)$ e $cov_e(x,y)$

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

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

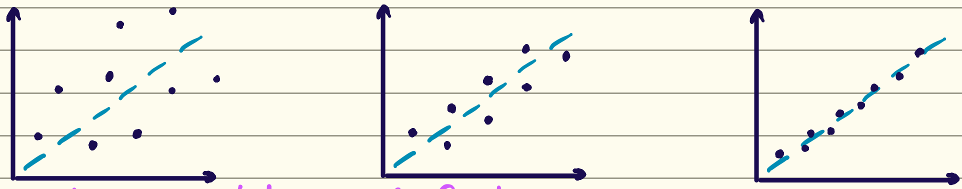
RECAP: $\sigma(x)$ = deviazione standard

Coefficiente di correlazione

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

OSS: coefficiente <u>non</u> cambia sostituendo con $cov_e(\cdots)$ e $\sigma_e(\cdots)$

La correlazione può essere:
- Debole: r vicino a 0
- Forte: r vicino a -1, oppure vicino a 1
- Perfetta: $r=1$ o $r=-1$



Correlazione debole   C. forte   C. perfetta

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 $\min\limits_{q,m\in\mathbb{R}} \sum\limits_{i=1}^{n} (y_i - q - mx_i)^2$

Teorema

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

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

Inoltre $\min\limits_{q,m\in\mathbb{R}} \sum\limits_{i=1}^{n}(y_i - q - mx_i)^2$ è uguale a $\sum\limits_{i=1}^{n}(y_i - \bar{y})^2 \cdot \left[1 - r(x,y)^2\right]$

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 $\min = 0$, 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 INDIPENDENZA

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

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

TEORIA DEGLI INSIEMI

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

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: esce un numero ≤ 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
0 imposs. 1 certo

Definizione classica (Laplace)
$$P(A) = \frac{\#\text{casi favorevoli di } A}{\#\text{casi possibili}} = \frac{\# A}{\# \Omega}$$

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

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

Definizione Frequentista

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

Presi quindi:
- $w_1, \dots, w_k$ eventi elementari incompatibili non equiprobabili
- $n_1, \dots, n_k$ frequenza assoluta

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

Pro e contro:
- Adatta a popolazioni ideali
- 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.rel $= \frac{529}{1000} = 0{,}529$
altri 1000 lanci, viene testa 493 volte $\Longrightarrow \frac{529+493}{2000} = 0{,}511$
Al tendere all'infinito delle prove dovrei 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) $P(\Omega) = 1$
  3) se $A_1, \dots, A_n$ è una successione di eventi incompatibili $(A_i \cap A_j = \emptyset \ \forall i \neq j)$
     allora $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$ additività
     e $P(\bigcup_{i=1}^{\infty} A_i) = \lim_{n \to \infty} \sum_{i=1}^{n} P(A_i)$ se la sommatoria è infinita

Oss: la probabilità frequentista soddisfa gli assiomi

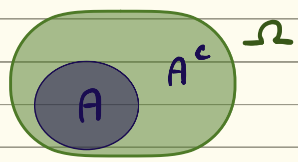
Spazio di probabilità

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

Proprietà di una probabilità

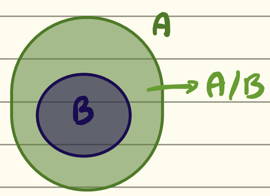
Sia (Ω, P) uno spazio di probabilità $P: \mathcal{P}(\Omega) \to \mathbb{R}$, allora:

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



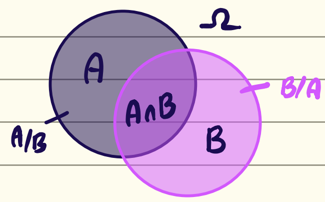
dimostrazione:
$P(A \cup A^c) = P(\Omega) = 1$
$\quad = P(A) + P(A^c)$

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



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:
$P(A \cup B) = P(A \setminus B \cup A \cap B \cup B \setminus A) = P(A) + P(B) - P(A \cap B)$
($A \setminus B \cup A \cap B$ = P(A); $A \cap B \cup B \setminus A$ = P(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 rate né sconto.

$A \cup B$ = "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 $\Longrightarrow P(A) = \frac{\# A}{\# \Omega}$
Inoltre, con un insieme $\{1, \dots, n\}$:

Calcolo combinatorio

- il # di sequenze ordinate con ripetizione di k oggetti è n^k
- il # di sequenze ordinate senza ripetizione di k oggetti è $\frac{n!}{(n-k)!}$
- il # 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 $\Rightarrow \frac{\# A \cap B}{\# B} = \frac{1}{3}$

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

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

Condizionamento ripetuto

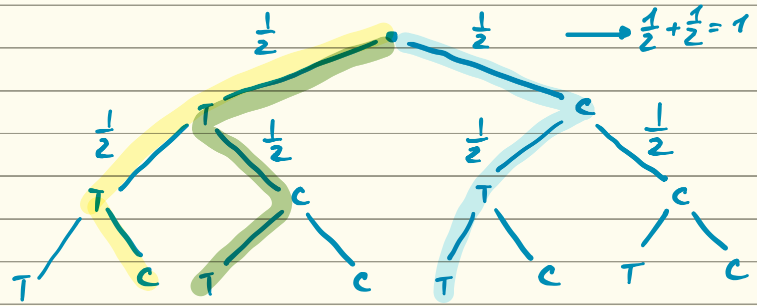
Siano $A_1, \dots, A_n$ eventi in Ω se $A_1 \cap \dots \cap A_n$ non trascurabile, allora:
$P(A_1 \cap \dots A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \dots \cdot P(A_n|A_1 \cap \dots \cap A_{n-1})$

Rappresentazione ad albero

Diagramma in cui:
- $\forall$ stadio (nodo) 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

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



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

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

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

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

Regola della catena

P(ramo) = prodotto delle probabilità dei sottorami

Regola della fattorizzazione

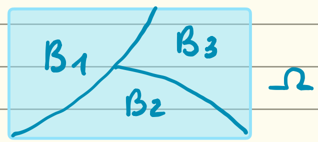
P(unione rami) = somma delle probabilità dei rami

Sistema di alternative per esperimento con spazio campionario Ω

Dato un esperimento con spazio campionario Ω, è il dato di $B_1, \dots, B_n$ eventi tali che:
- $B_i \cap B_j = \emptyset \quad \forall i \neq j$
- $\bigcup_{i=1}^{n} B_i = \Omega$
- ciascun B_i è non trascurabile $(P(B_i) \neq 0)$

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

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

es. $Urna_1 = \begin{cases} 5 \text{ palline rosse} \\ 5 \text{ palline blu} \end{cases}$ $\qquad Urna_2 = \begin{cases} 8 \text{ palline rosse} \\ 2 \text{ palline blu} \end{cases}$

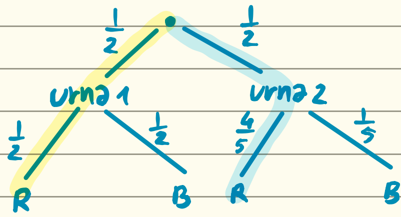
Scegliere a caso una pallina da una delle due urne, qual è la probabilità che la pallina scelta sia rossa?

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

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

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

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



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

$P(U_2 \cup R) = \frac{1}{2} \cdot \frac{4}{5} = \frac{4}{10}$

$\Downarrow$

$P(R) = \frac{1}{4} + \frac{4}{10} = \frac{13}{20}$

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

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

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

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

EVENTI INDIPENDENTI

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

Proposizione

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

Osservazioni

1) Se $P(A) = 0$ oppure $P(A) = 1 \implies$ 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 $P(A) = 1 \implies P(A^c) = 0 \implies A^c$ 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 $A \cap B = \emptyset \implies P(A \cap B) = 0$

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?

$P(A) = \frac{4}{40} = \frac{1}{10}$ $P(B) = \frac{10}{40} = \frac{1}{4}$
$P(A \cap B) =$ "esce l'asso di denari" $= \frac{1}{40} = \frac{1}{10} \cdot \frac{1}{4} \implies$ 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 <u>entrambe</u> le condizioni

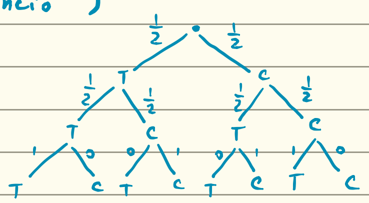
Indipendenza di n eventi

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

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: "T al 1° lancio"
B: "T al 2° lancio"
C: "T al 3° lancio"

Verificare che A, B, C sono 2 a 2 indipendenti ma non sono globalmente indipendenti



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

$P(A \cap B \cap C) = \frac{1}{2} \cdot \frac{1}{2} \cdot 1 = \frac{1}{4} \neq \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}$ No! non globale

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 <u>indipendenti</u>, la probabilità è:

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

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

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

$$P(1111110000) = \left(\tfrac{1}{2}\right)^{6} \cdot \left(\tfrac{1}{2}\right)^{4} = \left(\tfrac{1}{2}\right)^{10}$$

No! L'evento che ci interessa non è 1111110000 ma è l'unione di tutti gli eventi con quel risultato

$$\Downarrow$$

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

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

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

ESERCIZI NELLA LEZIONE 7

VARIABILI ALEATORIE

Aleatorio $\Longrightarrow$ casuale

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

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

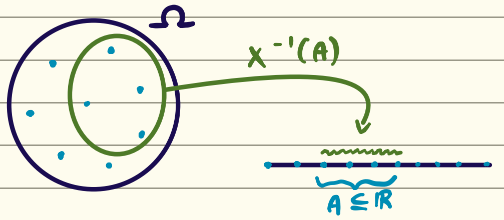
X: "# teste in n lanci" $X: \Omega \to \mathbb{R} \Longrightarrow X(w_1, \dots, w_n) = \sum_{i=1}^{n} w_i$

aggiungo un'altra variabile aleatoria per il numero delle croci y = "# croci in n lanci"
$y: \Omega \to \mathbb{R} \Longrightarrow (w_1, \dots, w_n) = n - \sum w_i \Longrightarrow y = n - X$

Z = "# coppie teste consecutive" idea = $z((\underbrace{1,1}_{1\cdot 1=1 \text{ coppia}}, \overbrace{1, 0}^{1\cdot 0=0 \text{ no coppia}}, 0, 0, 1)) = 1$

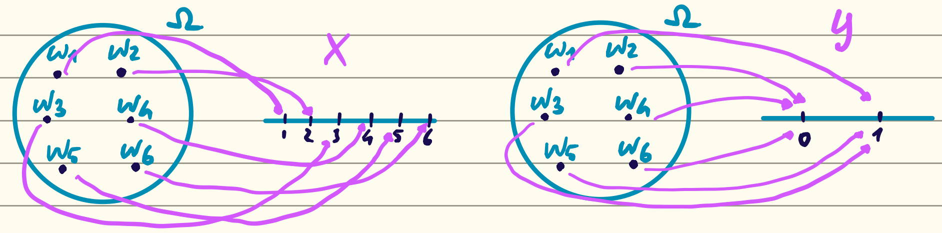
$Z: \Omega \to \mathbb{R}$ $(w_1, \dots, w_n) = w_1 \cdot w_2 + w_2 \cdot w_3 + \dots + w_{n-1} \cdot w_n = \sum_{i=1}^{n-1} w_i \cdot w_{i+1}$

Legge di probab. di una variabile aleatoria: Sia $X: \Omega \to \mathbb{R}$ una variabile aleatoria, definiamo legge (o distribuzione) di X la funzione $P_X: \mathcal{P}(\mathbb{R}) \to \mathbb{R}$ data da:
$P_X(A) = P(X \in A)$ $\forall A \subseteq \mathbb{R}$ 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"

$X: \Omega \to \mathbb{R}$ dove X = variabile che esprime # puntini associati al lancio del dado
$y: \Omega \to \mathbb{R}$ dove $w_i = 1$ se i è dispari, $w_i = 0$ se i è pari

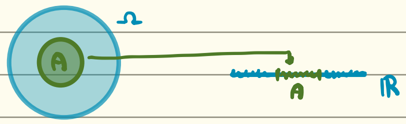


$y^{-1}(0) = \{w_2, w_4, w_6\}$
$y^{-1}(1) = \{w_1, w_3, w_5\}$

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

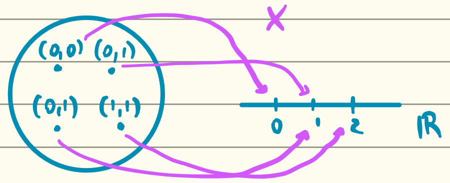
es. $A = \{3,4\} \subseteq \mathbb{R}$ $P(X^{-1}(\{3,4\})) = P(w_3 \cup w_4) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$
es$_2$. $P_y(\{0\}) \doteq P(y^{-1}(0)) = P(w_2 \cup w_4 \cup w_6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$

Proposizione
- X^{-1} commuta con le operazioni insiemistiche
es. $X^{-1}(A \cup B) = X^{-1}(A) \cup X^{-1}(B)$ $X^{-1}(A^C) = (X^{-1}(A))^C$
- $P_X: P(x) \to \mathbb{R}$ è una probabilità su $\mathbb{R}$



es. 2 lanci di moneta equilibrata

$\Omega = \{C,T\}^2 = \{(\omega_1,\omega_2) \mid \omega_1,\omega_2 \in \{C,T\}\} = \{0,1\}^2 \longrightarrow$ 0 = esce croce, 1 = esce testa



X = "# 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 $\quad \Omega = \{C,T\}^2 = \{0,1\}^2$

$X = \text{"\#teste"} = \sum_{i=1}^{2} \omega_i = \omega_1 + \omega_2 \qquad P_X = \mathcal{P}(\mathbb{R}) \to \mathbb{R} \quad A \longmapsto P(X^{-1}(A))$

$Y = \text{"\#croci"} = n - \sum_{i=1}^{2} \omega_i = 2 - X \qquad P_Y = \mathcal{P}(\mathbb{R}) \to \mathbb{R} \quad A \longmapsto P(Y^{-1}(A))$

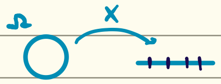
$P_X = P_Y$ quindi le v.a. sono equidistribuite (ma non uguali)

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. 

assume 4 valori es. numero di errori in un testo

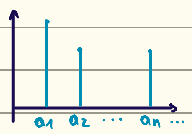
Funzione di massa

Supponiamo di avere una variabile aleatoria discreta $X: \Omega \to \mathbb{R}$, la funzione di massa (o densità discreta) di X è $P_X: \{a_1, \dots, a_n, \dots\} \to \mathbb{R}$ definita come

$P_X(a_i) = P(X = a_i) = P_X\{a_i\} = P(X^{-1}(a_i))$

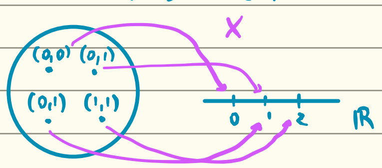
È poi possibile estendere P_X per il numero definito solo sull'insieme discreto dei valori di X a tutto $\mathbb{R}$, ponendo $P_X(a) = 0 \ \forall a \in \{a_1, \dots\}$

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



es. esperimento = 2 lanci di moneta equilibrata

$\Omega = \{C,T\}^2 = \{0,1\}^2 \qquad X = \#$ teste in 2 lanci



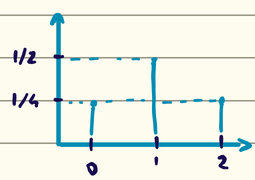
$X: \Omega \to \mathbb{R}$ prende solo i valori 0, 1, 2

La sua funzione di massa è: $P_X: \{0,1,2\} \to \mathbb{R}$

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

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

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



Calcolo delle probabilità di X usando P_X

Presa 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?

$\Omega = \{0,1\}^4 \quad X = \#\text{teste} \quad \Omega \to 0 \; 1 \; 2 \; 3 \; 4$

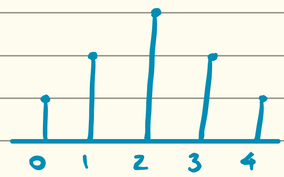
$P_X(0) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{16}$

$P_X(1) =$ TCCC CTCC CCTC CCCT $= \frac{4}{16} = \frac{1}{4}$

$P_X(2) =$ TTCC TCTC TCCT CTCT CTTC CCTT $= \frac{6}{16} = \frac{3}{8}$

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

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



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

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) $P_X(a_i) \doteq P(X=a_i) = P(X^{-1}(a_i)) \geq 0$ P è una probabilità
2) $\sum_i P_X(a_i) = \sum_i P(X=a_i) = P(\bigcup X^{-1}(a_i)) = P(\Omega) = 1$

Proposizione

Se $q: \{a_1, a_2, \dots\} \to \mathbb{R}$ è una funzione che soddisfa le proprietà 1 e 2 (sopra) allora $\exists$ una v.a. X di cui q è la funzione di massa.

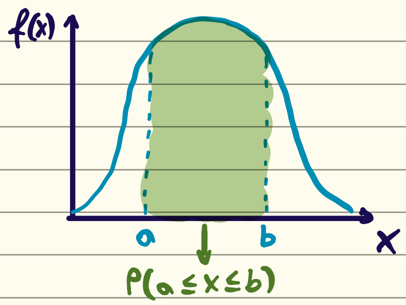
es. ($q = P_X$ per qualche X v.a. discreta)

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 $f: \mathbb{R} \to \mathbb{R}$ 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

$P(X \in A) = P(a \leq x \leq b) = \int_a^b f(x)\,dx$ (area sottesa del grafico di f 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) la misura dell'area sotto al grafico di tutte f è $= 1$

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

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

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

Definisco la v.a. X con densità 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 $b = \frac{1}{4}$ $\quad P(0 \leq x \leq \frac{1}{4}) = \int_0^{1/4} 1\,dx = \frac{1}{4}$

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

($b = 5$: fuori intervallo)

Interpretazioni delle variabili aleatorie

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

- X = carattere quantitativo discreto, considero $X(\omega) =$ "# figli di ω"

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

$P_X(1)$: probabilità che un cittadino italiano abbia un figlio unico

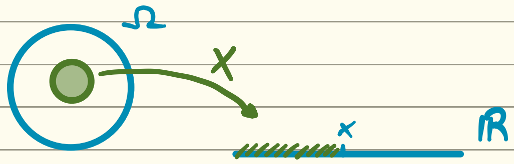
- X = carattere quantitativo con densità, considero $X(\omega)$ = altezza di ω

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

Funzione di ripartizione

Prese X v.a. e $P_X : \mathcal{P}(\mathbb{R}) \to \mathbb{R}$ legge di probabilità, la funzione di ripartizione di X è la funzione $F_X : \mathbb{R} \to [0,1]$ definita come $F_X(x) = P(X \leq x) = P_X((-\infty, x])\ \forall x \in \mathbb{R}$
assegna ad ogni $x \in \mathbb{R}$ la probabilità 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 $(x_1 \leq x_2 \Longrightarrow F(x_1) \leq F(x_2))$

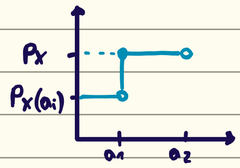
2) $\lim_{x \to -\infty} F_X(x) = 0$; $\lim_{x \to +\infty} F_X(x) = 1$

3) F_X è continua a destra $(x_n \to x \Longrightarrow F(x_n) \to F(x))$

F_X per X v.a. discreta

Se X è una v.a. discreta, allora $F_X(x) \doteq P_X\{X \le x\} = \sum_{a_i \le x} P_X(a_i)$

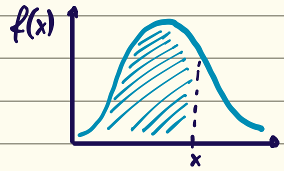
Ne deriva una funzione a gradini:
- ha un salto di ampiezza $P_X(a_i)$ in corrispondenza di ogni valore
- è costante tra due valori consecutivi
- Se X assume un valore finito di valori $a_1, \dots, a_n \Rightarrow F(x) = 1 \ \forall x \ge x_n$



F_X per X v.a. continue

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

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



$F_X(x) =$ l'area

Calcolo della probabilità di intervalli:

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

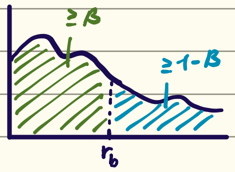
dimostrazione

$F_X(b) - F_X(a) = P\{X \le b\} - P\{X \le a\} = P(\{X \le b\} \setminus \{X \le a\}) = P(a \le x \le b)$

B-quantile

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

Ovvero, il b-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 $\forall \beta \in (0,1)$ $\exists!$ βquantile e $r_\beta = F_X^{-1}(\beta)$

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

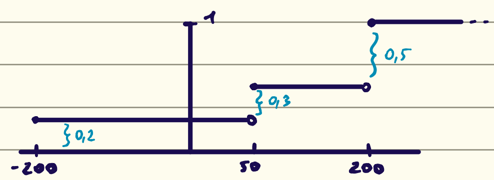
X v.a. "profitto del negoziante" $\begin{cases} 200 & P = 0.5 \\ 50 & P = 0.3 \\ -200 & P = 0.2 \end{cases}$

$P_X(-200, 50, 200) \to \mathbb{R}$

$P_X(-200) = 0{,}2 \quad P_X(50) = 0{,}3 \quad P_X(200) = 0{,}5$

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

$F_X(x) = P\{X \le x\} = P_X(X^{-1}(-\infty, x])$



2) X = "# figli maschi in famiglie con 4 figli"

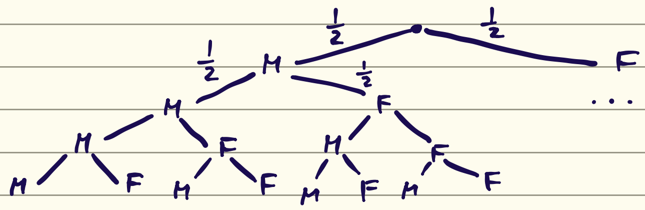
a. costruire una tabella della distribuzione di prob. di X

b. rappresentare graficamente la distribuzione

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

$P_X : \{0,1,2,3,4\} \to \mathbb{R} \quad a_i \longrightarrow P(X=a_i) = P(X^{-1}(a_i))$

Consideriamo M e F equiparabili $\Longrightarrow P(M) = P(F) = \frac{1}{2}$



Tutti i casi hanno probab. $\frac{1}{2}$

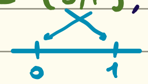
$\Longrightarrow \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{16}$

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 $\Omega = \{S, F\}$, $X : \Omega \to \mathbb{R}$



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 $B(0,p)$

es. $\Omega = \{T, C\}$ $\quad X : \Omega \to \mathbb{R}$
$T \to 1 \quad C \to 0 \quad p = 0.5$
$P_X(1) = 0.5$
$P_X(0) = 1 - 0.5 = 0.5$
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 $\Omega = \{0,1\}^n = \{(\omega_1, \dots, \omega_n) \mid \omega_i \in \{0,1\}\}$ e p probabilità di successo nella singola prova. Definisco X = "# successi in n prove ripetute". La probabilità di avere una sequenza di k successi è $P_X(k) = P\{X=k\} = P(X^{-1}(k))$

$P(X^{-1}(0)) = (1-p)^5$

$P(X^{-1}(5)) = p^5$

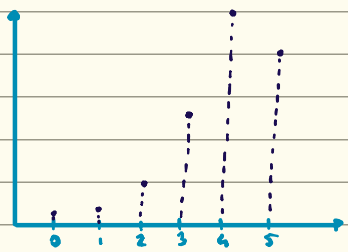
Il caso generale è

$$\underbrace{\sum_{\substack{\text{sequenze} \\ \text{con } k \text{ successi}}}}_{\binom{n}{k}} p^k \cdot (1-p)^{n-k} = P_X(k)$$

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 = "# teste in 5 lanci"
$\Longrightarrow B(5, 0.8)$ 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	$(0.2)^5 = 0{,}0003$
1	$\binom{5}{1} \cdot (0.8)^1 \cdot (0.2)^4 = 0.0006$
2	$\binom{5}{2} (0.8)^2 (0.2)^3 = 0.05$
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	$\binom{5}{5} (0.8)^5 = 0.35$



Esempi di v.a. binomiali

Sono esempi di v.a. binomiali:

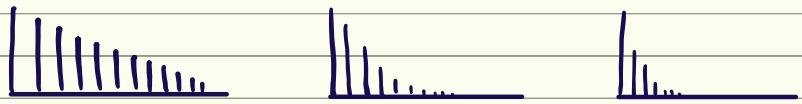
- # teste in n lanci di monete $\longrightarrow B(n, \frac{1}{2})$
- # di "6" in n lanci di un dado $\longrightarrow B(n, \frac{1}{6})$
- # biglie rosse estratte, in n estrazioni con rimpiazzo, da un'urna con biglie rosse e blu

Variabile casuale geometrica

Esperimento con esito dicotomico $\{S, F\}$ ripetuto infinite volte, con probabilità p.
Allora $\Omega = \{0,1\}^{\mathbb{N}^+} = \{(a_1, a_2, \dots) \mid a_i \in \{0,1\} \ \forall i \in \mathbb{N}_+\}$ e $T = \min\{n \in \mathbb{N}_+ \mid \omega_n = 1\}$
ovvero l'istante ($n°$ della prova) in cui si verifica il primo successo. $0, 0, 0, 0, 1, 1, 0, \dots$
$\Longrightarrow$ T è una v.a. discreta a valori in $\mathbb{N}_+$

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

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



$p = 0.1$ $\quad$ $p = 0.3$ $\quad$ $p = 0.5$

All'aumentare di p la funzione probabilità decresce più rapidamente

es. probabilità che un docente che ha ricevuto la copia omaggio adotti il libro è 0.11
determinare la distribuzione del numero di prof a cui dare la copia omaggio, affinché l'ultimo adotti il libro.
X = "# di prof. a cui dare il libro affinché l'ultimo lo adotti" $\sim G(0.11)$

$P_X\{(\underbrace{0,0,0}_{k-1}, \underset{k}{1})\} = p \cdot (1-p)^{k-1} \Longrightarrow P\{X=1\} = p(1-p)^0 = 0{,}11$
$\Longrightarrow P\{X=2\} = p(1-p)^1 = 0{,}11 \cdot 0{,}89 = 0{,}098$
$\Longrightarrow P\{X=3\} = p(1-p)^2 = 0{,}11 \cdot (0{,}89)^2 = 0{,}087$

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 $P(T = n+k \mid T > n) = P(T = k) \quad \forall n, k \in \mathbb{N}_+$
Traduzione: la distribuzione del 1° successo dalla (n+1)esima prova non cambia sapendo che il successo non si è verificato nelle prime n prove

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

Poiss(λ) con $\lambda > 0$ è una v.a. discreta che assume i valori in $\mathbb{N}$.
La funzione di massa è data da $P_X(k) = P\{X = k\} = \dfrac{\lambda^k \cdot e^{-\lambda}}{k!} \quad k \in \mathbb{N}$

- λ rappresenta il # medio di eventi nell'intervallo Poiss(λ) $\sim B(n,p)$
- Poiss(λ) conta # successi in prove ripetute
- Il fattore $e^{-\lambda}$ serve per far sì che la P_X sia una funzione di massa

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

$\displaystyle\sum_{k=0}^{\infty} \frac{1}{k!} = e$

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" $\longleftrightarrow$ SUCCESSO
- N−F "non favorevoli" $\longleftrightarrow$ INSUCCESSO

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

X = "# unità favorevoli nelle n prove"
X = <u>V.a. ipergeometrica</u> di parametri $IP(N, F, n)$

$$P_X(k) = P(X=k) = \frac{\#\text{casi favorevoli}}{\#\text{casi possibili}} = \frac{\#\text{n-uple in cui ho k elementi di F}}{\#\text{n-uple che posso estrarre}}$$

In particolare:
- # casi favorevoli = # n-uple in cui ci sono k elementi di F e n−k di N−F = $\binom{F}{k}\binom{N-F}{n-k}$
- # casi possibili = # n-uple di N elementi = $\binom{N}{n} = \frac{N!}{n!(N-n)!}$
- <u>Ne deriva</u> $P\{X=k\} = \frac{\binom{F}{k}\binom{N-F}{n-k}}{\binom{N}{n}}$

es. Un'azienda riceve da un fornitore un pacco con 20 pezzi.
L'azienda ne prova 5 estraendoli dal pacco, se nessuno è difettato 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 difettato ovvero: $P\{X=0\}$ dove $X \sim IP(20,6,5)$

$$P\{X=0\} = \frac{\binom{6}{0}\binom{20-6}{5-0}}{\binom{20}{5}} = \frac{1\cdot\binom{14}{5}}{\binom{20}{5}} = \frac{14!}{\cancel{5!}\,9!} \cdot \frac{\cancel{5!}\,15!}{20!} = \frac{14\cdot13\cdot12\cdot11\cdot10}{20\cdot19\cdot18\cdot17\cdot16} = 0.125$$

Osservazione

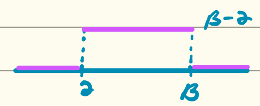
Le v.a. viste finora sono <u>discrete</u>, le prossime sono v.a. con densità, in <u>generale</u>:
Sia X una v.a. con densità f, se f = 0 al di fuori di un certo intervallo [ovvero su $(c,d)^c$] $\Rightarrow$

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

V.A. Uniformi

Assumono valori in un intervallo e sono costanti in esso.
Definiamo X v.a. uniforme sull'intervallo (α,β) (notazione $U(\alpha,\beta)$)

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



Si indica con $\mathbb{1}_{(\alpha,\beta)}$

$$P\{a\le X\le b\} = \int_a^b \frac{1}{\beta-\alpha}\cdot\mathbb{1}_{(\alpha,\beta)}(x)\,dx = \frac{1}{\beta-\alpha}\int_{(a,b)\cap(\alpha,\beta)} 1\,dx = \frac{\text{Lungh}((\alpha,\beta)\cap(a,b))}{\beta-\alpha} \rightarrow (=\text{lungh}(\alpha,\beta))$$

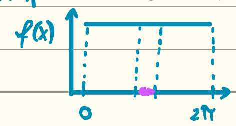
es. Nel gioco ruota della fortuna un ago viene fatto ruotare in modo tale da non privilegiare nessun punto della circonferenza quando si arresta, quindi scelta di un punto $(0,2\pi)$ senza preferenze.

X = "angolo ottenuto quando l'ago si ferma" $\Rightarrow$ v.a. che prende valori in $(0,2\pi)$ senza preferenza
$\Rightarrow X \sim U(0,2\pi)$ 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 VINCE il turno.
Determinare la probabilità che il giocatore PERDA il turno (non cade in quello spicchio).

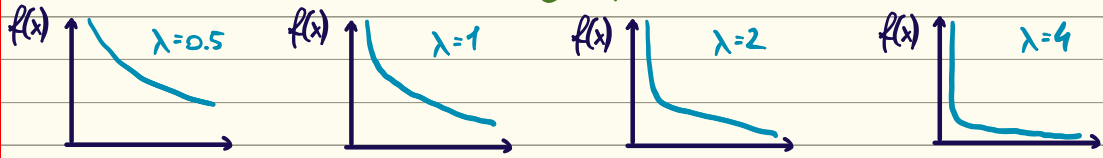
$P\{X \in (\text{quello spicchio})^c\}$ = area al di sotto del grafico del complementare dello spicchio

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



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

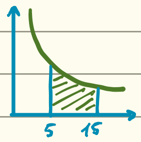
$Exp(\lambda)$: v.a. con densità $f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{per } x > 0 \\ 0 & \text{per } x \leq 0 \end{cases} = \lambda e^{-\lambda x} \cdot \mathbb{1}_{(0,+\infty)}(x)$



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 = "tempo impiegato per assistere i clienti" $\sim Exp(\lambda)$
Con $\lambda = \frac{1}{10}$ minuti, calcolare la probabilità che per assistere un cliente il commesso impieghi tra 5 e 15 minuti.

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



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

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

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

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

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

$$P\{X > 8 \mid X > 2\} = P\{X > 6\} = 1 - P\{X \leq 6\} = 1 - \left(1 - e^{-3/4}\right) = e^{-3/4}$$

assenza di memoria

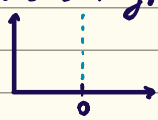
Trasformazione della v.a. con densità

Sia $X: \Omega \to \mathbb{R}$ una v.a. con densità $f_X: \mathbb{R} \to \mathbb{R}$ ($P\{X \in A\} = \int_A f(x)dx$)
Sia poi $h: \mathbb{R} \to \mathbb{R}$ una funzione (es. $h(x) = x^2$)
Allora $Y = h \circ X : \Omega \to \mathbb{R}$ è anche lei una v.a.

Non è detto che Y ammetta densità, e anche quando la ammette,
$\nexists$ una regola generale per calcolarla.

· $\equiv 0$ allora $Y = h \circ X$ -

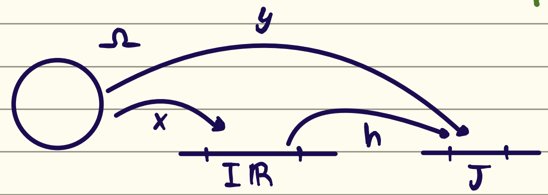
Se, per esempio, h è la funzione , e a sua volta $\equiv 0$ e non c'è modo di trovare una funzione tale che $P(Y \in A)$ sia uguale all'area al di sotto del grafico



- Lo stesso è vero $\forall$ funzione $h:\mathbb{R}\to\mathbb{R}$ 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 biunivoca, 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 $h: I\to J$ un DIFFEOMORFISMO (C^1 invertibile con inversa C^1) allora la v.a. $Y = h\circ X$ ammette densità data da
$$f_Y(y) = \begin{cases} f_X(h^{-1}(y))\left|\frac{dh^{-1}}{dy}(y)\right| & \text{con } y\in J \\ 0 & \text{con } y\notin J\end{cases}$$
$$= f_X(h^{-1}(y))\cdot\left|\frac{dh^{-1}}{dy}(y)\right|\cdot \mathbb{1}_J(y)$$



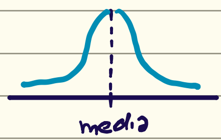
V.A. gaussiane (o normali)

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.
Z v.a. normale standard $\doteq$ 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)

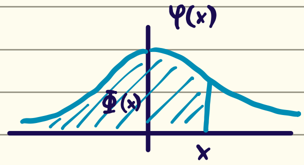
↳ Notazione $N(0,1)$

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



Funzione di ripartizione di N(0,1)

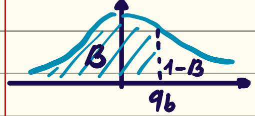
Si dimostra che è $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt$

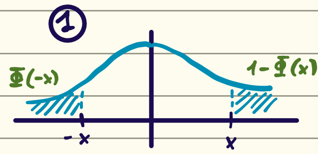


ovvero $P(Z\le x)$

β-quantile di N(0,1)

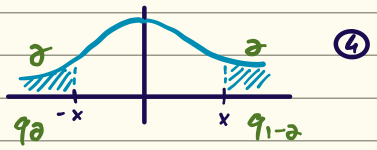
Si indica con q_β il β-quantile di $N(0,1)$, ovvero il numero tale che $\Phi(q_\beta) = \beta$ con $\beta\in(0,1)$





Proprietà di N(0,1)

1) $\Phi(-x) = 1-\Phi(x)$
2) $P\{-x\le Z\le x\} = \Phi(x)-\Phi(-x) = \Phi(x) - [1-\Phi(x)] = 2\Phi(x)-1$
3) $\Phi(0) = P\{Z\le 0\} = \frac{1}{2}$
4) $q_{1-a} = -q_a \Rightarrow$ l'area fino a q_a è a, fino a q_{1-a} l'area è $1-a$, quella dopo è a



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 $\Phi: P\{a<Z<b\} = \int_a^b\varphi(x)dx = \Phi(b)-\Phi(a)$

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 $q_a = -q_{1-a}$

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

$P\{-2 \leq Z \leq 2\} \approx 0.94$

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

Variabile Gaussiana Generale

Se $Z \sim N(0,1)$, posso ottenere un'altra v.a. con densità componendo con $h: \mathbb{R} \to \mathbb{R}$ $x \mapsto \sigma x + m$

La variabile Y ottenuta è $Y = \sigma Z + m$ ed è possibile calcolare la sua densità tramite la formula di cambio variabile

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

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

Calcoli e standardizzazione

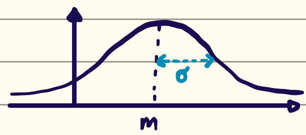
Dato $N(m, \sigma^2)$, per fare i calcoli ci si riconduce a $N(0,1)$ tramite il cambiamento di variabile inverso rispetto a quello visto nella simmetria (STANDARDIZZAZIONE)

$Z \sim (0,1) \Longleftrightarrow Y = \sigma Z + m \sim N(m, \sigma^2)$, quindi:

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

In particolare:

- $P\{m-\sigma \leq Y \leq m+\sigma\} = P\{\frac{m-\sigma-m}{\sigma} \leq Z \leq \frac{m+\sigma-m}{\sigma}\} = P\{-1 \leq Z \leq 1\} \approx 0.68$
- $P\{m-2\sigma \leq Y \leq m+2\sigma\} = P\{-2 \leq Z \leq 2\} \approx 0.94$
- $P\{m-3\sigma \leq Y \leq m+3\sigma\} = P\{-3 \leq Z \leq 3\} \approx 0.997$



Proprietà di riproducibilità

$$\left.\begin{array}{l}\text{Sia } Y \text{ una v.a. gaussiana} \sim N(m, \sigma^2) \\ \text{Sia } V = aY + \beta \text{ con } a, \beta \in \mathbb{R}, a \neq 0\end{array}\right\} \Longrightarrow V \sim N(am+\beta, a^2\sigma^2)$$

dimostrazione:

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

$$V = aY + \beta \Longleftrightarrow Y = \frac{V-\beta}{a} \Longrightarrow Z = \frac{\frac{V-\beta}{a} - m}{\sigma} = \frac{V-\beta-am}{a\sigma} = \frac{V-(am+\beta)}{a\sigma}$$

$\Rightarrow V \sim N(am+\beta, \sigma^2 a^2)$

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

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

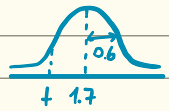
$\Rightarrow P\{Y < 1.6\} = P\{0.6Z + 1.7 < 1.6\} = P\{Z < \frac{-0.1}{0.6}\} = \Phi(-0.17) =$

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

b) Determinare un valore $t \in \mathbb{R}$ t.c. il 95% delle viti abbia lunghezza almeno t
↳ $P\{Y \geq t\} = 0.95$

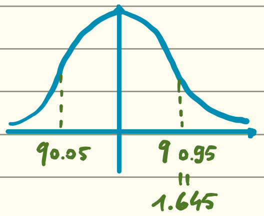
$P\{Y \geq t\} \underset{\text{stand.}}{=} P\{0.6Z + 1.7 \geq t\} = P\{Z \geq \frac{t-1.7}{0.6}\} = 1 - P\{Z \leq \frac{t-1.7}{0.6}\} = 0.95$

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



Cerco quindi il quantile $q_{0.05}$

$q_{1-a} = -q_a \Longrightarrow q_{0.05} = q_{1-0.95} = -q_{0.95} \approx 1.645 \Longrightarrow \frac{t-1.7}{0.6} \approx -1.645 \Longrightarrow t \approx 0.713$



STATISTICA DESCRITTIVA 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

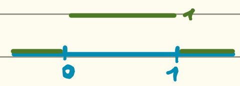
Indice che fornisce informazioni sulla posizione di una v.a. X, si indica con $E[X]$ oppure μ ed è definito come segue:

- X v.a. discreta $\Rightarrow E[X] \doteq \sum_i a_i \cdot p(a_i)$ [in caso gli a_i siano infiniti $\sum_{i=1}^{\infty} a_i \cdot p(a_i) = \lim_{n\to\infty} \sum_{i=1}^{n} a_i \cdot p(a_i)$]
- X v.a. con densità $f \Rightarrow E[X] = \int_{-\infty}^{+\infty} x f(x)dx \quad \left(= \lim_{b\to+\infty} \int_{-\infty}^{b} x f(x) dx\right)$

es. v.a. discreta
X = esito di un lancio di un dado equilibrato
$\{a_1, \dots, a_6\} = \{1, 2, \dots, 6\} \quad P\{X=k\} = \frac{1}{6} \quad \forall k \in \{1, \dots, 6\}$

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

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

es. v.a. con densità
Sia $X \sim U(0,1)$ $\longrightarrow$  $\left(\Rightarrow f(x) = \begin{cases} 1 & \text{per } x \in (0,1) \\ 0 & \text{altrimenti} \end{cases}\right)$

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

$\hookrightarrow$ Il valore atteso non è sempre finito e non sempre definito.

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

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 ∞

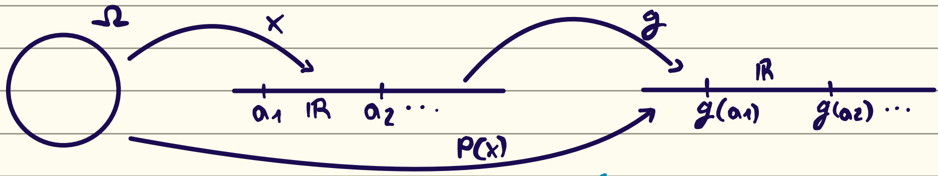
infatti: $E[X] = \begin{cases} \sum_k a_k \cdot p(a_k) & \text{caso discreto} \longrightarrow \text{se i valori sono tutti positivi ha sempre senso} \\ \int_{\mathbb{R}} x f(x)dx & \text{caso con densità} \longrightarrow \int_0^{\infty} x f(x)dx \in [0,\infty] \text{ l'integrale indefinito converge o tende a } +\infty \end{cases}$

- Se X v.a. generica ammette $E[X]$ se:
  - caso discreto: $\sum_k |a_k| p_X(a_k) < \infty \longrightarrow$ Se X assume un # finito di valori allora ammette valore atteso (somma finita)
  - caso con densità: $\int_{-\infty}^{+\infty} |x| f(x)dx < \infty \longrightarrow$ 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 $\int_a^b \dots$

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

Proposizione X v.a. discreta, oppure con densità f, $g:\mathbb{R}\to\mathbb{R}$ t.c. $g(X)$ sia v.a.
Allora:
1) Se X è discreta con funzione di massa $p \Rightarrow E[g(X)] = \sum g(a_k)\cdot p(a_k)$

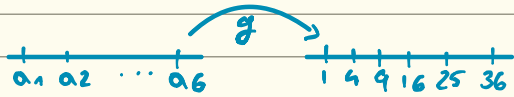
2) Se X ha densità $f \Rightarrow E[g(X)] = \int_{-\infty}^{+\infty} g(x) f(x)\,dx$



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

$$\Longrightarrow E[g(X)] = E[X^2] = \sum g(a_k)\cdot \underset{\substack{\| \\ 1/6}}{p(a_k)}$$

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



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 $E:\{\text{v.a. con val. atteso}\}\longrightarrow\mathbb{R}$ 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"

$$E[X]=E[Y]=\sum_{k=1}^{6} k\cdot\frac{1}{6} = \frac{1+2+3+4+5+6}{6} = \frac{7}{2}$$
$$E[X+Y] = E[X]+E[Y] = 7/2 + 7/2 = 7$$

Momento di ordine n di X $E[X^n]$ Data X v.a., chiamo momento di ordine n di X il valore atteso di X^n, ovvero $E[X^n]$ (ad esempio, $E[X]$ = momento di ordine 1)

. caso <u>discreto</u>: $E[X^n] = \sum_k a_k^n \cdot p(X=a_k)$

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

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

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

$$\text{var}(X) = E[(X-E[X])^2] = E[X^2 - 2XE[X] + E[X]^2] \underset{①②}{=} E[X^2] - 2E[X]\cdot E[X]\cdot E[E[X]^2]$$
$$\Longrightarrow E[X^2] - E[X]^2$$

Deviazione stand. di X $\sigma(X)$ $\sigma(X) = \sqrt{\text{var}(X)}$

Significato di var(X) Come per $E[X]$, 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 $var(X)$
$var(X) = E[X^2] - E[X]^2 = \underbrace{91/6 - (7/2)^2}_{\text{dall'esempio precedente}} = 35/12$

Proprietà di Scaling
- $Var(aX+b) = a^2 \cdot var(X)$
- In particolare, $Var(b) = 0$

dimostrazione
$var(aX+b) = E[(aX+b)^2] - (E[aX+b])^2 = a^2[E[X^2] - E[X]^2] = a^2 Var(X)$

Disuguaglianza di Chebyshev
X v.a. con momento secondo, $d>0$. Allora:
$P\{|X - E[X]| \geq d\} \leq \frac{1}{d^2} \cdot var(X)$

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

Corollario
$Var(X) = 0 \iff X$ costante

dimostrazione
$(\Rightarrow)$ $Var(X) = 0 \Rightarrow P\{|X - E[X]| \geq d\} \leq \frac{1}{d^2} \cdot 0 = 0$
$(\Leftarrow)$ X è costante $= E[X] \Rightarrow Var(X) = E[X^2] - E[X]^2 = cost^2 - cost^2 = 0$

VARIANZA E VALORE ATTESO DELLE V.A. NOTEVOLI

$B(n,p)$

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

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

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

- $P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$
- $E[X] = E[X_1 + X_2 + \dots + X_n] = \sum_i E[X_i] = n \cdot p$
- $var(X) = E[X^2] - E[X]^2 = np(1-p)$

$T \sim Geom(p)$ $\quad T$ = "# prova in cui si verifica il 1° successo"

- $P_T(\{\underbrace{0,\dots,0}_{k-1},\underset{k}{1}\}) = p \cdot (1-p)^{k-1} \ \forall k \in \mathbb{N}^+$
- $E[T] = \frac{1}{p}$
- $Var[T] = \frac{1-p}{p^2}$
- $T \geq 0 \Rightarrow \exists E[T]$

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.

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

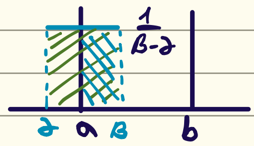
- $P(X=x) = \frac{e^{-\lambda}\cdot\lambda^k}{x!}$
- $E[X] = \sum_k a_k\, p(X=a_k) = \sum_k k\,\frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_k k\cdot\frac{\lambda^k}{k!} = e^{-\lambda}\sum_k \frac{\lambda^k}{(k-1)!} = \cancel{e^{-\lambda}}\cdot\lambda\cdot\cancel{\boxed{\sum_k \frac{\lambda^{k-1}}{(k-1)!}}}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

$\Rightarrow Var(x) = \lambda$

$X \sim IP(N,F,n)$ "※ successi in n estrazioni in blocco"
N unità totali, F favorevoli, n unità estratte

- $P\{X=k\} = \frac{\binom{F}{k}\binom{N-F}{n-k}}{\binom{N}{n}}$
- $E[X] = n\cdot\frac{F}{N}$
- $var(x) = n\cdot\frac{F}{N}\left(1-\frac{F}{N}\right)\frac{N-n}{N-1}$

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



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

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

- $P\{a \le X \le b\} = \frac{lungh((\alpha,\beta)\cap(a,b))}{lungh(\alpha,\beta)}$

Essendo f supportata su un intervallo $(\alpha,\beta) \Rightarrow$ ha momenti di ogni ordine (quindi $\exists E[X], \exists E[X^n]\ \forall n$)

- $E[X] = \int_{-\infty}^{+\infty} x\,f(x)\,dx = \int_\alpha^\beta x\cdot\frac{1}{\beta-\alpha}\,dx = \frac{1}{\beta-\alpha}\cdot\left[\frac{x^2}{2}\right]_\alpha^\beta = \frac{\beta^2-\alpha^2}{2(\beta-\alpha)} = \frac{\beta+\alpha}{2}$
- $E[X^n] = \int_{-\infty}^{+\infty} x^n f(x)\,dx = \int_\alpha^\beta x^n\cdot\frac{1}{\beta-\alpha}\,dx = \frac{1}{\beta-\alpha}\cdot\frac{1}{n+1}\left[x^{n+1}\right]_\alpha^\beta$

$\Rightarrow E[X^2] = \frac{1}{\cancel{\beta-\alpha}}\cdot\frac{1}{3}\cdot\cancel{(\beta^3-\alpha^3)} = \frac{1}{3}(\alpha^2+\alpha\beta+\beta^2)$

- $Var(x) = E[x^2]-E[x]^2 = \frac{1}{3}(\alpha^2+\alpha\beta+\beta^2) - \left(\frac{\beta+\alpha}{2}\right) = \frac{1}{3}\alpha^2+\frac{1}{3}\alpha\beta+\frac{1}{3}\beta^2-\frac{\beta^2}{4}-\frac{\alpha^2}{4}-\frac{\beta\alpha}{2} = \frac{(\beta-\alpha)^2}{12}$

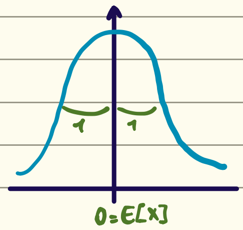
$X \sim Exp(\lambda)$ $f(x) = \lambda e^{-\lambda x}\cdot \mathbb{1}_{(0,\infty)}(x)$

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

- $E[X] = \int_{-\infty}^{+\infty} x\cdot\lambda e^{-\lambda x}\,\mathbb{1}_{(0,\infty)}(x)\,dx = \int_0^\infty x\cdot\lambda\cdot e^{-\lambda x}dx = -xe^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x}dx = -\frac{x}{e^{\lambda x}}\Big|_0^\infty + \left(-\frac{1}{\lambda}e^{-\lambda x}\right)\Big|_0^\infty = \frac{1}{\lambda}$
- $E[x^2] = \frac{2}{\lambda^2}$
- $Var(x) = E[x^2]-E[x]^2 = \frac{2}{\lambda} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

$X \sim N(m, \sigma^2)$

. Caso gaussiana standard $\quad \varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}, \; x \in \mathbb{R}$



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

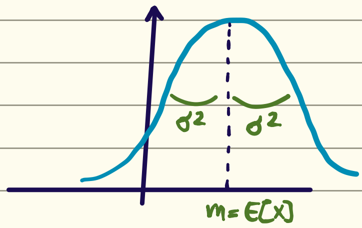
$\Rightarrow$ Se n è dispari è una funzione dispari

Se n è pari

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

- $Var(X) = E[X^2] - E[X]^2 = 1$

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

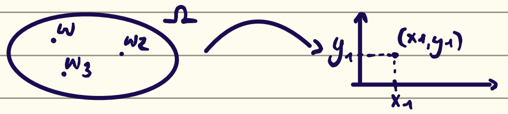


- $E[X] = E[\sigma Z + m] = \sigma E[Z] + E[m] = E[m] = m$
- $Var(X) = \sigma^2 var(Z) = \sigma^2$

$\Rightarrow$ 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 $(x,y) \in \mathbb{R}^2$



$(X,Y): \Omega \longmapsto \mathbb{R}^2$

$\omega \longmapsto (X(\omega), Y(\omega))$

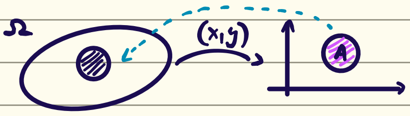
es. Ω = laureati triennali nel 2023

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

$\Rightarrow (X,Y)(\omega)$ = (voto di laurea, mesi tra laurea e 1° impiego)(ω)

Legge congiunta di X e Y

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



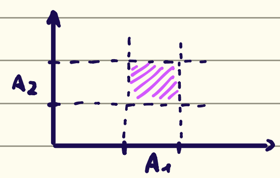
Significato: esprime la probabilità con la quale contemporaneamente entrambe le variabili assumono specifici valori. È quindi indispensabile per l'analisi dei legami e delle forme di dipendenza.

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:

$P\{(X,Y) \in A_1 \times A_2\} = P\{X \in A_1, Y \in A_2\}$
$= P(\{X \in A_1\} \cap P\{Y \in A_2\})$
$= P$("si verificano congiuntamente $X \in A_1$ e $Y \in A_2$")



es. X = voto Y = mesi laurea – I impiego

$P\{(X,Y) \in [105,110] \times [0,2]\} = P\{$"laureato con voto $\in [105,110]$ 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 ~~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)

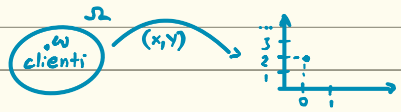
Preso $(X,Y): \Omega \longmapsto \mathbb{R}^2$, $\forall (x_i, y_j)$ si può calcolare $P(x_i, y_j) = P\{X = x_i, Y = y_j\}$.

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 = v.a. che conta il ※ acquisti



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

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

La fun. di prob. marginale della X si ottiene dalla $P_{(X,Y)}$. Precisamente:

$P_X = \{X = x_i\} = P(X^{-1}(x_i)) = P(\bigcup_j (X,Y)^{-1}(x_i, j)) = P((X,Y)^{-1}(x_i,1) \cup (X,Y)^{-1}(x_i,2) \cup (X,Y)^{-1}(x_i,3))$

$= P((X,Y)^{-1}(x_i,1)) + P((X,Y)^{-1}(x_i,2)) + P((X,Y)^{-1}(x_i,3))$

$\Longrightarrow P_X\{X=0\} = P((X,Y)^{-1}(0,1)) + P((X,Y)^{-1}(0,2)) + P((X,Y)^{-1}(0,3)) = 0.2 + 0.15 + 0.05 = 0.40$

$\Longrightarrow P_X\{X=1\} = 0.2 + 0.25 + 0.15 = 0.60$ $\longrightarrow$ prob. che un cliente possieda la carta fedeltà

Analogamente per P_Y

Osservazioni

1) $P(x_i, y_j) \geq 0 \quad \forall i,j$

$\sum_i \sum_j P(x_i, y_j) = 1$

dim: • $p(x_i, y_j) = P\{X = x_i, Y = y_j\} \geq 0 \longrightarrow$ probabilità

• $\sum_i \sum_j p(x_i, y_j) = P\{\bigcup_i \bigcup_j (X,Y)^{-1}(x_i, y_j)\} = P(\Omega) = 1$

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 $P_X: \mathcal{P}(\mathbb{R}) \to \mathbb{R}$ e $P_Y: \mathcal{P}(\mathbb{R}) \to \mathbb{R}$

dimostrazione:

• $P_X(x_i) = P\{X = x_i\} \doteq P(X^{-1}(x_i)) = P\{\bigcup_j (X,Y)^{-1}(x_i, y_j)\} \overset{\text{disgiunti}}{=} \sum_j P\{(X,Y)^{-1}(x_i, y_j)\} = \sum_j P_{(X,Y)}(x_i, y_j)$

• $P_Y(y_j) = P\{Y = y_j\} \doteq P(Y^{-1}(y_j)) = P\{\bigcup_i (X,Y)^{-1}(x_i, y_j)\} \underset{\text{disgiunti}}{=} \sum_i P_{(X,Y)}(x_i, y_j)$

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

$$X_i = \begin{cases} 1 \longleftrightarrow T \\ 0 \longleftrightarrow C \end{cases}$$ dell'i-mo lancio, $i = 1,2 \Longrightarrow P_{X_i} \sim B(1, 1/2)\ \forall i = 1,2$

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

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

V.A. doppia con densità

La v.a. doppia (X,Y) si dice una v.a. doppia con densità $f:\mathbb{R}^2 \to \mathbb{R}$ (congiunta) se:
$P_{(X,Y)}(A) = P\{(X,Y) \in A\} = \iint_A f(x,y)\,dx\,dy \quad \forall A \subseteq \mathbb{R}^2$ (es. $A = [a,b] \times [c,d]$)

V.A. indipendenti

(Ω, P) spazio di probabilità e $X,Y: \Omega \to \mathbb{R}$ 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:
$P\{X \in A, Y \in B\} = P\{X \in A\} \cdot P\{X \in B\}$

Generalizzando, $X_1, \dots, X_n : \Omega \to \mathbb{R}$ 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 $P\{X_1 \in A_1, \dots, X_n \in A_n\} = P\{X_1 \in A_1\} \cdot \ldots \cdot P\{X_n \in A_n\}$

es. n lanci di moneta $X_i = \begin{cases} 1 & \text{se T all'i-esimo lancio} \\ 0 & \text{altrimenti} \end{cases} \quad i = 1, \dots, n$
$\Longrightarrow$ le X_i sono indipendenti

Criterio di indipendenza per V.A. discrete

X, Y v.a. discrete, allora X, Y sono indipendenti $\iff P_{(X,Y)}(x_i, y_j) = P_X(x_i) \cdot P_Y(y_j) \quad \forall (x_i, y_j)$
dimostrazione:
$(\Rightarrow)$ X,Y 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 $X_i = \begin{cases} 1 \longleftrightarrow \text{T all'i-esimo lancio} \\ 0 \quad \text{altrimenti} \end{cases}$
Allora $P_{(X_1,X_2)}(a,b) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P_{X_1}(a) \cdot P_{X_2}(b) \Longrightarrow$ sono indipendenti

es2. 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 <u>non</u> 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 $h, k: \mathbb{R} \to \mathbb{R} \Longrightarrow h(x)$ e $k(y)$ sono indipendenti

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

Proprietà di riproducibilità

Nei 3 casi considero la v.a. univariata $h: \mathbb{R}^2 \to \mathbb{R}$ $(X,Y) \longrightarrow X+Y$
- $X \sim B(m,p)$ e $Y \sim B(h,p)$ indipendenti $\Longrightarrow X+Y \sim B(m+h, p)$
- $X \sim P(\lambda)$ e $Y \sim P(\mu)$ indipendenti $\Longrightarrow X+Y \sim P(\lambda + \mu)$
- $X \sim N(m_1, \sigma_1^2)$ e $Y \sim N(m_2, \sigma_2^2)$ indipendenti $\Longrightarrow X+Y \sim N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$

es. # errori commessi da un server in un giorno $\sim P(2.5)$
errori in un dato giorno indipendente dal # errori in altri giorni
Probabilità di avere almeno 3 errori in 2 giorni consecutivi?

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

La v.a. $X+Y$ conta gli errori totali nei 2 giorni consecutivi, quindi $P\{X+Y \geq 3\}$?
X, Y indipendenti $\Longrightarrow X+Y \sim P(2.5 + 2.5) = P(5)$
Quindi $P\{X+Y \geq 3\} = 1 - P\{X+Y < 3\} = 1 - P\{X+Y = 0\} - P\{X+Y = 1\} - P\{X+Y = 2\}$

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

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

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 $P_{2x}(2) = p$, $P_{2x}(0) = 1-p$ $\quad x = \begin{cases} 1 & \text{con prob. } p \\ 0 & \text{con prob. } 1-p \end{cases}$ $\quad 2x = \begin{cases} 2 & p \\ 0 & 1-p \end{cases}$

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 $X, Y: \Omega \longmapsto \mathbb{R}$ v.a. e sia $g: \mathbb{R}^2 \to \mathbb{R}$ con $g(x,y) = g \circ (x,y): \Omega \longmapsto \mathbb{R}$.
Allora $g(X,Y)$ è una v.a. univariata per cui è possibile provare a calcolare il valore atteso.
Se X e 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 e Y discrete, $E[X \cdot Y] \doteq \sum_{i,J} x_i, y_J \cdot P\{X = x_i, Y = y_J\} \overset{\text{indip.}}{=} \sum_{i,J} x_i \cdot y_J \cdot P_X(x_i) \cdot P_Y(y_J)$
$= \sum_i x_i \cdot P_X(x_i) \cdot \sum_J y_J \cdot P_Y(y_J) = E[X] \cdot E[Y]$

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

Reminder

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

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

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

Covarianza di 2 V.A. con momento secondo

Date X e Y v.a. con momento secondo, si definisce covarianza di X e Y il numero $cov(X,Y) = E[(X - E[X])(Y - E[Y])]$

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

Coefficiente di correlazione

Date X e Y v.a. con momento secondo e con $var(x) \neq 0$ e $var(y) \neq 0$ è il numero $\rho \doteq \frac{cov(x,y)}{\sqrt{var(x) var(y)}}$

Variabili Scorrelate

- Se $\rho(x,y) = 0$ (equivalente a $cov(x,y) = 0$) $\Rightarrow$ X e Y sono scorrelate
- Inoltre, se X e Y con momento secondo e indipendenti $\Rightarrow$ X e Y scorrelate

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

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 Z v.a. con momento secondo e $a, b \in \mathbb{R}$. Allora:
1) $cov(aX + bY + c, Z) = a \cdot cov(X,Z) + b \cdot cov(Y,Z)$ bilineare
2) $cov(X,Y) = cov(Y,X)$ e $cov(X,X) = var(X)$ simmetrica

Varianza della somma

X, Y v.a. Allora $var(X+Y) = var(X) + var(Y) + 2cov(X,Y)$
In particolare, se X e Y sono scorrelate $\iff var(X+Y) = var(X) + var(Y)$

dimostrazione: $var(X+Y) \overset{①}{=} cov(X+Y, X+Y) \overset{②}{=} cov(X,X) + cov(X,Y) + cov(Y,X) + cov(Y,Y)$
$\overset{①}{=} var(X) + cov(X,Y) + cov(X,Y) + var(Y)$ (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 $W_n = \sum_{i=1}^{n} x_i$

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

Teorema

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

a) $|\rho(X,Y)| \leq 1$
b) $\min_{a,b \in \mathbb{R}} E[(Y - a - bX)^2] = Var(Y) \cdot (1 - \rho(X,Y)^2)$

Traduzione:
- Presi a b che realizzano il minimo in ⓑ, la retta $Y = a + bX$ è la migliore approssimazione lineare tra X e Y
- Il valore del minimo è proporzionale a $1 - \rho(X,Y)^2$
  $\Rightarrow$ la relazione tra X e Y è tanto meglio approssimata da una retta quanto più $\rho(X,Y)$ è vicino a ± 1
  $\Rightarrow \rho$ è una misura della DIPENDENZA LINEARE tra X e Y
  - $\rho \approx \pm 1$ alta dipendenza lineare
  - $\rho \approx 0$ bassa dipendenza lineare
- Se X, Y sono due caratteri di n individui di una popolazione, $\rho(X,Y)$ è il coefficiente di correlazione lineare campionaria tra i caratteri di TUTTA la popolazione (e non solo di un campione)

V.A. Indipend. ed equamente 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 $\forall i$)

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 $\overline{X}_n = \frac{X_1, \dots, X_n}{n}$

Convergenza

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

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 $\overline{X}_n$ converge in probabilità a $E[X_i] = \mu$.
Ovvero $\forall \varepsilon > 0 \lim_{n \to \infty} P\{|\overline{X}_n - \mu| > \varepsilon\} = 0$

Trad: 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
$\overline{X}_{1000} \approx \mu = 3{,}5 \quad E[X_i] = 3{,}5$

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

$\Rightarrow \overline{X}_n$, secondo LGN, converge in probabilità $n \to \infty$ a $E[X_i] = p$

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

S^2

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

$\downarrow$

Proprietà

$X_1, \dots, X_n, \dots$ successione di v.a. i.i.d. con momento secondo, $\sigma^2 = Var(X_i)$.
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 e Y.
Supponiamo che F sia continua.
Allora diciamo che Y_n converge in legge (o distribuzione) a Y [si scrive $Y_n \xrightarrow[n \to \infty]{(d)} Y$]
Se $\lim_{n \to \infty} F_n(t) = F(t) \ \forall t \in \mathbb{R}$

Teorema centrale del limite (TCL/TLC)

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 $\sigma^2 = Var(X_i)$, con $\sigma > 0$ (cioè X_i non costanti). Sia inoltre $\overline{X}_n = 1/n \sum_{i=1}^{n} X_i$.
Allora $\sqrt{n}\, \frac{\overline{X} - \mu}{\sigma} \xrightarrow[n \to \infty]{(d)} Z \sim N(0,1)$

Cioè $\forall -\infty \leq a < b \leq +\infty \quad \lim_{n \to \infty} P\{a \leq \sqrt{n}\, \frac{\overline{X}_n - \mu}{\sigma} \leq b\} = P\{a \leq Z \leq b\} = \Phi(b) - \Phi(a)$

Dimostrazione:

$X_1, \dots, X_n$ v.a. i.i.d., la media è $\overline{X}_n = 1/n \sum_{i=1}^{n} X_i$.
So anche che $E[\overline{X}_n] = E\left[\frac{X_1 + \dots + X_n}{n}\right] = 1/n \cdot E[X_1 + \dots + X_n] = 1/n \left(\underset{\mu}{E[X_1]} + \dots + \underset{\mu}{E[X_n]}\right) = \frac{n\mu}{n} = \mu$
$Var[\overline{X}_n] = Var\left[\frac{X_1 + \dots + X_n}{n}\right] = 1/n^2 \, Var[X_1 + \dots + X_n] = 1/n^2 \left(\underset{\sigma^2}{Var(X_1)} + \dots + \underset{\sigma^2}{Var(X_n)}\right) = \frac{n\sigma^2}{n} = \sigma^2$

Standardizzazione della v.a. $\overline{X}_n \longrightarrow Z_n = \frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = \sqrt{n}\, \frac{\overline{X}_n - \mu}{\sigma}$

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

Caso particolare → Approssimazione normale delle binomiale

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

$X_1, \dots, X_n$ v.a. i.i.d. $\quad \underbrace{X_1 + \dots + X_n}_{\downarrow} \sim B(n,p)$

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

$E[Y_n] = nE[\overline{X}_n] = n \cdot p = n \cdot \mu$

$Var(Y_n) = n^2 \cdot p(1-p) = n \cdot np(1-p) = n\sigma^2$

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

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

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

es. 1'000 lanci moneta equilibrata,

a) probabilità di avere almeno 480 teste?

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

$$\frac{Y - np}{\sqrt{np(1-p)}} = \frac{Y - 1'000 \cdot 1/2}{\sqrt{1'000 \cdot 1/2 (1 - 1/2)}} = \frac{Y - 500}{\sqrt{250}} \sim Z = N(0,1) \qquad 250 > 10$$

$$P\{Y \geq 480\} = P\left\{\frac{Y-500}{\sqrt{250}} \geq \frac{480-500}{\sqrt{250}}\right\} = P\left(\frac{Y-500}{\sqrt{250}} - 1,26\right) \overset{TCL}{\approx} P(Z \geq -1,26) =$$

$= 1 - \Phi(-1,26) = \Phi\, 1,26 \approx 0,896$ 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 $k < 480$ perchè $0,95 > 0,896$ (risultato punto a)

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

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

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

EQUAZIONE X TROVARE K

si arrotonda x difetto x garantire che prob sia almeno 0,95

STATISTICA INFERENZIALE

Scopo Ottenere info sulla distribuzione di una popolazione studiando dati rilevanti 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 → indipendenza
b) l'estrazione di ogni individuo che andrà a costruire il campione è fatta in modo identico al precedente → 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. $X : \Omega \to \mathbb{R}$ 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 taglia 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;
- $\beta \in H$: H indica l'insieme dei valori ammissibili;
- P_X^β: la β indica che la prob. dipendente da un parametro.

Statistica campionaria Funzione $g(X_1, \dots, X_n)$ dal campione $(X_1, \dots, X_n)$

Stimatori di un parametro β Una statistica campionaria $P(X_1, \dots, X_n)$ atta a stimare β

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

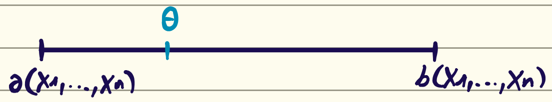
Stimatore corretto (non distorto)	$g(x_1,\dots,x_n)$ stimatore del parametro β si dice corretto se $E[g(x_1,\dots,x_n)]=\beta$
Proposizione	$X_1,\dots,X_n$ campione statistico con momento secondo, allora:
	① $\overline{X}_n$ è uno stimatore corretto per $E[X_i]=E[X]$;
	② S_n^2 è uno stimatore corretto per $Var(X_i)=Var(X)$
	Dimostrazione:
	① $\overline{X}_n$ stimatore corretto per $E[X_i] \iff E[\overline{X}_n]=E[X]$
	$E[\frac{X_1+\dots+X_n}{n}]=1/n\sum_{i=1}^{n}X_i=1/n\cdot n\,E[X]=E[X]$
	② S_n^2 stimatore corretto per $Var(X) \iff E[S_n^2]=Var(X)$
	DA FARE DIM.
Stimatore Consistente	Parametro β della distribuzione, $X_1,\dots,X_n$ $n\in\mathbb{N}$ campione di ∞ v.a. i.i.d. di X
	$g_n(X_1,\dots,X_n)$ è uno stimatore consistente di β se per $n\to\infty$ vale $g_n(X_1,\dots,X_n)\to\beta$ in probabilità
	es. $X_1,\dots,X_n$ campione di v.a. i.i.d. con momento secondo
	Allora $\overline{X}_n$ è uno stimatore consistente per $E[X_i]$ $(=E[X]=\mu)$
	e S_n^2 è uno stimatore consistente per $var(X_i)$
	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 $var(g(X_1,\dots,X_n))\le var(h(X_1,\dots,X_n))$
	es. $X_1,\dots,X_n$ v.a. i.i.d. $\Rightarrow$ $\overline{X}_n$ è tanto più efficiente quanto più n aumenta
	Dim. $Var(\overline{X}_n)=var(\frac{X_1+\dots+X_n}{n})=\frac{1}{n^2}\cdot(var(X_1)+\dots+var(X_n))=\frac{1}{n^2}\cdot n\,var(X_i)=\frac{var(X_i)}{n}$
	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	$L_\beta:\mathbb{R}^n\to\mathbb{R}$ definita come:
	. caso discreto: $P_\beta(X_1,\dots,X_n)=P\{X_1=x_1\}\cdot\ldots\cdot P\{X_n=x_n\}=P\{X_1=x_1,\dots,X_n=x_n\}$
	. caso con densità: $f_\beta(X_1)\cdot\ldots\cdot f_\beta(X_n)$
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 $L_{\hat{\beta}}(X_1,\dots,X_n)=\max_{\beta\in M} L_\beta(X_1,\dots,X_n)$ $\forall(X_1,\dots,X_n)\in\mathbb{R}^n$
	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, $\theta \in H \in \mathbb{R}$
$\alpha \in (0,1)$ e $a,b:\mathbb{R}^n \to \mathbb{R}$ 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 $1-\alpha$ se

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



La prob. che θ stia in questo intervallo è $\geq 1-\alpha$