Statistica e Analisi dei dati

- → STATISTICA DESCRITTIVA → compito di aggregare dati e trasformarli in un indice numerico o trasformarli graficamente
  - indice numerico → analisi quantitativa
  - trasformarli graficamente → analisi qualitativa
- → CALCOLO DELLE PROBABILITÀ
- → STATISTICA INFERENZIALE

Statistica descrittiva

Popolazione: insieme di individui a cui sto facendo riferimento, che voglio studiare e su cui voglio trarre conclusione

Campione: il sottoinsieme d. popolazione su cui baso i miei studi scelto oculatamente, che sia rappresentativo della popolazione

Campione casuale: campione preso senza dipendenze da caratteristiche presenti nei campioni, che influenza la scelta del campione successivo, ogni elem. della popolaz. ha la stessa prob. di essere preso nel campione
↓
preso senza BIAS (pregiudizio)

Esempio: Studenti scuola

I anno	300
II anno	500
III anno	600
IV anno	600
	2000

Frequenza assoluta: numero di occorrenze di quel valore
↓
se divido per la taglia ottengo la **frequenza relativa** (relativa alla taglia del campione)

$$\left.\begin{matrix}300\\500\\600\\600\end{matrix}\right\} /2000 \quad \left\{\begin{matrix}0.15 \cdot 100 = 15\\0.25 \cdot 100 = 25\\0.3 \cdot 100 = 30\\0.3 \cdot 100 = 30\end{matrix}\right.$$

campione casuale stratificato
↓
stratificazione: si fa campionamento su diversi sottogruppi

POSIZIONE / CENTRALITÀ (caratteristica per vedere la grandezza tendenziale se i miei valori sono alti o bassi)

- → MEDIA
- → MEDIANA
- → MODA

} CAMPIONARIA

n = dimensione / taglia del campione
$\{x_1, x_2, \dots, x_n\}$ elementi del campione

scarti: differenze tra ciascun valore dei dati e la media campionaria
Il valore dell' i-esimo scarto é $x_i - \bar{x}$

Media campionaria: media aritmetica degli elementi del campione
$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ (non necessariamente equivale a ció che ho osservato)
↓
Proprietá della media campionaria

- → $\forall i \quad y_i = x_i + b \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}(x_i + b) = \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n} b = \bar{x} + \frac{1}{\cancel{n}} \cdot b \cdot \cancel{n}$
  la media campionaria dei valori traslati é uguale alla media originale traslata dello stesso valore $\bar{y} = \bar{x} + b$
- → $\forall i \quad y_i = a x_i \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} a x_i = a \cdot \frac{1}{n}\sum_{i=1}^{n} x_i = a\bar{x}$
  la media campionaria dei valori scalati é uguale alla media campionaria dei valori originali, scalata
- → $\forall i \quad y_i = a x_i + b \qquad \bar{y} = a\bar{x} + b$ la media campionaria é un operatore lineare

Altri modi per calcolare la media campionaria

Esempio:

valore	frequenza (assoluta)
3	2
4	1
5	3

(Tabella delle frequenze assolute)

$\{3,3,4,5,5,5\}$ scrittura estensiva

Media campionaria: $\frac{3+3+4+5+5+5}{6} = \frac{3\cdot 2 + 4\cdot 1 + 5\cdot 3}{6} = 3\cdot\frac{2}{6} + 4\cdot\frac{1}{6} + 5\cdot\frac{3}{6} = \frac{25}{6}$

freq. ass. — valore — freq relative

Esempio 2: campione di n elementi

$v_1, v_2, \dots, v_k$ valori distinti osservabili

$x_1, x_2, \dots, x_k \qquad k \leq n$

$\underbrace{v_1, v_1, \dots, v_1}_{f_1}, \underbrace{v_2, v_2, \dots, v_2}_{f_2}, \dots, \underbrace{v_n, v_n, \dots, v_n}_{f_k}$ (n)

f_i: frequenza assoluta del valore v_i (numero di volte che v_i occorre nel mio campione)

$\bar{x} = \frac{\underbrace{v_1 + \dots + v_1}_{f_1} + v_2 + \dots + v_2 + \dots + v_k + \dots + v_k}{n} = \frac{v_1 \cdot f_1 + v_2 \cdot f_2 + \dots + v_k \cdot f_k}{n}$

$\bar{x} = \frac{1}{n}\sum_{j=1}^{k} f_j v_j = \sum_{j=1}^{k} \underbrace{\frac{f_j}{n}}_{\text{freq. relativa}} v_j = \sum_{j=1}^{k} f_j' v_j$

In base a come sono organizzati i dati posso scrivere la media campionaria in 3 modi diversi;

Esempio: 2, 110, 5, 7, 6, 7, 3 media campionaria: 20

valori fuori scala / outlier

↓ la loro presenza mi falsa le conclusioni che posso trarre

↓ media campionaria non è uno strumento robusto

⇓

MEDIANA CAMPIONARIA ⟶ prima cosa da fare, ordinare i dati (ordinare il campione)

↓ robusta rispetto agli outlier

2 3 5 | 6 | 7 7 110

si prende il valore centrale (mediana campionaria)

Se ho un numero pari di elementi

$1\ 2\ 3\ |\ \underbrace{5\ \ 6}\ |\ 7\ 7\ 110$

media aritmetica tra i due: 5.5

$\{x_1, \dots, x_n\}$ CAMPIONE

$x_{(1)}, \dots, x_{(n)}$ CAMPIONE ORDINATO

MEDIANA

$x_{\left(\frac{n+1}{2}\right)}$ se n disp.

$\frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n+1}{2}\right)}}{2}$ se n pari

Non tutti i dati sono espressi in numeri e come potremmo calcolare media aritmetica? ⟹ PROPRIETÀ CATEGORICHE: (dati qualitativi / categorici)

- ordinabili, li posso ordinare ↓ mediana valida
- non ordinabili? "colore occhi" (es) ⟹ **MODA CAMPIONARIA**

↓ campione che si verifica con più frequenza (guardo la freq. assoluta, ma anche relativa) osservazioni

Esempio: campione A $\quad 1, 2, 5, 6, 6 \quad \bar{a} = \frac{1+2+5+6+6}{5} = 4$

campione B $\quad -10, 0, 5, 20, 35 \quad \bar{b} = \frac{-10+0+5+20+35}{5} = 4$

hanno la stessa centralità / misura di posizione

ma

NOTEVOLE DIFFERENZA: range di valori $\longrightarrow$ CAMPIONE B più sparso/esteso/disperso

$\Downarrow$

DISPERSIONE/VARIABILITÀ: se non ho variabilità (osservazioni tutte uguali) $\longrightarrow$ media campionaria uguale

le osservazioni possono oscillare (più o meno intorno al valore centrale)

può essere utile cogliere quanto un campione è disperso

ma

come calcolare la dispersione rispetto alla media campionaria?

scarti da $\bar{x}$ $\quad \sum_i (x_i - \bar{x}) = \sum_i x_i - \sum_i \bar{x} = 0 \longrightarrow$ non va bene

$\bar{x} = \frac{1}{n}\sum_i x_i \qquad n\bar{x} = \sum_i x_i$

$\Downarrow$

$\sum_i |x_i - \bar{x}|$

(non va ancora bene, il modulo introduce casini)

$(x_i - \bar{x})^2$ scarto quadratico

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

VARIANZA CAMPIONARIA

Esempio: varianza campionaria CAMPIONE A

$$s_a^2 = \frac{9+4+1+4+4}{4} = 5.5$$

varianza campionaria CAMPIONE B

$$s_b^2 = 792.5$$

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_i (x_i - \bar{x})^2}$$ DEVIAZIONE STANDARD CAMPIONARIA

Formula alternativa varianza campionaria

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_i (x_i^2 - 2\bar{x}x_i + \bar{x}^2) = \sum_i x_i^2 - 2\bar{x}\underbrace{\sum x_i}_{n\bar{x}} + \underbrace{\sum_i \bar{x}^2}_{n\bar{x}^2} = \sum_i x_i^2 - n\bar{x}^2$$

$\{x_1, \dots, x_n\}$ $\quad \bar{x} \quad \forall i = 1, \dots, n \quad y_i = a x_i \quad \bar{y} = a\bar{x}$

$$s_y^2 = \frac{1}{n-1}\sum_i (y_i - \bar{y})^2 = \frac{1}{n-1}\sum_i (a x_i - a\bar{x})^2 = a^2 \frac{1}{n-1}\sum (x_i - \bar{x})^2 = s_x^2$$

$\boxed{s_y^2 = a^2 s_x^2}$ $\quad \forall i = 1, \dots, n \quad y_i = x_i + b \quad \bar{y} = \bar{x} + b$ $\quad$ una traslaz. è un'operaz. che non ha effetto sulla dispersione dei dati

$$\boxed{s_y^2 = \frac{1}{n-1}\sum_i (x_i + \cancel{b} - \bar{x} - \cancel{b})^2 = s_x^2} \longrightarrow s_y^2 = s_x^2 \longrightarrow s_y = s_x$$

$$s_y^2 = a^2 s_x^2 \longrightarrow s_y = |a| s_x$$

la scalatura invece agisce sulla dispersione abbiamo un difetto se operiamo con valori associati a unità di misura, entra in gioco la deviazione standard campionaria

Mediana campionaria m é il valore del campione che é $\geq$ di almeno metá delle osservazioni e $\leq$ di almeno metá delle osservazioni

quell'elemento che una volta ordinati i dati me li spacca a metá

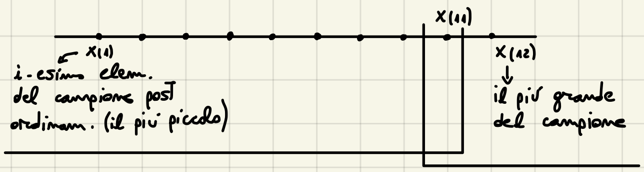
(**quantile campionario**) → generalizzaz. del concetto di mediana

⟹ **QUANTILE di livello $q \in [0,1]$**: é il valore del campione che é $\geq$ di almeno mq delle osservazioni e $\leq$ di almeno $m(1-q)$ delle osservazioni

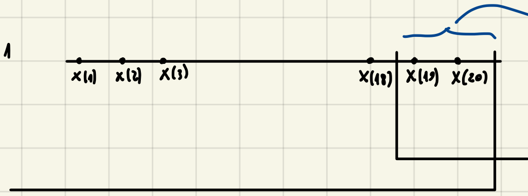
(m = dimensioni del campione)

$m = 12 \quad q = 0.9$

$mq = 10.8 \quad m(1-q) = 1.2$



$m = 20 \quad q = 0.95 \quad mq = 19 \quad m(1-q) = 1$



in tal caso come per la mediana facciamo la media aritmetica $\frac{X_{(19)} + X_{(20)}}{2}$

PERCENTILI: equivalenti ai quantili, ma espressi in percentuali (es di prima: 95%) → $\geq$ di almeno $\frac{p}{100}m$ osservazioni

$\leq$ di almeno $\left(1 - \frac{p}{100}\right)m$ oss

↓

DECILI (1-10)

↓

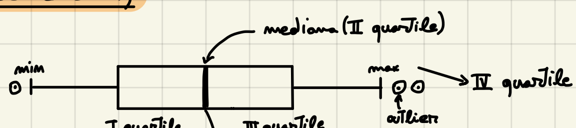
QUARTILI (4 livelli) I quartile II quartile → 2/4 a dx e 2/4 a sx → mediana

III quartile IV quartile

BOX PLOT / PROGRAMMA A SCATOLA (SCATOLA e BAFFI)

↓

I, II, III QUARTILE Rappresentazione:



più larga la scatola

più dati sono dispersi

Indice di centralità e indice di dispersione

RANGE INTERQUATILE (indice di dispersione)

↓

differenza tra III e I quartile

RANGE (indice di dispersione)

↓

differenza tra max e min

$\frac{s}{|\bar{x}|} = s^*$ **COEFFICIENTE DI VARIAZIONE** (q.tà adimensionale)

agisce come fattore di correzione

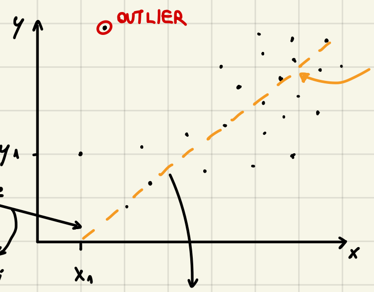
Consideriamo un dataset non con 1, ma con 2 colonne (2 attributi)

possibile relazione tra attributo della prima colonna e attributo della seconda colonna → se c'é una relazione tra due attributi possiamo usarlo a nostro favore

Cambiamo quindi notazione per fare ciò

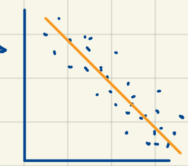
$$\left.\begin{array}{l} 1^a \text{ colonna}: x_1, \dots, x_n \\ 2^a \text{ colonna}: y_1, \dots, y_n \end{array}\right\} \text{ sono da considerare a coppie } \{(x_1,y_1), (x_2,y_2), \dots, (x_n,y_n)\}$$

Strumento qualitativo grafico → **SCATTER PLOT/DIAGRAMMA DI DISPERSIONE**



RELAZIONE DIRETTA (retta crescente)

RELAZIONE INDIRETTA (1 componente grande l'altra piccola)



ci interessano relazioni di tipo tendenziale (es: al crescere di x cresce y e viceversa)

con tale retta posso fare un'estrapolazione/una previsione (esistono altri modi: esponenziale/logaritmico)

Strumento quantitativo

$x_1, \dots, x_n$ $\bar{x}$ $y_1, \dots, y_n$ $\bar{y}$ $\forall i$ x_i é "grande" se $x_i \geq \bar{x}$, é piccolo se $x_i \leq \bar{x}$ (Analogo per le y)

ma vuole dire che $x_i - \bar{x} \geq 0$ e $x_i - \bar{x} < 0$

RELAZIONE TENDENZ. DIRETTA

x_i é grande, y_i é grande ($x_i - \bar{x} \geq 0$ e $y_i - \bar{y} \geq 0$)
OPPURE OPPURE
x_i é piccolo, y_i é piccolo ($x_i - \bar{x} < 0$ e $y_i - \bar{y} < 0$)

$$(x_i - \bar{x})(y_i - \bar{y}) \geq 0$$

$$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) > 0$$

RELAZIONE TENDENZ. INDIRETTA

x_i grande, y_i piccolo ($x_i - \bar{x} \geq 0$ e $y_i - \bar{y} < 0$)
OPPURE OPPURE
x_i piccolo, y_i grande ($x_i - \bar{x} < 0$ e $y_i - \bar{y} \geq 0$)

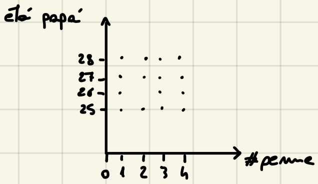
$$(x_i - \bar{x})(y_i - \bar{y}) < 0$$

$$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) < 0$$

COVARIANZA CAMPIONARIA (COV): $\dfrac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \gtrless 0$

Esempio: $x_i = \#$ penne

$y_i =$ età vostro padre quando siete nati



NON HA SENSO PARLARE DI RELAZIONE DIRETTA/INDIRETTA

$\mathrm{cov} = \emptyset$, tendono a cancellarsi le parti

$\downarrow$

COEFFICIENTE DI CORRELAZIONE CAMPIONARIA/INDICE CORRELAZIONE LINEARE:

$$r = \frac{\frac{1}{n-1}\left[(x_i-\bar{x})(y_i-\bar{y})\right]}{S_x \cdot S_y}$$

valore compreso tra $-1 \le r \le 1$ (INDIRETTA -1, DIRETTA 1)

(se $r = \emptyset \longrightarrow$ attributi detti INDIPENDENTI o STATISTICAMENTE INDIPENDENTI)

$$r = \frac{1}{n-1} \frac{\sum_i (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{\sum_i (x_i-\bar{x})^2}{n-1} \cdot \frac{\sum_i (y_i-\bar{y})^2}{n-1}}} = \frac{\sum_i (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_i (x_i-\bar{x})^2 \cdot \sum_i (y_i-\bar{y})^2}}$$

(stesso segno della covarianza)

Proprietà del coefficiente di correlazione campionaria

$x_1, \dots, x_n \quad \forall i = 1, \dots, n \quad y_i = a + bx_i \quad \bar{y} = a + b\bar{x}$

$y_i - \bar{y} = \cancel{a} + bx_i - \cancel{a} - b\bar{x} = b(x_i - \bar{x}) \qquad S_y^2 = b^2 S_x^2 \qquad S_y = |b| S_x$

$$r = \frac{1}{n-1} \frac{\sum_i (x_i-\bar{x})(y_i-\bar{y})}{S_x \cdot S_y} = \frac{1}{n-1} \frac{\sum_i (x_i-\bar{x}) b (x_i-\bar{x})}{|b| S_x \cdot S_x} = \frac{b}{|b|} \cancel{\frac{1}{n-1} \frac{\sum_i (x_i-\bar{x})^2}{S_x^2}}^{\,1}$$

$r = \frac{b}{|b|} \longrightarrow$ se tra x_i e y_i c'è una relazione deterministica lineare

$$\begin{cases} +1 & \text{se } b > 0 \\ -1 & \text{se } b < 0 \end{cases}$$

$x_i \quad x_i' = a + bx_i \quad \bar{x}' = a + b\bar{x} \quad S_x' = |b| S_x \qquad (x_i' - \bar{x}') = \cancel{a} + bx_i - \cancel{a} - b\bar{x} = b(x_i - \bar{x})$

$y_i \quad y_i' = c + dy_i \quad \bar{y}' = c + d\bar{y} \quad S_y' = |d| S_y \qquad (y_i' - \bar{y}') = \cancel{c} + dy_i - \cancel{c} - d\bar{y} = d(y_i - \bar{y})$

$$r = \frac{1}{n-1} \frac{\sum (x_i'-\bar{x}')(y_i'-\bar{y}')}{S_x' S_y'} = \frac{1}{n-1} \frac{b \cdot d \sum_i (x_i-\bar{x})(y_i-\bar{y})}{|b||d| S_x \cdot S_y} = \frac{b \cdot d}{|b| \cdot |d|} \cdot r$$

$$\begin{cases} +1 & bd \text{ concordi di segno} \\ -1 & bd \text{ discordi di segno} \end{cases}$$

r è insensibile alle trasformazioni lineari

Altro modo per calcolare l'indice di correlazione

$$S_x^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_i x_i^2 - n\bar{x}^2\right)$$

$$(x_i-\bar{x})(y_i-\bar{y}) = (x_i y_i - x_i\bar{y} - \bar{x} y_i + \bar{x}\bar{y}) \longrightarrow \sum_i (x_i-\bar{x})(y_i-\bar{y}) = \sum_i x_i y_i - \bar{y}\sum x_i - \bar{x}\sum_i y_i + n\bar{x}\bar{y}$$

$$= \sum_i x_i y_i - n\bar{x}\bar{y} - \cancel{n\bar{x}\bar{y}} + \cancel{n\bar{x}\bar{y}}$$

$$= \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{n-1}$$

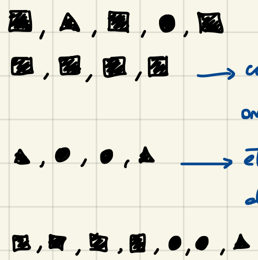
$$r = \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sqrt{\sum x_i^2 - n\bar{x}^2} \cdot \sqrt{\sum y_i^2 - n\bar{y}^2}}$$

Eterogeneità / Omogeneità

Pensiamo di lavorare con attributi che non sono scalari/numerici
$\downarrow$
Tocca abbandonare indici legati a concetti aritmetici

si applica a campioni di tipo categorico (es: colori, ...)

Esempio: Forme geometriche



→ campione altamente omogeneo poco eterogeneo

→ eterogeneo, legato alle frequenze d. oggetti

Indice di GINI per eterogeneità

Partiamo con un campione di n elementi $x_1, \dots, x_n$ $\qquad k \leq n$
$\downarrow$
possono assumere diversi valori: $v_1, v_2, \dots, v_k$

Possiamo scriverci le varie freq. relative

v_1	f_1
v_2	f_2
$\vdots$	$\vdots$
v_k	f_k

$$I = 1 - \sum_{j=1}^{k} f_j^2$$

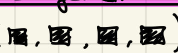
- $\forall j \quad 0 \leq f_j \leq 1$
- $\forall j \quad f_j \geq 0$
- $\sum_{j=1}^{k} f_j = 1 \longrightarrow \exists j' : f_{j'} > 0$

$$f_{j'}^2 > 0 \longrightarrow \sum_{j=1}^{k} f_j^2 > 0 \implies I = 1 - \sum_j f_j^2 < 1 \qquad \boxed{I < 1}$$

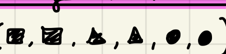
Possiamo dargli una limitazione anche a sx?

$$\forall j \quad 0 \leq f_j \leq 1 \longrightarrow f_j^2 \leq f_j$$
$$\Downarrow$$
$$\sum_j f_j^2 \leq \sum_j f_j = 1 \implies I = 1 - \sum_j f_j^2 \geq 0 \qquad \implies \boxed{0 \leq I < 1}$$

Eterogeneità minima: $I = 1 - \sum_j f_j^2 = 0$

()

Eterogeneità massima: $\forall j \quad f_j = \frac{1}{k} \qquad I = 1 - \sum_{j=1}^{k} \frac{1}{k^2} = 1 - \cancel{k}\frac{1}{k^{\cancel{2}}} = 1 - \frac{1}{k} = \frac{k-1}{k}$

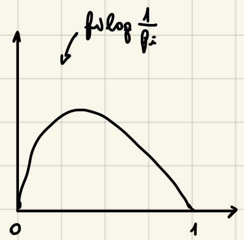
()

Variante dell'indice → **Indice di GINI NORMALIZZATO:** $I' = \frac{k}{k-1} \cdot I \qquad 0 \leq I' \leq 1$ in questo caso abbiamo sia una limitazione a sx sia a dx

Entropia

Dato un campione $\{x_1, \dots, x_m\}$ in cui occorrono k valori distinti $v_1, \dots, v_k$ e indicando con f_i la frequenza relativa dell'elemento v_i per $i = 1, \dots, k$ la quantità:

$$H = \sum_{i=1}^{k} f_i \log \frac{1}{f_i} = - \sum_{i=1}^{k} f_i \log f_i$$ è detta **indice di entropia del campione**



$H = \sum f_i \log \frac{1}{f_i} \geq 0$ ma quando un addendo è uguale a $\emptyset$? ossia $f_i \log \frac{1}{f_i} = \emptyset$ → $f_i = \emptyset$ oppure $\log \frac{1}{f_i} = \emptyset$, $\frac{1}{f_i} = 1$, $f_i = 1$

E per essere tutto questo pari a zero? $\forall i \quad f_i \log \frac{1}{f_i} = \emptyset$

cioè vuol dire che tutte le freq. relative devono valere $\emptyset$ e 1

Quando ciò si verifica? ⇒ Quando tutte le freq. relative sono a $\emptyset$ TRANNE UNA che si trova ad 1 (CASO DI ETEROGENEITÀ MINIMA)

Nel caso di ETEROGENEITÀ MAX?

$\forall i \quad f_i = \frac{1}{k} \longrightarrow H = \sum_i \frac{1}{k} \log k = \cancel{k} \cdot \frac{1}{\cancel{k}} \log k$

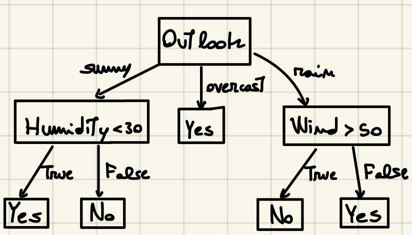
Quindi l'entropia ⟶ $\boxed{0 \leq H \leq \log k}$

Posso introdurre una normalizzazione dell'entropia
↓
$H' = \frac{H}{\log k}$ $\quad \boxed{0 \leq H' \leq 1}$

Se entropia calcolata con log in BASE 2
↓
misura in bit

Alberi di decisione ⟶ strumento seppur semplice, molto attuale nell'IA, Machine Learning
↓
mi permettono di formalizzare un processo decisionale
↓
dividendolo in tanti passi

Esempio: Lettura albero di decisione : Si parte dalla radice
↓
la radice come ogni nodo che NON è una foglia contiene una domanda
↓
In base alla risposta per quella domanda seguiamo un sottoalbero particolare
↓
Reitero fino al raggiungimento di una foglia
↓
contiene istruzioni su cosa fare come decisione



Il procedimento che a partire dai dati permette di ottenere un albero di decisione é basato sul concetto di eterogeneità

punto a fare domande che mi permettono di ottenere sottogruppi meno eterogenei possibili in modo da semplificarmi il lavoro e ridurre il num di domande da fare

Indici di concentrazione

si usa di solito per la ricchezza

Immaginiamo di avere n osservazioni $a_1, \dots, a_n$, ognuna di queste osservazioni mi va a dire quanto é ricco un elemento del campione considerato (immaginiamo pure di aver ordinato queste osservazioni da più piccola a più grande)

$$\bar{a} = \frac{1}{n} \sum_{i=1}^{n} a_i \qquad TOT = n\bar{a} = \sum_{i=1}^{n} a_i$$

Possiamo avere due situazioni: 1) caso di concentrazione minima $\longrightarrow$ Tutti gli elementi del campione assumono lo stesso valore $a_1 = a_2 = \dots = a_n = \bar{a}$

2) caso di concentrazione massima $\longrightarrow$ Tutti gli elementi del campione assumono valore $\emptyset$, a parte uno $a_1 = a_2 = \dots = a_{n-1} = \emptyset$ e $a_n = n\bar{a}$

Andiamo a definire due freq. interessanti

- frequenza relativa cumulata degli individui fino all'i-esima osservazione: $F_i = \frac{i}{n}$ per $i = 1, \dots, n$
- quantità (ricchezza) relativa cumulata fino all'i-esima osservazione: $Q_i = \frac{\sum_{k=1}^{i} a_k}{TOT}$

Proprietà:
- $0 \leq F_i \leq 1$ e $0 \leq Q_i \leq 1$
- $Q_i \leq F_i$ siccome le osservaz. sono ordinate in modo cresente
- $Q_i = F_i$ nel caso di concentrazione minima
- $Q_n = F_n$

Per $i = 1, \dots, n$ le coppie (F_i, Q_i) indicano che il $100 F_i\%$ della popolazione detiene $100 Q_i\%$ della quantità considerata

Indice di concentrazione di GINI: $G = \frac{\sum_{i=1}^{n-1} F_i - Q_i}{\sum_{i=1}^{n-1} F_i}$

Trasformazione dei dati

Partiamo da una tabella delle frequenze

VALORI	FREQUENZE
v_1	f_1
⋮	⋮
v_k	f_k

Sono interessato a considerare dei valori possibili per trasformare i miei valori

↓

delle funzioni $g: X \longrightarrow X'$ $\quad g(v_j) = v_j'$

sono interessato a ragionare su delle funzioni che non mi cambiano le frequenze ⟶ CONSIDERO TRASFORMAZ. che godono di INIETTIVITÀ

Considereremo in particolar modo trasformaz. lineari

↓

$g(x) = ax + b$

$a, b \in \mathbb{R}$

TRASLAZIONE

$k > 0$ $\quad v \longrightarrow v' = v - k$

$v \longrightarrow v' = v + k$

Perché traslare i dati? ⟶ Traslazione diventa interessante quando io ho dei dati che sono tutti molto grandi ma relativamente poco dispersi tra loro ⟶ mi permette di graficarli meglio

↓

ma

che effetto ha sugli indici?

- MEDIA, MEDIANA, QUANTILI traslati
- RANGE, RANGE INTERQUANTILE, VAR, DEV STD invariati

Altra trasformaz. lineare

↓

DILATAZIONE / CONTRAZIONE (SCALATURA)

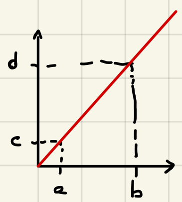
$h \in \mathbb{R}^+$ $\quad v \longrightarrow \frac{v}{h}$ $\quad$ Utile se dati fortemente dispersi ⟶ **effetto su indici?**

$h > 1$ CONTRAZIONE $\quad h < \min v_j \quad v_j' > 1$

$h < 1$ DILATAZIONE $\quad h > \max v_j \quad v_j' < 1$

- MEDIA, MEDIANA, QUANTILI scalati
- RANGE, IQR, DEV STD scalati
- VAR scalata di $1/h^2$

Supponiamo abbiamo dati che variano nell'intervallo $(a,b) \longmapsto (c,d)$



$$\frac{v'-c}{d-c} = \frac{v-a}{b-a}$$

$$v' = \frac{d-c}{b-a}(v-a) + c$$

CASI PARTICOLARI $(a,b) \mapsto (0,1)$

$\mapsto (-1,1)$

PARTICOLARE TRASFORMAZIONE ⟶ STANDARDIZZAZIONE

↓

Mi baso sulle osservazioni del mio campione $x_1, \dots, x_n$ $\quad \bar{x} \quad s_x$

$$g(v) = \frac{v - \bar{x}}{s_x} \Longrightarrow x_1', \dots, x_n'$$

$$\bar{x}' = \emptyset, \; s_{x'} = 1$$

TRASFORMAZIONE LOGARITMICA

$v \longmapsto v' = \log v$

Abbiamo visto come sia possibile costruire un <u>albero di decisione</u>

↓

serve per ottenere un classificatore: un metodo matematico che, dato un insieme di valori, mi individua un'osservazione

vogliamo che ci indichi l'appartenenza o meno di un individuo ad una certa classe

Esistono tanti modi per fare un classificatore

↓

noi abbiamo visto **un classificatore binario** → separiamo elementi che stanno in una classe da elementi che non stanno in quella classe (buoni / non buoni)

Come valutare un classificatore?

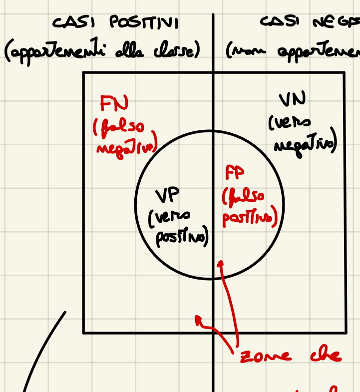
↓

<u>o meglio come performa</u>?
- Abbiamo le predizioni che un classificatore fa (quello che mi dice, può essere giusto o sbagliato)
- Abbiamo una verità di base (le etichette nel mio dataset)
- Non lavoriamo su singolo caso di test, ma su tanti casi

↓

dobbiamo trovare un modo per aggregare le risposte (corrette/sbagliate) che il nostro classificatore ci dà

<u>Ci possiamo trovare in 4 casi possibili</u>



predizioni relative al classificatore

Dentro al $\boxed{\text{cerchio}}$: mettiamo le predizioni positive

Fuori dal $\boxed{\text{cerchio}}$: mettiamo le predizioni negative

non c'è un'equipartizione delle zone

zone che rappresentano gli errori che il classificatore può fare

Cosa abbiamo fatto per valutare il nostro albero di decisione?

→ abbiamo guardato se ci beccava o no

PREDIZIONI \ ETICHETTA	+	−
+	VP	FP
−	FN	VN
	TP (tot pos)	TN (tot neg)

Matrice di CONFUSIONE (2x2)

Possiamo valutare classificatore dicendo quante volte sbaglia e quante ci becca

Accuratezza: somma d. numero di veri positivi (VP) e veri negativi (VN) diviso tot. positivi e tot negativi

$$\frac{VP + VN}{TP + TN}$$

↓

compresa tra Ø e 1

ma

alle volte usare accuratezza per valutare classificatore non è il massimo (es: malattie un FN è gravissimo)

Dobbiamo pesare diversamente errori sui positivi ed errori sui negativi

Creiamo così due nuovi indici

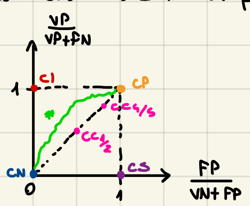
→ **SENSIBILITÀ**: $\frac{VP}{TP} = \frac{VP}{VP+FN}$

→ **SPECIFICITÀ**: $\frac{VN}{TN} = \frac{VN}{VN+FP}$

Consideriamo la coppia $(1\text{-SPECIFICITÀ}, \text{SENSIBILITÀ}) = \left(1 - \frac{VN}{TN}, \frac{VP}{TP}\right) = \left(\frac{FP}{TN}, \frac{VP}{TP}\right) = \left(\frac{FP}{VN+FP}, \frac{VP}{VP+FN}\right)$

↓

voglio visualizzare in un piano cartesiano



Es. di classificatori:
- classificatore costante positivo (CP)
- classificatore costante negativo (CN)
- classificatore ideale (CI)
- classificatore sbagliato (CS)
- classificatore casuale (CC) $(CC_{\frac{1}{2}}, CC_{4/5})$

CP → matrice di confusione

		ETICHETTA +	ETICHETTA −
PRED	+	TP	TN
	−	∅	∅

$SENS = \frac{TP}{TP} = 1 \qquad SPEC = \frac{VN}{TN} = \varnothing$

$(1-SPEC, SENS) = (1,1)$

CS → matrice di confusione

		ETICHETTA +	ETICHETTA −
PRED	+	∅	TN
	−	TP	∅

$SENS = \frac{0}{TP} = 0 \qquad SPEC = \frac{0}{TN} = 0$

$(1-SPEC, SENS) = (1, \varnothing)$

CN → matrice di confusione

		ETICHETTA +	ETICHETTA −
PRED	+	∅	∅
	−	TP	TN

$SENS = \frac{VP}{TP} = \varnothing \qquad SPEC = \frac{VN}{TN} = \frac{TN}{TN} = 1$

$(1-SPEC, SENS) = (\varnothing, \varnothing)$

$CC_{\frac{1}{2}}$ → matrice di confusione

		ETICHETTA +	ETICHETTA −
PRED	+	TP/2	TN/2
	−	TP/2	TN/2

$SENS = \frac{TP/2}{TP} = \frac{1}{2} \qquad SPEC = \frac{TN/2}{TN} = \frac{1}{2}$

$(1-SPEC, SENS) = (1/2, 1/2)$

CI → matrice di confusione

		ETICHETTA +	ETICHETTA −
PRED	+	TP	∅
	−	∅	TN

$SENS = \frac{TP}{TP} = 1 \qquad SPEC = \frac{TN}{TN} = 1$

$(1-SPEC, SENS) = (\varnothing, 1)$

$CC_{4/5}$ → matrice di confusione

		ETICHETTA +	ETICHETTA −
PRED	+	$4TP/5$	$4TN/5$
	−	$TP/5$	$TN/5$

$SENS = \frac{4TP/5}{TP} = \frac{4}{5} \qquad SPEC = \frac{TN/5}{TN} = \frac{1}{5}$

$(1-SPEC, SENS) = (4/5, 4/5)$

Ma è così male avere un classificatore sbagliato? E mi conviene discostarmi da esso?

↓

Non proprio è l'ESATTO OPPOSTO DELL'IDEALE, giro le etichette e da CS divento CI

Ipotizziamo ora di avere un classificatore che emette non positivo e negativo, ma un numero reale

↓

vogliamo convertirlo in positivo/negativo tramite procedimento di sogliatura

AL VARIARE DELLA SOGLIA VARIANO LE PERFORMANCE DEL MIO CLASSIFICATORE

↓

- SE SOGLIA AL MASSIMO → sempre negativo (CN)
- SE SOGLIA AL MINIMO → sempre positivo (CP)

DESCRIVE UNA CURVA

↓

* CURVA ROC (RECEIVER OPERATING CHARACTERISTICS)

Usato ai tempi dei sonar

Immaginiamo avere 2 diversi classificatori a soglia → difficile confrontarli

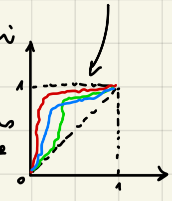
Allora calcoliamo curva ROC per i due classificatori

Possiamo definire una quantità (AUC) → l'area sotto la curva di ROC

$0 \leq AUC \leq 1$

più vicino all'ideale

ma così visivo è una tecnica QUALITATIVA



Anova (Analysis of Variance)

Immaginiamo di avere il nostro campione, le cui osservazioni sono peró suddivise in gruppi

Tot. di n osservazioni
G gruppi
ognuna delle mie n osservazioni appartiene a UNO E UNO SOLO di questi G gruppi

di ampiezza $n_1, \dots, n_G \Rightarrow \sum_{i=1}^{G} n_i = n$
Quante delle mie n osservazioni stanno nel primo gruppo

Perché fare ció? Per capire se c'é una qualche differenza tra i gruppi

Esempio: campo medico, n persone ad un gruppo do medicina vera ed altro gruppo (di controllo) do un placebo
voglio vedere poi risultati

Posso calcolare media campionaria dei due gruppi (tipo il colesterolo)

Analisi d. varianza da forza a questo modo di procedere rispetto ad un semplice confronto tra valori medi

Introduciamo notazione: x_i^g = i-esima osservazione nel g-esimo gruppo $x_1^1, x_2^1, \dots, x_{n_1}^1$ $\quad x_1^G, \dots, x_{n_G}^G$
$x_1^2, x_2^2, \dots, x_{n_2}^2$

$$\boxed{\bar{x}^g = \frac{1}{n_g}\sum_{i=1}^{n_g} x_i^g} \longrightarrow n_g\bar{x}^g = \sum_{i=1}^{n_g} x_i^g \longrightarrow \sum_{g=1}^{G} n_g\bar{x}^g = \sum_{g=1}^{G}\sum_{i=1}^{n_g} x_i^g \to \frac{1}{n}\sum_{g=1}^{G} n_g\bar{x}^g = \frac{1}{n}\text{ SOMMA DI TUTTE LE OSSERVAZIONI} = \bar{x}$$

aggiunta sommatoria membro a membro → se divido membro a membro cosa ottengo?

banalmente é la media campionaria di tutto il campione

$$\boxed{\bar{x} = \frac{1}{n}\sum_{g=1}^{G} n_g\bar{x}^g}$$

IMPORTANTI

$$\boxed{SS_T = \sum_{g=1}^{G}\sum_{i=1}^{n_g}\left(x_i^g - \bar{x}\right)^2}$$

SUM OF SQUARES → somma di tutti gli scarti quadratici dalla media campionaria → a partire da questa somma possiamo calcolare la varianza campionaria

$$\boxed{S^2 = \frac{SS_T}{n-1}}$$

$$SS_W = \sum_{g=1}^{G}\sum_{i=1}^{n_g}\left(x_i^g - \bar{x}^g\right)^2 \longrightarrow \text{VARIANZA ENTRO I GRUPPI} = \frac{SS_W}{n-G}$$

calcolo d. varianza separatamente all'interno di ogni gruppo

$$SS_B = \sum_{g=1}^{G} n_g\left(\bar{x}^g - \bar{x}\right)^2 \longrightarrow \text{VARIANZA TRA I GRUPPI} = \frac{SS_B}{G-1}$$

moltiplico per n_g poiché gruppi hanno ampiezza variabile

$SS_T = SS_W + SS_B$ ⟶ (DIMOSTRAZ. DISPENSE)

$$\underbrace{\frac{SST}{n-1}}_{(1)} = \frac{n-G}{n-1}\underbrace{\frac{SS_W}{n-G}}_{(2)} + \frac{G-1}{n-1}\frac{SS_B}{G-1}$$

Partiamo dall'ipotesi che la maggior parte delle volte si aspetta di non trovare

Esempio precedente: persone da curare con farmaco ⟶ gruppo persone curate e gruppo persone non curate

Ma perché fare ció? Se non ci sono diff. fondamentali tra i due gruppi — siamo interessati a questo di default — ma con questo approccio partiamo dall'ipotesi che la cura non abbia fatto effetto

calcolare varianza tot e varianza entro singolo gruppo mi aspetto di trovare la stessa cosa, ma se valuto separatamente (1) e (2) e le confronto (o sono approx uguali no diff gruppo tutto insieme)
(o sono diversi, non posso dire che non ci sono diff tra gruppi)

⇓

OVVIO SIAMO CONTENTI SE POI POSSIAMO DIRE CHE QUESTA IPOTESI NON É RAGIONEVOLE

Calcolo combinatorio

Idea base abbiamo un insieme finito di oggetti $\longrightarrow$ e dei criteri che mi permettono di raggruppare questi oggetti

Voglio andare a contare il numero diverso di raggruppamenti possibili

Principio di enumerazione

$\longrightarrow$ ho tanti esperimenti con diversi esiti

I esperimento con m esiti
II esperimento con n esiti

I esperimento: 1 2 ... m
II esperimento: 1 2 ... n

I esp ha dato come esito il primo degli m esiti e il II esp ha dato come esito il primo degli n esiti $\longrightarrow (1,1)$

per ogni posto della matrice avrò diverse coppie di esiti congiunti

MA IN QUANTI MODI DIVERSI POTRÒ OTTENERE QUESTI ESITI CONGIUNTI?

$m \cdot n$ COPPIE ORDINATE POSSIBILI DI ESITI

in quanti modi diversi si possono presentare i risultati di questi 2 esperimenti considerato congiuntamente

Immaginiamo di avere n esperimenti

m_i = # esiti dell'esperimento i-esimo $\Rightarrow$ in quanti modi diversi potrò ottenere delle tuple ordinate che mi indicano esiti fino all'n-esimo

$$m_1 \cdot m_2 \cdot \ldots \cdot m_n = \prod_{i=1}^{n} m_i$$

Disposizioni

Disposizioni: Tipo di raggruppamento in cui io ho fissato un certo numero di posti e voglio mettere i miei oggetti in fila in questi posti (NOTA: non è detto che num di posti è uguale a numero di oggetti)

ordine conta $\longrightarrow$ è una sequenza ordinata $\Rightarrow$ DISPOSIZIONI DI n OGGETTI IN k POSTI

ma

possiamo scegliere uno stesso oggetto più volte oppure no

si parlerà di
- DISPOSIZIONI CON RIPETIZIONE
- DISPOSIZIONI SENZA RIPETIZIONE

Disposizioni con ripetizione

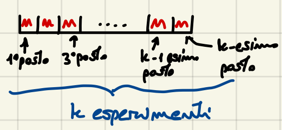
$\{a, b, c, d\}$ $\longrightarrow$ non ci interessa considerare la natura specifica degli elementi, ci basta sapere solo quanti sono

Es: dispos. con ripet. in 2 posti

aa ba ...
ab bb
ac bc
ad bd

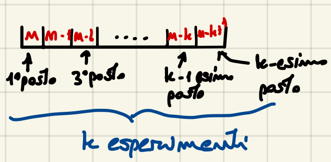
Ma quanti sono?

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



Disposizioni senza ripetizioni

$\{a, b, c, d\}$



Quanti sono? $d_{n,k} = n(n-1)(n-2)\ldots(n-k+1)$

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

Permutazioni

Se noi abbiamo una sequenza di oggetti, una sua permutazione non è altro che quello che si può ottenere prendendo gli oggetti e cambiando in qualche modo l'ordine

↓

facciamo riferimento ad un solo indice non più a 2, parleremo quindi di:

Permutazioni di n oggetti → diciamo che equivale a fare una disposizione senza ripetiz. di n oggetti in n posti

$P_n = d_{n,n} = n!$ | n | n-1 | ... | ... | ... | 2 | 1 |

Perché $0! = 1$? Possiamo vederlo come, devo sistemare ∅ elementi e non fare nulla è un solo modo per sistemarli

In info $n! = n(n-1)!$ nella ricorsione

$1 = 1! = 1 \cdot 0! = 0!$

Combinazioni di n oggetti in k

↓

ordine non conta a differenza delle disposizioni

$\{a,b,c,d\}$ $\{a,b\}, \{a,c\}, \{a,d\}$

$\{a,b,c\}$ $d_{3,2}$ $\boxed{a,b \; b,a}$ $\boxed{a,c \; c,a}$ $\boxed{b,c \; c,b}$

$\{a,b\}$ $\{a,c\}$ $\{b,c\}$

$C_{3,2} = \frac{d_{3,2}}{2}$

$C_{n,k} = \frac{d_{n,k}}{k!} = \frac{n!}{(n-k)!} \cdot \frac{1}{k!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$ **COEFFICIENTE BINOMIALE** → mi dice in quanti modi posso estrarre dei sottoinsiemi di k elementi da un insieme che ne contiene n

ma

che succede se k più grande di n?

1) Come faccio a tirare fuori più elementi di quanti ne ho
2) A livello di conti c'è un problema

Esempio: vogliamo mettere in fila dei libri su uno scaffale

4 libri matematica (M)
3 " chimica (C)
2 " storia (S)
1 " lingue (L)

$4M + 3C + 2S + 1L = 10$ libri

↓

in quanti modi diversi posso affiancare questi 10 libri su uno scaffale?

Permutazioni → non va bene se no non avremmo messo l'info che specifica il tipo di libri

. Libri di matematica in quanti modi posso ordinarli? $4!$
. " " chimica " " " " " ? $3!$
. " " storia " " " " " ? $2!$
. " " lingue " " " " " ? $1!$

Cose sbaglio?

$M_1 M_2 M_3 M_4$ $C_1 C_2 C_3$
$M_2 M_4 M_1 M_3$ $C_3 C_1 C_2$
⋮ ⋮

Se la mia idea è di permutare i libri in un qualunque modo → $10!$ ok

ma

Se voglio libri M tutti vicini tra loro, libri C vicini, ecc... → introduciamo ordinamento di secondo livello quello tra discipline dei libri → $(4!\,3!\,2!\,1!)\,4! = 6912$

M C S L — TRA DISCIPLINE

Anagrammi: SEDIA $5!$ "anagrammi" TAPPO $\frac{5!}{2!} = 60$

MAMMA $\frac{5!}{3!\,2!}$

Probabilità → quantificazione dell'incertezza di un evento

→ un qualcosa che può verificarsi oppure no, e io non so a priori se si verifica o no

Ci sono tanti modi di andare a fare una formalizzazione o meglio una quantificazione di questo concetto di incertezza

→ normalmente si fissa una scala possibile di valori

→ estremi coincidono con certezza totale che succede una cosa, e certezza totale che NON succede una cosa

Ipotesi, ragioniamo in termini di percentuali

Es: Geologo dice: Prob. 60% di trovare petrolio

POSSO INTERPRETARE QUESTA COSA IN TANTI MODI

ACCEZIONE SOGGETTIVA

ACCEZIONE FREQUENTISTA → Si basa su ragionamento a posteriori
Su 100 volte che ho scavato 60 volte ho trovato petrolio

Importante avere strumenti formali matematici che ci permettono di trattare con delle situazioni complesse

→ Usare degli assiomi (Teoria assiomatica)

Prima bisogna introdurre un po' di formalizzazione

Teoria degli insiemi

Insieme: aggregazione di individui che soddisfano una certa regola

vogliamo usarle per modellare l'esito di un esperimento che è incerto

→ ogni volta che osservo l'esperimento viene fuori un esito diverso

Ci possiamo concentrare su tutti gli esiti possibili, diversi tra loro di un esperimento e metterli all'interno di un particolare insieme

→ Ω insieme universo (insieme degli esiti possibili / spazio degli eventi)

Es:
- determinazione genere di nascituro $\Omega = \{M, F\}$
- corsa di 7 cavalli $\Omega = \{\text{permutazioni di 7 elementi}\}$ $|\Omega| = 7!$
- determinazione dosaggio minimo di farmaco $\Omega = \mathbb{R}^+ = (0, +\infty)$

Prendiamo generico elemento dentro Ω (un esito possibile)

$\forall \omega \in \Omega$ EVENTO ELEMENTARE/ESITO $\{\omega\}$

Es:
- $\{F\}$
- $\{(1,2,3,4,5,6,7)\}$
- $\{42\}$

Evento $E \subseteq \Omega$

Insieme $\Omega \longrightarrow$ evento certo

Insieme vuoto $\longrightarrow$ evento impossibile

<u>Esempio</u>: . $\{F\}$ è un evento, perché è un sottoinsieme di Ω

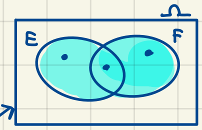
. $\{(1,2,3,4,5,6,7),(1,3,2,4,5,6,7)\}$

oppure mi interessa che il cavallo 3 vinca

considero le permutazioni con cavallo 3 primo $6!$ permutaz. (non 7 perché

3 è fissato in prima posizione)

. $[3,7]$ dosaggio minimo farmaco



. $E \cup F$ (unione degli eventi E ed F) $\longrightarrow$ Quando si verifica? Quando almeno uno tra elemento E ed elemento F si verifica

. $E \cap F$ (intersezione d. eventi E ed F) $\longrightarrow$ Quando si verifica? Sse entrambi gli elementi E ed F si verificano

. $E \cap F = \emptyset$ <u>insiemi disgiunti</u>

$\downarrow$

EVENTI DISGIUNTI / MUTUAMENTE ESCLUSIVI

. Ω evento certo

. $\emptyset, \{\}$ evento impossibile

. $E \setminus F \longrightarrow$ evento che si verifica quando si verifica E e NON si verifica F

. $\Omega \setminus E = \bar{E} = E^c \longrightarrow$ evento che si verifica quando NON si verifica E

. $E \subseteq F$ $\forall x \in \Omega$ $x \in E \longrightarrow x \in F$ $\Rightarrow$ se si verifica E allora si verifica F

. $E \subseteq F \wedge F \subseteq E \longrightarrow E = F$

. $E_1 \cup E_2 \cup \ldots \cup E_n = \bigcup_{i=1}^{n} E_i$ $\quad \bigcap_{i=1}^{n} E_i$

<u>Per op. di intersezione e unione valgono:</u>

$\rightarrow$ **COMMUTATIVA** $\forall E,F \subseteq \Omega$ $\quad E \cup F = F \cup E$, $E \cap F = F \cap E$

$\rightarrow$ **ASSOCIATIVA** $\forall E,F,G \subseteq \Omega$ $\quad (E \cup F) \cup G = E \cup (F \cup G)$ (vale anche per intersezione)

$\rightarrow$ **DISTRIBUTIVA** $\forall E,F,G \subseteq \Omega$ $\quad E \cup (F \cap G) = (E \cup F) \cap (E \cup G)$

$E \cap (F \cup G) = (E \cap F) \cup (E \cap G)$

$\rightarrow$ **LEGGI DI DEMORGAN** $\overline{E \cup F} = \bar{E} \cap \bar{F}$ $\quad \overline{E \cap F} = \bar{E} \cup \bar{F}$

$\overline{E \cup F} \subseteq \bar{E} \cap \bar{F}$

$\bar{E} \cap \bar{F} \subseteq \overline{E \cup F}$

Es: $x \in \overline{E \cup F} \longleftrightarrow x \notin E \cup F \longleftrightarrow x \notin E \wedge x \notin F \leftrightarrow x \in \bar{E} \wedge x \in \bar{F}$

$\longleftrightarrow x \in \bar{E} \cup \bar{F}$

Concetto di probabilità formalizzato tramite una funzione che mappa eventi in numeri reali

Dominio

Algebra degli eventi, idealmente vogliamo mettere insieme tutti gli eventi a cui siamo interessati

$\mathcal{A} = \{E_i \subseteq \Omega\}$ (insieme di eventi/collezione di eventi) ma la dobbiamo definire effettivamente attraverso una serie di proprietà che questa collezione deve soddisfare

- $\Omega \in \mathcal{A}$ (non posso fare a meno tra tutti gli eventi che metto in un'algebra dell'evento certo)
- $\forall E \subseteq \Omega \quad E \in \mathcal{A} \longrightarrow \bar{E} \in \mathcal{A}$ (chiusura rispetto alla complementazione, se voglio ragionare in term. di un evento che si verifichi, devo anche poter ragionare in termini del fatto che quell'evento non si verifichi)
- $\forall E, F \subseteq \Omega \quad E \in \mathcal{A}, F \in \mathcal{A} \longrightarrow E \cup F \in \mathcal{A}$ (chiusura rispetto all'unione, se voglio poter ragionare in termini di due eventi separatamente devo anche poter ragionare in termini dell'evento che è la loro unione)

si estende ad un numero finito di eventi

Da queste 3 proprietà non posso estenderlo al caso di successioni infinite di eventi $\longrightarrow$ dovrei considerare oggetti che non sono algebre degli eventi $\longrightarrow$ se dovesse valere che:

$\boxed{E_1, \dots, E_n, \dots \quad \forall i \, E_i \in \mathcal{A} \longrightarrow \bigcup_{i=1}^{\infty} E_i \in \mathcal{A}}$

si dice che:

$\mathcal{A}$ è una σ-algebra degli eventi

Es di algebra: $\mathcal{A} = \{\Omega, \emptyset\}$

(ma) se volessi un qualcosa di più, considero $\mathcal{P}(\Omega)$ (insieme d. parti di Ω)

soddisfa ovviamente le 3 proprietà

Se ho uno spazio degli eventi FINITO

NON HO MODO DI USARE UN'ALGEBRA PIÙ PICCOLA DELL'INSIEME DELLE PARTI

Se Ω è infinito

il mio insieme delle parti potrebbe contenere più cose di quelle che mi interessano

Funzione di probabilità

$P: \mathcal{A} \longrightarrow \mathbb{R}$

Teoria assiomatica d. probabilità:

A1: $\forall E \in \mathcal{A} \quad 0 \leq P(E) \leq 1$

A2: $P(\Omega) = 1$

A3: $\forall E_1, \dots, E_n \quad \forall i \; E_i \in \mathcal{A}$ A DUE A DUE DISGIUNTI $\Longrightarrow P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$

$\forall i \neq j \quad E_i \cap E_j = \emptyset$

ASSIOMI DI KOLMOGOROV

Tali assiomi catturano perfettamente il concetto di prob. in senso frequentista

freq. relative

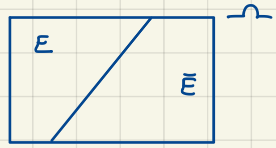
Teorema: $\forall E \in \mathcal{A} \quad P(\bar{E}) = 1 - P(E)$

$\downarrow$

Usiamo A2 $\quad 1 = P(\Omega) = P(E \cup \bar{E})$

$E \cap \bar{E} = \emptyset$

$E \cup \bar{E} = \Omega$



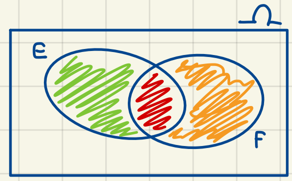
$\downarrow$ applico A3

$1 = P(E) + P(\bar{E}) \implies P(\bar{E}) = 1 - P(E)$

Teorema 2: $\forall E, F \in \mathcal{A} \quad P(E \cup F) = P(E) + P(F) - P(E \cap F)$

non necessariamente disgiunti

ci vuole questo nel caso non siano disgiunti



$\underbrace{E \cap \bar{F} \quad E \cap F \quad \bar{E} \cap F}_{\text{disgiunti a due a due}}$

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

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

$\downarrow$

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

$\downarrow$ 3ASSIOMA

$= \underbrace{P(E \cap \bar{F}) + P(E \cap F)}_{A3\ =} + P(\bar{E} \cap F)$

MAGHEGGIO ALGEBRICO

$P((E \cap \bar{F}) \cup (E \cap F)) = P(E) + \underbrace{P(\bar{E} \cap F) + P(E \cap F)}_{A3\ =} - P(E \cap F)$

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

$\Downarrow$

$P(E \cup F) = P(E) + P(F) - P(E \cap F)$

Teorema 3: $P(\emptyset) = 0$

A2 $P(\Omega) = 1$

per il

Teorema 1 $\longrightarrow P(\bar{E}) = 1 - P(E)$

$\Longrightarrow P(\emptyset) = P(\bar{\Omega}) = 1 - P(\Omega) = 0$

La freq. approssima la probabilità se si riferiscono allo stesso evento

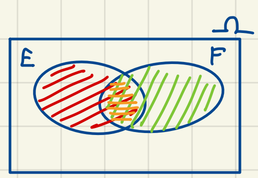
Esempi: Immaginiamo che 28% maschi americani fuma sigarette E

7% " " fuma sigaro F

5% " " fuma entrambi $E \cap F$

percentuale di non fumatori?

$P(E) = 0.28$



$P(F) = 0.07$

$P(E \cap F) = 0.05$

$P(\text{non fuma}) = 1 - P(\overline{\text{non fuma}})$

$= 1 - P(\text{fuma})$

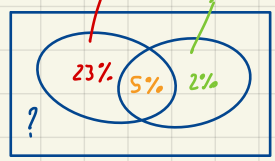
$= 1 - P(E \cup F)$

$= 1 - (P(E) + P(F) - P(E \cap F))$

$= 1 - (0.28 + 0.07 - 0.05) = 0.7$

28 - 5

7 - 5



$100 - (23+5+2) = 70\%$

spazio delle probabilità

è una Terna

non è un concetto unico, a partire da uno stesso spazio probabilizzabile io posso costruire delle diverse funzioni di probabilità

$\Omega, \mathcal{A}$

questa coppia prende nome di spazio probabilizzabile → ci posso costruire una funzione di probabilità P

$$\Downarrow$$

$$(\Omega, \mathcal{A}, P)$$

Prima categoria di spazi d. probabilità che vediamo

↓

Spazi equiprobabili ⟶ (Tutti eventi elementari hanno stessa probabilità costante p)

→ evento che corrisponde ad un insieme singoletto e quindi un evento che contiene uno e un solo esito

Spazio d. eventi finito (Ω contiene N esiti)

$$\Omega = \{e_1, e_2, \dots, e_N\}$$

$$P(\{e_1\}) = P(\{e_2\}) = \dots = P(\{e_N\}) = p$$

N (parametro)

disgiunti

$$\Omega = \bigcup_{i=1}^{N} \{e_i\}$$

$$\underbrace{1 = P(\Omega)}_{A2} = \underbrace{P\left(\bigcup_{i=1}^{N}\{e_i\}\right) = \sum_{i=1}^{N} P(\{e_i\})}_{A3} = \sum_{i=1}^{N} p = Np \quad \Rightarrow \quad 1 = Np \qquad p = \frac{1}{N}$$

E se omega fosse infinito? Non possono esistere spazi equiprobabili con numero infinito di esiti

↓

uscirebbe un $\lim_{n\to\infty} \Rightarrow p \to 0$ ma ciò è impossibile perché la loro somma dovrebbe fare 1 poiché $P(\Omega) = 1$ VIOLEREMMO UN' ASSIOMA DI KOLMOGOROV

Potrei essere interessato a calcolare la probabilità di un generico evento nella mia algebra

$$\forall E \in \mathcal{A} \longrightarrow E \subseteq \Omega \longrightarrow \exists k \le n : \ E = \{e_{i_1}, \dots, e_{i_k}\}$$

$$E = \{e_{i_1}\} \cup \dots \cup \{e_{i_k}\} \qquad P(E) = \sum_{j=1}^{k} P(\{e_{i_j}\}) = \sum_{j=1}^{k} \frac{1}{N} = \frac{k}{N} = \frac{|E|}{N}$$

#casi favorevoli

#casi possibili

Esercizio:

urna: 6 palle bianche estraggo 2 palle senza reimmissione
5 palle nere

#casi possibili = numero di modi in cui posso estrarre due palle da un'urna che ne contiene 11

↓

disposiz. senza ripetiz. = $d_{11,2} = 11 \cdot 10$

#casi favorevoli

B N
$6 \cdot 5 = 30$

N B $+ = 60$
$5 \cdot 6 = 30$

$$P(\text{due colori diversi}) = \frac{60}{110}$$

Es: commissione di 5 persone, vanno scelte da 1 Team di esperti

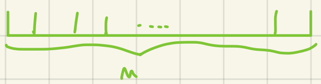
$\downarrow$

6 Uomini + 9 Donne

$$P(3U+2D) = \frac{\binom{6}{3}\cdot\binom{9}{2}}{\binom{15}{5}} = \frac{240}{1001}$$

$\binom{15}{5} \rightarrow$ combinazioni

Es: probl. del compleanno



casi possibili $= D_{365,m}$

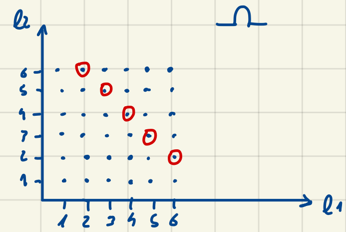
casi favorevoli $= d_{365,m}$

$$P(\text{no compleanni comuni}) = \frac{d_{365,m}}{D_{365,m}} = \frac{365\cdot 364\cdot\ldots\cdot 365-m+1}{365^m} = \frac{365!}{m!\,365^m}$$

$$P(\text{compleanno comune}) = 1 - P(\text{no compleanni comuni})$$

Es: lanciamo 2 dati bilanciati

ℓ_1, ℓ_2 = i risultati $\qquad P(\ell_1+\ell_2 = 8) = \frac{5}{36}$



i dadi li sto considerando distinguibili

$(1,6) \neq (6,1)$ per me in questo caso

Probabilità condizionata

$\downarrow$

Il fatto di conoscere un'informazione parziale cambia le probabilità che noi possiamo dare ad un evento?

$\rightarrow$ Dipende $\rightarrow$ iniziamo a formalizzare

È un modo per andare a parlare della prob. di eventi

ma

sapendo che qualcosa è successo

Esempio: lancio 2 dadi bilanciati

$\downarrow$

equiprobabilità

$F = \{3 \text{ al primo dato}\}$

$E = \{\text{somma dei dati} = 8\}$

Ipotesi:

$\downarrow$

so che primo dado $= 3$ $\xrightarrow[\text{(Tengo conto di sx)}]{\text{(sapendo questo)}}$ $P(\text{somma} = 8) = ? = \frac{1}{6}$

Casi possibili: $(3,1)\ (3,2)\ (3,3)\ (3,4)\ (3,5)\ (3,6)$ casi possibili

caso favorevole: $(3,5)$

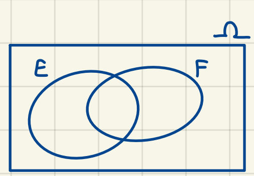
Questa def richiede $P(F) \neq 0$

$P(F) = \frac{1}{6}$

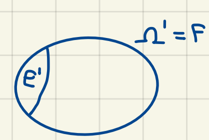
$P(E\cap F) = P(\{3,5\}) = \frac{1}{36}$

$P(E|F) = \frac{1/36}{1/6} = \frac{1}{6}$

Lo scriveremo come $P(E|F) = \dfrac{P(E\cap F)}{P(F)}$



da qui è come se il mio spazio di prob. si restringesse, lo spazio campionario non è più Ω, ma F



$E, F \in \mathcal{A}$

E: evento condizionato

F: evento condizionante

Esempio: confezione di 40 pennarelli (diversi): - 5 guasti
- 10 difettosi
- 25 accettabili

E = {accettabile}

F = non guasto = {difettoso ∪ accettabile}

↓

quando calcolo intersezione un pennarello non può essere simultaneamente accettabile e difettoso

$$P(E|F) = \frac{P(E\cap F)}{P(F)} = \frac{P(\text{accettabile})}{1 - P(\text{guasto})} = \frac{25/40}{1-5/40} = \frac{25/40}{35/40} = \frac{5}{7}$$

Modo alternativo di calcolare, considero spazio campionario aggiornato sapendo pennarello non guasto ⟶ 10 D, 25 A

$$P(\text{accettabile}) = \frac{25}{35} = \frac{5}{7}$$

Esempio 2: $\Omega = \{(m,m),(m,f),(f,m),(f,f)\}$

$A = \{2 \text{ figli maschi}\} = \{(m,m)\}$

$B = \{\text{almeno un figlio maschio}\} = \{(m,m),(m,f),(f,m)\}$

$$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{1/4}{3/4} = \frac{1}{3}$$

Dalla formula della probabilità condizionata $P(A|B) = \frac{P(A\cap B)}{P(B)} \longrightarrow P(A\cap B) = P(A|B)\cdot P(B)$ REGOLA DI FATTORIZZAZIONE

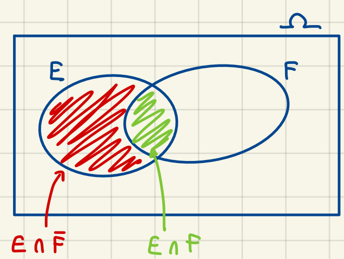
Esempio:

U = {apre nuovo ufficio} $P(U) = 0.3$

M = {sig. x diventa direttore del nuovo ufficio} $P(M|U) = 0.6$

$$P(U\cap M) = P(U)\cdot P(M|U) = 0.18$$

Regola d. fattorizzazione è alla base di un risultato teorico interessante



$E, F \in \mathcal{A}$

$(E\cap\bar{F})\cap(E\cap F) = (E\cap E)\cap(F\cap\bar{F}) = E\cap\emptyset = \emptyset$ sono disgiunti

Se li unisco? $(E\cap\bar{F})\cup(E\cap F) = E\cap(F\cup\bar{F}) = E\cap\Omega = E$

Posso applicare A3 ho detto che lo posso esprimere come unione di due eventi disgiunti

$$\begin{aligned}
P(E) &= P((E\cap\bar{F})\cup(E\cap F)) \\
&\overset{A3}{=} P(E\cap\bar{F}) + P(E\cap F) \\
&\overset{\text{REG DI FATTORIZZ.}}{=} P(E|\bar{F})P(\bar{F}) + P(E|F)P(F) \\
&= P(E|F)P(F) + P(E|\bar{F})P(\bar{F}) \\
&= P(E|F)P(F) + P(E|\bar{F})(1-P(F))
\end{aligned}$$

→ Teorema delle probabilità totali

↓

per applicarlo devo soddisfare delle ipotesi

→ $P(F) \neq 0$ e $P(\bar{F}) \neq 0$

Perché fare ciò? A volte più easy calcolare prob. condizionate che quelle incondizionate

Esempio: Due Tipi di clienti nelle assicurazioni

inclini a incidenti
$P(\text{incidente}) = 0.4$
30 % clienti inclini

non inclini a incidenti
$P(\text{incidente}) = 0.2$
70 % clienti non inclini

$A = \{\text{incidente nel prox anno}\}$
$H = \{\text{cliente incline agli incidenti}\}$

$P(A|H) = 0.4 \qquad P(H) = 0.3$
$P(A|\bar{H}) = 0.2 \qquad P(\bar{H}) = 0.7$

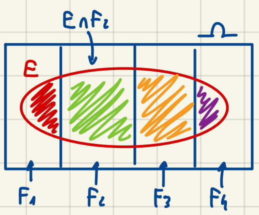
$P(A) = ?$

$P(A) = P(A|H) \cdot P(H) + P(A|\bar{H}) \cdot P(\bar{H})$
$\quad = 0.4 \cdot 0.3 + 0.2 \cdot 0.7 = 0.26$

Estensione a forma generale Teorema d. prob. totali

Tiriamo dentro il concetto di partizione d. spazio campionario (partizione di Ω)

insieme di insiemi $\quad F_1, \dots, F_n \subseteq \Omega \quad \bigcup_{i=1}^{n} F_i = \Omega \quad \forall i \neq j \;\; F_i \cap F_j = \emptyset$



$\bigcup_{i=1}^{n} (E \cap F_i) = E \cap \bigcup_{i=1}^{n} F_i = E \cap \Omega = E$

$(E \cap F_i) \cap (E \cap F_j) =$ con $(i \neq j)$
$= (E \cap E) \cap (F_i \cap F_j) = E \cap \emptyset = \emptyset$

$P(E) = P(\bigcup_{i=1}^{n} (E \cap F_i)) = \sum_{i=1}^{n} P(E \cap F_i) = \sum_{i=1}^{n} P(E|F_i) P(F_i)$ FORMA PIÚ GENERALE DEL TEOREMA DELLE PROBABILITÁ TOTALI

Ipotesi da Tenere in considerazione

- $\forall i \;\; P(F_i) \neq 0$
- $F_1, \dots, F_n$ partizioni di Ω

Esempio: fabbrica con 3 macchinari:

	% pezzi difettosi	% pezzi prodotti sul Totale
A	2	60
B	3	30
C	4	10

$P(D) = ?$

D = pezzo difettoso
A = pezzo prodotto dal macchinario A
B = 〃 〃 〃 〃 B
C = 〃 〃 〃 〃 C

$P(D) = P(D|A) \cdot P(A) + P(D|B) \cdot P(B) + P(D|C) \cdot P(C)$
$\quad = 0.02 \cdot 0.6 + 0.03 \cdot 0.3 + 0.04 \cdot 0.1 = 0.025$

$P(D|A) = 0.02 \qquad P(A) = 0.6$
$P(D|B) = 0.03 \qquad P(B) = 0.3$
$P(D|C) = 0.04 \qquad P(C) = 0.1$

Applicazione d. forma generale del Teorema delle prob Totali ai sondaggi

$\downarrow$

100 studenti — sfera da $\# \leq 70 \longrightarrow$ domanda delicata

sfera da $\# > 70 \longrightarrow$ domanda di controllo

25 sí

$$P(\text{sí}) = \underbrace{P(\text{sí} \mid \text{domanda delicata})}_{=\,?} \cdot \overbrace{P(\text{domanda delicata})}^{0.7} + \underbrace{P(\text{sí} \mid \text{domanda di controllo})}_{0.5} \cdot \overbrace{P(\text{domanda di controllo})}^{0.3}$$

$P(\text{sí}) \sim 0.25$

$$P(\text{sí} \mid \text{domanda delicata}) \approx \frac{0.25 - 0.5 \cdot 0.3}{0.7} \approx 0.14$$

Esempio: E = esito positivo

M = sono malato

se malato, il 99% d. volte il test "ci becca"

1% falsi positivo

cosé $P(M)$? prob. che prendendo un individuo sia malato (Tasso incidenza malattia) $P(M) = 0.005$

$P(E \mid M) = 0.99$ VP

$P(E \mid \bar{M}) = 0.01$ FP

$P(\bar{E} \mid \bar{M}) = 1 - P(E \mid \bar{M}) = 0.99$ valgono le proprietà

REGOLA FATTORIZZ.

TEO PROB. TOT SU E

$$P(M \mid E) = ? = \frac{P(M \cap E)}{P(E)} = \frac{P(E \cap M)}{P(E)} = \frac{P(E \mid M)\,P(M)}{P(E)} = \frac{P(E \mid M)\,P(M)}{P(E \mid M)P(M) + P(E \mid \bar{M}) \cdot P(\bar{M})} = \frac{0.99 \cdot 0.005}{0.99 \cdot 0.005 + 0.01 \cdot 0.995} \approx 0.3322$$

Teorema di Bayes

calcolo alternativo

$P(M) = 0.005 = \frac{1}{200} = 0.5\%$ su 200 persone

1 malato $\longrightarrow 1 \cdot 0.99$ esito positivo

199 sani $\longrightarrow 199 \cdot 0.01 = 1.99$ esiti pos.

$$P(M \mid E) = \frac{0.99}{0.99 + 1.99} \approx 0.3322$$

generalizzato

Se mi trovo nelle stesse ipotesi del Teorema d. prob. Totali

- $F_1, \dots, F_n$ partizioni di Ω
- $\forall i \; P(F_i) \neq 0$
- $E \in \mathcal{A}$

$$P(F_j \mid E) = \frac{P(E \mid F_j) \cdot P(F_j)}{\sum_{i=1}^{n} P(E \mid F_i)\,P(F_i)}$$

$$= \frac{P(F_j \cap E)}{P(E)} = \frac{P(E \mid F_j)\,P(F_j)}{\sum_{i=1}^{n} P(E \mid F_i)\,P(F_i)}$$

Esempio: $P(C) = 0.6$ sospetto colpevole $\longrightarrow$ quantificaz. soggettiva

$\downarrow$

Teo di Bayes strum. che si può utilizzare per andare a rivedere questa quantificazione dell'incertezza alla luce di nuove info

$P(M \mid C) = 1$

$P(M) = 0.2$

$$P(C \mid M) = ? = \frac{P(M \mid C)\,P(C)}{\underbrace{P(M \mid \bar{C})\,P(\bar{C}) + P(M \mid C) \cdot P(C)}_{\approx P(M)}} \approx 0.68$$

Esempio: aereo scomparso 3 zone stessa prob.

$P(R_1) = 1/3$

$P(R_2) = 1/3$

$P(R_3) = 1/3$

α_i = P(non trovare aereo se cerco nella zona i)

E = aereo non trovato in zona 1

$P(E \mid R_1) = \alpha_1 \qquad P(E \mid R_2) = 1 \qquad P(E \mid R_3) = 1$

$$P(E) = \sum_{i=1}^{3} P(E \mid R_i)\,P(R_i) = \frac{\alpha_1}{3} + \frac{2}{3}$$

$$P(R_1 \mid E) = \frac{P(E \mid R_1)\,P(R_1)}{P(E)} = \frac{\alpha_1/3}{(\alpha_1+2)/3} = \frac{\alpha_1}{\alpha_1+2}$$

$$P(R_2 \mid E) = \frac{1/3}{(\alpha_1+2)/3} = \frac{1}{\alpha_1+2} = P(R_3 \mid E)$$

Prob condizionata $\longrightarrow$ imparentata con classificatori

M: supereroe marvel
N: occhi neri

$$P(M|N) = \frac{P(N|M) \cdot P(M)}{P(N)}$$

C: colore dei capelli
O: colore degli occhi

m·n prob

$$P(M \mid C = c \cap O = \sigma) = \frac{\boxed{P(C=c \cap O=\sigma \mid M)}\, P(M)}{P(C=c \cap O=\sigma)}$$

assumiamo che $= \dfrac{\boxed{P(C=c|M) \cdot P(O=\sigma|M)} \cdot P(M)}{P(C=c \cap O=\sigma)}$

assunzione da prendere con le pinze ingenua

m+n prob

la facciamo perché se ho
m colori occhi
n colori capelli $\Rightarrow$ m·n prob.

OTTENIAMO UN CLASSIFICATORE NAIVE BAYES

---

m possibili classi
$Y = y_k$

n attributi
$X_i = x_i$

$$P(Y=y_k \mid X_1 = x_1 , \ldots , X_n = x_n) = \frac{P(X_1=x_1,\ldots,X_n=x_n \mid Y=y_k)\, P(Y=y_k)}{P(X_1=x_1,\ldots,X_n=x_n)} \approx \frac{\prod_{i=1}^{n} P(X_i = x_i \mid Y=y_k)\, P(Y=y_k)}{\underbrace{P(X_1=x_1,\ldots,X_n=x_n \mid Y=y_k)}_{\text{vogliamo togliere questo}}}$$

virgole = intersez.

$$\text{argmax}_k \frac{\prod_{i=1}^{n} P(X_i = x_i \mid Y=y_k)\, P(Y=y_k)}{\underbrace{P(X_1=x_1,\ldots,X_n=x_n \mid Y=y_k)}_{\text{non dipende } k}} \longrightarrow \text{argmax}_k \prod_{i=1}^{n} P(X_i = x_i \mid Y=y_k)\, P(Y=y_k)$$

Abbiamo visto la prob. condizionata $P(E|F)$

ma

cosa succede $P(E|F) = P(E)$? vuol dire che non sappiamo niente che ci aiuti

Tale situazione è detta condizione di indipendenza tra E ed F ($E \perp F$)

ma

se $P(E|F) = P(E)$ quindi scambiando i due insiemi non cambia nulla

Es: F = lancio dado bilanciato e ottengo 3
E = domani piove

$$\underbrace{\frac{P(E \cap F)}{P(F)}}_{P(E|F)} = P(E) \Rightarrow \boxed{P(E \cap F) = P(E) \cdot P(F)}$$

INDIPENDENZA tra E ed F

Es: mazzo di 52 carte (mischiato): pesco una carta

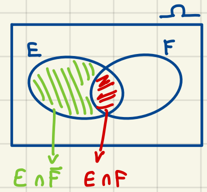
A = pesco un asso   C = pesco cuori

$P(A) = \frac{4}{52}$ $P(C) = \frac{13}{52}$ $A \cap C$ = pesco asso di cuori

$$P(A \cap C) = \frac{1}{52} = \frac{4}{52} \cdot \frac{13}{52} = P(A) \cdot P(C)$$

Proprietà: E, F indipendenti $\longrightarrow E, \bar{F}$ indipendenti

$$E = (E \cap F) \cup (E \cap \bar{F})$$
$$P(E) = P(E \cap F) + P(E \cap \bar{F})$$
$$P(E \cap \bar{F}) = P(E) - P(E \cap F)$$
$$= P(E) - P(E) \cdot P(F) = P(E)(1 - P(F))$$
$$= P(E) \cdot P(\bar{F})$$

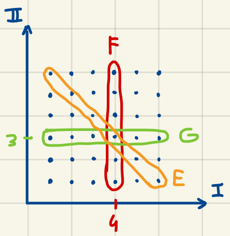


Viene naturale estendere questo concetto di indipendenza a insiemi di più di due eventi, prendendoli a coppie verificando indipendenza e dire che è un insieme di eventi indipendenti

ma

è un ragionamento fallace

Es: Lancio di due dadi



E = somma vale 7

F = 4 al primo dado

G = 3 al secondo dado

$$P(E) = P(F) = P(G) = \frac{1}{6}$$

$$P(E \cap F) = P(E \cap G) = \frac{1}{36} \longrightarrow \begin{array}{l} P(E \cap F) = P(E) \cdot P(F) \\ P(E \mid F \cap G) = 1 \end{array}$$

Non va bene quindi ragionare per coppie di eventi

Che si fa?

Andiamo a richiedere la fattorizzazione per tutti modi che io ho di calcolare l'intersezione per questi eventi

$\downarrow$

E, F, G sono indipendenti se e solo se
$$\begin{cases} P(E \cap F) = P(E) \cdot P(F) \\ P(E \cap G) = P(E)\, P(G) \\ P(F \cap G) = P(F) \cdot P(G) \\ P(E \cap F \cap G) = P(E) \cdot P(F) \cdot P(G) \end{cases}$$

Voglio dimostrare che se E, F e G sono indipendenti allora sono indipendenti E e $F \cup G$

$\downarrow$

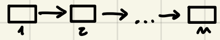
Applichiamo la definizione $\longrightarrow$
$$P(E \cap (F \cup G)) = P((E \cap F) \cup (E \cap G))$$
$$= P(E \cap F) + P(E \cap G) - P(\underbrace{(E \cap F) \cap (E \cap G)}_{E \cap F \cap G})$$
$$= P(E) \cdot P(F) + P(E) \cdot P(G) - P(E) \cdot P(F) \cdot P(G)$$
$$= P(E)(P(F) + P(G) - P(F) \cdot P(G))$$
$$= P(E)(\underbrace{P(F) + P(G) - P(F \cap G)}_{P(F \cup G)})$$
$$= P(E)\, P(F \cup G)$$

Immaginiamo di aver n eventi

$E_1, \dots, E_n$ sono indipendenti $\longrightarrow$ se e solo se $\forall r = 2, \dots, n$

$$\forall\ \underbrace{1 \leq \alpha_1 \leq \dots \leq \alpha_r}_{r \text{ indici distinti}} \leq n \qquad P\left(\bigcap_{i=1}^{r} E_{\alpha_i}\right) = \prod_{i=1}^{r} P(E_{\alpha_i})$$

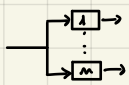
Sistemi in serie



$\forall i = 1, \dots, m \quad p_i = P(\text{l'i-esimo componente funziona})$

$$
\begin{aligned}
P(\text{il sistema funziona}) &= P(\text{Tutti i componenti funzionano}) \\
&= P\Big(\bigcap_{i=1}^{m} \text{componente } i\text{-esimo funziona}\Big) \\
&\quad \downarrow \text{ l'ipotesi di indipendenza mi permette di riscrivere} \\
&= \prod_{i=1}^{m} p_i
\end{aligned}
$$

Sistema in parallelo



$$
\begin{aligned}
P(\text{il sistema funziona}) &= 1 - P(\text{sistema non funziona}) \\
&= 1 - P(\text{Tutti i componenti non funzionano}) \\
&= 1 - P\Big(\bigcap_{i=1}^{m} \text{componente } i\text{-esimo non funziona}\Big) \\
&= 1 - \prod_{i=1}^{m} P(\text{comp. } i\text{-esimo non funziona}) \\
&= 1 - \prod_{i=1}^{m} (1 - p_i)
\end{aligned}
$$

Esempio: collezione figurine, album da m figurine
ogni volta prendo una bustina che ha 1 figurina

$p_i = P(\text{acquisto } i\text{-esima}) \qquad 0 \le p_i \le 1 \qquad \sum_{i=1}^{m} p_i = 1$

Ho comprato k figurine $\qquad P(\underbrace{\text{almeno un esemplare di } j}_{A_j} \mid \underbrace{\text{almeno un esemplare di } i}_{A_i})$

Se mi viene chiesto di calcolare la probabilità di un evento, e in questo evento occorre il termine almeno, quasi sempre vale la pena passare a evento complementare

$$P(A_j \mid A_i) = \frac{P(A_i \cap A_j)}{P(A_i)} = \frac{1 - (1-p_i)^k - (1-p_j)^k + (1-p_i-p_j)^k}{1 - (1-p_i)^k}$$

$$
\begin{aligned}
P(A_i) &= 1 - P(\bar{A}_i) = 1 - P(\text{su } k \text{ acquisti, mai la } i) \\
&= 1 - P\Big(\bigcap_{r=1}^{k} \text{al } r\text{-esimo acquisto non trovo la } i\Big) \\
&\quad \downarrow \text{ stiamo ipotizzando eventi indipendenti} \\
&= 1 - \prod_{r=1}^{k} P(\text{al } r\text{-esimo acquisto non trovo la } i) \\
&= 1 - \prod_{r=1}^{k} (1 - p_i) = 1 - (1 - p_i)^k
\end{aligned}
$$

$$
\begin{aligned}
P(A_i \cap A_j) &= 1 - P(\overline{A_i \cap A_j}) = 1 - P(\bar{A}_i \cup \bar{A}_j) \\
&= 1 - \big(P(\bar{A}_i) + P(\bar{A}_j) - P(\bar{A}_i \cap \bar{A}_j)\big) \\
&= 1 - \big((1-p_i)^k + (1-p_j)^k - \underline{P(\bar{A}_i \cap \bar{A}_j)}\big) \\
&= 1 - (1-p_i)^k - (1-p_j)^k + (1-p_i-p_j)^k
\end{aligned}
$$

$$
\begin{aligned}
P(\bar{A}_i \cap \bar{A}_j) &= P(\text{su } k \text{ acquisti, mai né la } i \text{, né la } j) \\
&= P\Big(\bigcap_{r=1}^{k} \text{al } r\text{-esimo acquisto non trovo né la } i \text{, né la } j\Big) \\
&= \prod_{r=1}^{k} (1 - p_i - p_j) = (1 - p_i - p_j)^k
\end{aligned}
$$

Variabile aleatoria: l'idea è quella di avere una quantità variabile in modo aleatorio, e varia su un insieme numerico

Si introducono per codificare degli esiti di un esperimento aleatorio in termini numerici

Dato uno spazio delle probabilità $(\Omega, \mathcal{A}, P)$

spazio eventi — algebra — funz. probabilità

se lo posso definire una funzione $X: \Omega \longrightarrow \mathbb{R}$ (le immagini di X sono in qualche modo una codifica numerica dei possibili esiti di un esperimento aleatorio)

si dice che sto definendo una variabile aleatoria X

la casualità del nostro esperimento induce un'incertezze sul valore di X

Es: $P(X=\alpha) \equiv \{\omega \in \Omega : X(\omega) = \alpha\}$

Perché si fa tutto ciò?

Esempio: Ω: lancio di due dadi non truccati $\quad X :=$ somma dei due esiti

$$P(X=3) \equiv \{\omega \in \Omega \;\; X(\omega)=3\} = \{(1,2),(2,1)\} = 2/36$$

Possibili valori di X: da 2 a 12

Specificazione: indica uno dei valori che la mia variabile aleatoria può assumere

$$P(X=2) = P(\{1,1\}) = 1/36$$
$$P(X=3) = P(\{1,2\},\{2,1\}) = 2/36$$
$$P(X=4) = P(\{1,3\},\{2,2\},\{3,1\}) = 3/36$$
$$P(X=5) = P(\{1,4\},\{2,3\},\{3,2\},\{4,1\}) = 4/36$$
$$P(X=6) = P(\{1,5\},\{2,4\},\{3,3\},\{4,2\},\{5,1\}) = 5/36$$
$$P(X=7) = P(\{1,6\},\{2,5\},\{3,4\},\{4,3\},\{5,2\},\{6,1\}) = 6/36$$
$$P(X=8) = P(\{2,6\},\{3,5\},\{4,4\},\{5,3\},\{6,2\}) = 5/36$$
$$P(X=9) = P(\{3,6\},\{4,5\},\{5,4\},\{6,3\}) = 4/36$$
$$P(X=10) = P(\{4,6\},\{5,5\},\{6,4\}) = 3/36$$
$$P(X=11) = P(\{5,6\},\{6,5\}) = 2/36$$
$$P(X=12) = P(\{6,6\}) = 1/36$$

Si verifica facilmente che $\sum_{i=2}^{12} P(X=i) = P\left(\bigcup_{i=2}^{12} \{X=i\}\right) = P(\Omega) = 1$

A partire da uno stesso spazio di probabilità, quindi da uno stesso esperimento casuale, mica devo definire per forza un'unica variabile aleatoria, posso definirne tante

Es: I = esito del primo lancio $\quad P(I=i) = \frac{1}{6} \quad \forall i = 1, \dots, 6$

Funzione indicatrice / caratteristica: $A \subseteq \mathbb{R}$

$I_A : \mathbb{R} \longrightarrow \{0,1\}$ $\qquad \boxed{I_A(x) = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{altrimenti} \end{cases}}$

Lettera maiuscola = var. aleatoria
Stessa lettera MINUSCOLA = specificazione della var. aleatoria

Es:

$P(I=i) = \frac{1}{6} I_{\{1,\dots,6\}}(i)$

insieme di specificazioni (supporto della variab. aleatoria)

Esempio: Acquisto due componenti elettronici $\qquad$ d: difettoso $\quad$ f: funzionante

$P(d) = 0.3 \qquad P(f) = 0.7$

Nell'ipotesi che ci sia indipendenza tra i due componenti $\longrightarrow$ posso definire $\Omega = \{(d,d),(d,f),(f,d),(f,f)\}$

$X = \#$ componenti funzionanti

$P(X=0) = 0.3 \times 0.3 = 0.09$
$P(X=1) = (0.7 \times 0.3) \cdot 2 = 0.42$
$P(X=2) = 0.7 \times 0.7 = 0.49$

Definiamo $I = \begin{cases} 1 & \text{se almeno un comp. funziona} \\ 0 & \text{altrimenti} \end{cases}$

$P(I=1) = 0.42 + 0.49 = 0.91$
$P(I=0) = 0.09$

FUNZIONE INDICATRICE
DI UN EVENTO (almeno un componente funzionante)

Funzione di ripartizione / distribuzione cumulativa

Ogni variabile aleatoria ha la sua funzione di ripartizione

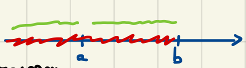
$F_X : \mathbb{R} \longrightarrow [0,1]$ com'è definita? Devo prendere un generico elemento del dominio e dire qual è il valore assunto

$F_X : \mathbb{R} \longrightarrow [0,1] \quad \forall x \in \mathbb{R} \quad F_X(x) = P(X \leq x)$

Una volta che io conosco la funzione di ripartizione so tutto quello che c'è da sapere su una certa variabile aleatoria, la conoscenza di F individua univocamente la variabile aleatoria

$X \sim F_X$

$\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\}$



UNIONE DISGIUNTA 3 ASSIOMA KOLMOGOROV

$P(X \leq b) = P(X \leq a) + P(a < X \leq b)$
$F_X(b) = F_X(a) + P(a < X \leq b)$
$P(a < X \leq b) = F_X(b) - F_X(a)$ $\longrightarrow$ Tutte le domande che posso farmi relative a eventi composti usando la variabile aleatoria X, queste probabilità posso calcolarle in termini della funzione di ripartizione

Consideriamo il caso semplice: X è una var. aleatoria il cui supporto (insieme delle sue specificaz.) è numerabile

ma

dire che è numerabile vuol dire che l'insieme è finito o è un insieme con stessa cardinalità dei numeri naturali

posso ragionare in termini di qualsiasi range di valori in tale insieme

Noi ci concentreremo per un po' sulle variabili aleatorie discrete

hanno supporto discreto

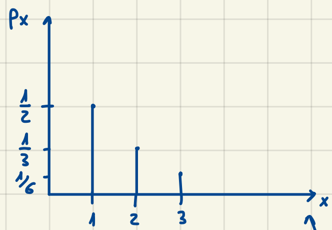
Possiamo affiancare un'altra funzione $\rightarrow$ **Funzione di massa di probabilità**: $p_X : \mathbb{R} \longrightarrow [0,1] \quad \forall x \in \mathbb{R} \quad p_X(x) = P(X = x)$

Un paio di proprietà:
- $p_X(x) \geq 0 \quad \forall x \in \mathbb{R}$

  $D_X = \{x_1, x_2, \dots\}$ (posso elencare il supporto della variabile aleatoria)
- $\sum_i p_X(x_i) = 1$

Se prendo una qualsiasi funzione da $\mathbb{R}$ in $[0,1]$ che soddisfa queste proprietà, allora tale funzione deve essere necessariamente la funzione di massa di probabilità di una qualche variabile aleatoria

Es: v.a X discreta $\quad D_X = \{1, 2, 3\} \quad P(X=1) = \frac{1}{2} \quad P(X=2) = \frac{1}{3}$



$p_X(1) + p_X(2) + p_X(3) = 1$

$p_X(3) = 1 - \frac{1}{2} - \frac{1}{3} = \frac{1}{6}$

Se volessi sapere quanto vale

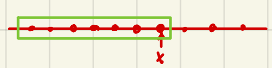
$P(X=6) = 0 \longrightarrow$ non è nel supporto

in genere la funz. di massa di probabilità si annulla in tutti i punti di $\mathbb{R}$ che non sono anche punti del supporto

Che relazione c'è tra funzione di massa di probabilità e funz. di ripartizione?

Se sto calcolando $F_X(x) = P(X \leq x) = P(\bigcup_{a \leq x} \{X = a\})$

UNIONE DI EVENTI DISGIUNTI

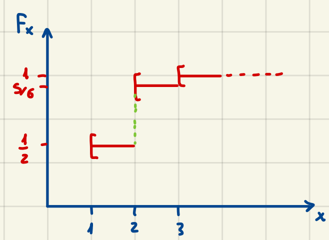


$= \sum_{a \leq x} P(X = a) = \sum_{a \leq x} p_X(a)$

Ho il grafico della funz. di massa di prob.

Se invece devo trovare la funzione di ripartizione

$$F_X(x) = \begin{cases} 0 & \text{se } x < 1 \\ 1/2 & \text{se } 1 \leq x < 2 \\ 5/6 & \text{se } 2 \leq x < 3 \\ 1 & \text{se } x \geq 3 \end{cases}$$



$F_X(2) - F_X(1)$

$\sum_{x \leq 2} p_X(x) - \sum_{x \leq 1} p_X(x) = \cancel{p_X(1)} + p_X(2) - \cancel{p_X(1)}$

Proprietà funz. ripartizione:
- $\forall x \in \mathbb{R} \quad F_X(x) \geq 0$
- $\lim_{x \to +\infty} F_X(x) = 1$
- F_X è continua da destra (dovuto al $\leq$ nella def. di funz di ripartizione)

Se prendo una funzione che soddisfa queste 3 proprietà, allora deve esistere una var. aleatoria di cui questa funzione è una funzione di ripartizione

Valore atteso di variabile aleatoria X : indice di centralità / posizione

Legato alla media campionaria $\{x_1, x_2, \dots\}$

$E(X) = \sum_i x_i \, p(X = x_i)$ VALORE ATTESO

posso mettere funz di massa di prob. volendo $p_X(x_i)$

questa somma converge? Supporto finito sì

ma

se avessi una specificazione in corrisp. di ogni numero naturale? Diventa una serie vera e propria

converge o no?

LANCIO MONETA BILANCIATA

Es: $D_X = \{0, 1\}$ $\quad p_X(0) = p_X(1) = \frac{1}{2}$

$E(X) = 0 \cdot p_X(0) + 1 \cdot p_X(1) = p_X(1) = \frac{1}{2}$

LANCIO MONETA NON BILANCIATA

$p_X(0) = \frac{2}{3} \quad p_X(1) = \frac{1}{3}$

$E(X) = 0 \cdot p_X(0) + 1 \cdot p_X(1) = \frac{1}{3}$

Es 2:

X = punteggio di un dado bilanciato $\quad D_X = \{1, 2, 3, 4, 5, 6\} \quad p_X(x) = \frac{1}{6} I_{\{1, \dots, 6\}}(x)$

$E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} = 3.5$

$= \sum_{x=1}^{6} x \, p_X(x) = \sum_{x=1}^{6} x \cdot \frac{1}{6} = \frac{1}{6} \sum_{x=1}^{6} x = \frac{1}{\cancel{6}} \cdot \frac{\cancel{6} \cdot 7}{2} = \frac{7}{2} = 3.5$ — NON È UN VALORE ASSUMIBILE DALLA MIA VAR ALEATORIA

ottengo ciò perché il dado è bilanciato ⟶ le prob. spingono verso di loro il valore medio

X = vinco in un gioco d'azzardo $\quad$ gioco n volte

x_i = quanto vinco

n_1 volte vinco x_1

n_2 volte vinco x_2

⋮

n_k volte vinco x_k

vincita media $= \dfrac{n_1 x_1 + \dots + n_k x_k}{n}$

$= \sum_{i=1}^{k} \frac{n_i}{n} x_i \approx \sum_{i=1}^{k} P(X = x_i) \cdot x_i$

A evento $\quad I_A = \begin{cases} 0 & \text{se A non si verifica} \\ 1 & \text{se A si verifica} \end{cases}$

$D_{I_A} = \{0, 1\} \qquad E(I_A) = P(A)$

Valore atteso gode di una proprietà: . come media campionaria è espresso in stessa unità di misura delle specificazioni

• $g: \mathbb{R} \longrightarrow \mathbb{R} \qquad g(X) = Y$

Es: $g(x) = x^2$

x	P(X=x)	y
0	0.2	0
1	0.5	1
2	0.3	4

P(Y=y) ⟶ dato dall'iniettività

$E(Y) = 0 \cdot 0.2 + 1 \cdot 0.5 + 4 \cdot 0.3 = 1.7$

$E(x^2) = 0^2 \cdot 0.2 + 1^2 \cdot 0.5 + 2^2 \cdot 0.3 = 1.7$

g $\quad$ X

$\{x_1, \dots, x_n\}$

$E(g(X)) = \sum_i g(x_i) \cdot P(X = x_i)$

$= \sum_i g(x_i) \, p_X(x)$

$g(x) = ax+b \qquad a,b \in \mathbb{R} \qquad$ v.a $X \qquad Y = aX+b$

$E(Y) = E(g(X)) = E(aX+b) = \sum_i \underbrace{(ax_i+b)}_{g(x_i)} P(X=x_i) = \sum_i (a x_i P(X=x_i) + b P(X=x_i))$

$= a \underbrace{\boxed{\sum_i x_i P(X=x_i)}}_{=E(x)} + b \underbrace{\boxed{\sum_i P(X=x_i)}}_{=1} = aE(x)+b$ Valore atteso gode delle proprietà di linearità

$a=0 \longrightarrow E(b)=b$

$b=0 \longrightarrow E(ax) = aE(x)$

Consideriamo 3 var. aleatorie

$W=0$ con prob 1 $\qquad Y = \begin{cases} -1 & \text{prob } 1/2 \\ 1 & \text{prob } 1/2 \end{cases} \qquad Z = \begin{cases} -100 & \text{prob } 1/2 \\ 100 & \text{prob } 1/2 \end{cases}$

$E(W) = 0 \qquad E(Y) = \emptyset \qquad E(Z) = \emptyset$

$\boxed{\text{Concetto di dispersione su v.a}}$

$\downarrow$

Quanto si discosta il valore atteso? Possiamo calcolare lo scarto? $g(x) = |x-E(x)| \qquad E(|x-E(x)|)$

ma avevamo detto nella statistica descrittiva che modulo porta problemi

$E((x-E(x))^2)$ VARIANZA della v.a X

$\boxed{\mu := E(x)}$

$var(x) = E((x-\mu)^2)$

APPLICATA LINEARITÀ $\longrightarrow$ $= E(x^2 - 2\mu x + \mu^2)$

$= E(x^2) - 2\mu E(x) + \mu^2$

$= E(x^2) - 2\mu^2 + \mu^2$

$= E(x^2) - \mu^2$

$\boxed{= E(x^2) - E(x)^2}$

Esempio: X, $D_X = \{1, \dots, 6\} \qquad p_X(x) = \frac{1}{6} I_{D_X}(x) \qquad E(x) = 3.5$

$var = E((x-\frac{7}{2})^2) = \sum_{i=1}^{6} (x-\frac{7}{2})^2 \cdot \frac{1}{6} = \dots$

$var = E(x^2) - E(x)^2 = E(x^2) - (\frac{7}{2})^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12} = \frac{36-1}{12}$

$E(x^2) = \sum_{x=1}^{6} x^2 \frac{1}{6} = \frac{1}{6} \sum_{x=1}^{6} x^2 = \frac{1}{\cancel{6}} \frac{\cancel{6} \cdot 7 \cdot 13}{6} = \frac{91}{6}$

$I_A = \begin{cases} 1 & \text{se A si verifica} \\ 0 & \text{altrimenti} \end{cases} \qquad E(I_A) = P(A) \qquad I_A$ gode di proprietà di idempotenza

$Var(I_A) = E(I_A^2) - E(I_A)^2 = E(I_A) - E(I_A)^2 = E(I_A)(1-E(I_A)) = P(A) \cdot P(\bar{A})$

Cosa succede se provo a trasformare linearmente la varianza?

$x \to ax+b$ $\qquad E(aX+b) = aE(X)+b$

$\text{Var}(\underbrace{aX+b}_{:=Y}) = E\big((\underbrace{aX+b}_{Y} - E(\underbrace{aX+b}_{Y}))^2\big) \leftarrow E((Y-E(Y))^2)$

$= E((aX+\cancel{b} - aE(X) - \cancel{b})^2)$

$= E(a^2(X-E(X))^2)$

$= a^2 E((X-E(X))^2) = a^2\,\text{Var}(X)$

b scomparso, è normale pure nella varianza in statistica descrittiva era così, equivale ad una traslazione geometrica (ininfluente)

Se impongo $a=0$ $\quad \text{Var}(b) = 0$

↓

variabili aleatorie degeneri hanno varianza nulla

Varianza nella variab. aleatoria, da lo stesso problema che dava in statistica descrittiva, avevamo perciò introdotto deviazione standard

↓

DEVIAZIONE STANDARD DI UN V.A X: $\sigma_X = \sqrt{\text{Var}(X)}$

In alcuni contesti conviene ragionare in termini di coppie di variabili aleatorie

Es: coppia X, Y

Dobbiamo estendere concetti di funz. di ripartizione e funz. di massa di probabilità

↓

Funzione di ripartizione congiunta: $F_{X,Y}(x,y) = P(X \leq x \overset{\text{intersez}}{,} Y \leq y)$

DISTRIBUZIONE MARGINALE

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

analogamente mandando $x \to +\infty$ avremmo ottenuto la ripartizione di Y con y

Funzione di massa di probabilità congiunta: $p_{X,Y}(x,y) = P(X=x, Y=y)$

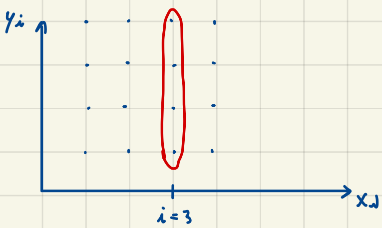
se anche uno solo tra x e y è un valore che non rispecchia una specificazione della corrispondente variabile aleatoria la funz. di massa di prob. congiunta è $\emptyset$

$D_X = \{x_1, \dots\}$ $\quad$ consideriamo $X = x_i$

$D_Y = \{y_1, \dots\}$ $\quad\quad\quad\quad\quad\quad\;\; Y = y_j$

$\boxed{\text{Se faccio } \bigcup_j \{X=x_i, Y=y_j\} = \{X=x_i\}}$

Cosa significa?



$\sum_j p_{X,Y}(x_i,y_j) = p_X(x_i)$ MARGINALE

Analogo fissando una specificaz. di Y

Indipendenza esteso a due variabili aleatorie

X e Y sono indipendenti se e solo se $\forall A, B \in \mathbb{R}$

$X \in A \quad Y \in B$

posso calcolare intersez. tra questi due eventi
e calcolarne la probabilità

$$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)$$

Equivale a:
- $\forall a, b \in \mathbb{R} \quad F_{X,Y}(a,b) = F_X(a) F_Y(b)$
- $\forall a, b \in \mathbb{R} \quad p_{X,Y}(a,b) = p_X(a) \cdot p_Y(b)$

Dimostriamo: $P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B) \implies p_{X,Y}(a,b) = p_X(a) \cdot p_Y(b)$

$A = \{a\} \quad B = \{b\}$ basta prendere i singoletti

$p_{X,Y}(a,b) = p_X(a) \cdot p_Y(b) \implies \forall A, B \quad P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)$

$$P(X \in A, Y \in B) = \sum_{a \in A} \sum_{b \in B} p_{X,Y}(a,b) = \sum_{a \in A} \sum_{b \in A} p_X(a) \cdot p_Y(b) = \underbrace{\boxed{\sum_{a \in A} p_X(a)}}_{P(X \in A)} \cdot \underbrace{\boxed{\sum_{b \in B} p_Y(b)}}_{P(Y \in B)}$$

Generalizziamo a più di due var. aleatorie

$X_1, X_2, \dots, X_n$

$$F_{X_1,\dots,X_n}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$$

$$p_{X_1,\dots,X_n}(x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n)$$

Indipendenza $\forall A_1, \dots, A_n \in \mathbb{R}$

$\downarrow$

$$P\left(\bigcap_{i=1}^{n} X_i \in A_i\right) = \prod_{i=1}^{n} P(X_i \in A_i)$$

Solo se vale indipendenza

$\downarrow$

$$F_{X_1,\dots,X_n}(x_1, \dots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$$

Variabile aleatoria multivariata: vettore aleatorio

altra opzione

Introdurre tante variab. aleatorie diverse quanti sono i valori (n var. aleatorie)

Cosa succede al valore atteso quando abbiamo 2 o più variabili aleatorie?

Es: calcolo valore atteso di una funz. di più variabili aleatorie

v.a X + funzione $g: \mathbb{R} \to \mathbb{R}$ $\quad E(X) = \sum_i x_i P(X = x_i) \quad E(g(X)) = \sum g(x_i) P(X = x_i)$

Supponiamo ora di avere:

v.a X, Y + funzione $f: \mathbb{R}^2 \to \mathbb{R}$ $\quad E(f(X,Y)) = \sum_{i,j} f(x_i, y_j) P(X = x_i, Y = y_j)$

Es: $f(x,y) = x + y$ $\quad E(X+Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i + y_j) P(X = x_i, Y = y_j)$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} x_i P(X = x_i, Y = y_j) + \sum_{i=1}^{n} \sum_{j=1}^{m} y_j P(X = x_i, Y = y_j)$$

$$= \sum_{i=1}^{n} x_i \boxed{\sum_{j=1}^{m} P(X = x_i, Y = y_j)} + \sum_{j=1}^{m} y_j \boxed{\sum_{i=1}^{n} P(X = x_i, Y = y_j)}$$

$$= \underbrace{\sum_{i=1}^{n} x_i P(X = x_i)}_{= E(X)} + \underbrace{\sum_{j=1}^{m} y_j P(Y = y_j)}_{E(Y)}$$

$E(X+Y) = E(X) + E(Y)$

$\downarrow$

la linearità del valore atteso si spinge oltre

ma

Ci sono delle ipotesi che non abbiamo detto su X e Y per rendere ciò possibile:

1) X, Y var. aleatorie discrete

2) Supporto di X e Y sono finiti

Questa cosa possiamo estenderla al valore atteso della somma di 3, 4, .. var aleatorie

$$E(X+Y+Z) = E((X+Y)+Z) = E(X+Y) + E(Z) = E(X) + E(Y) + E(Z)$$

Valore atteso della somma di n var. aleatorie è somma degli n valori attesi

Es: valore atteso della $\boxed{\text{somma dei risultati di due dadi non truccati}} := X$

$$E(X) = \sum_{i=2}^{12} i \, P(X = i) = \ldots$$

$\boxed{X = X_1 + X_2} \longrightarrow E(X) = E(X_1 + X_2) = E(X_1) + E(X_2) = \frac{7}{2} + \frac{7}{2} = 7$

$X_1, X_2 :=$ esito di un dado bilanciato

Es: ci sono n lettere ed n buste, ogni lettera associata ad una busta
se cadono e alcune lettere vengono mischiate, prob. che lettere sono in buste corrette?

n lettere + n buste (accoppiate a caso) → equiprobabilità
Quanti abbinamenti giusti?

$\forall i = 1, \dots, n \quad X_i = \begin{cases} 1 & \text{se } i\text{-esima lettera ok} \\ 0 & \text{altrimenti} \end{cases}$

var. aleatoria associata alla funz. indicatrice di un certo evento

perché? Facendo $\sum_{i=1}^{n} X_i$ se tutte giuste ottengo n

$\Rightarrow E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \frac{1}{n} = 1$

$E(X_i) = ? \longrightarrow E(X_i) = P(X_i = 1) = \frac{1}{n}$

Es 2: compro 10 confezioni
20 buoni diversi (1 per confezione)

valore atteso di $X = \#$buoni diversi

$\forall i = 1, \dots, 20 \quad X_i = \begin{cases} 1 & \text{se ho almeno un buono } i\text{-esimo} \\ 0 & \text{altrimenti} \end{cases}$

$$E\left(\sum_{i=1}^{20} X_i\right) = \sum_{i=1}^{20} E(X_i)$$

$E(X_i) = P(X_i = 1) = P($almeno un esemplare del buono i-esimo$)$
ragioniamo in termini di evento complementare

$= 1 - P($in nessuna delle 10 confez. appare il buono i-esimo$)$

$= 1 - P\left(\bigcap_{j=1}^{10} \text{nella confezione } j \text{ non compare il buono } i\right)$

APPLICO IPOTESI INDIPENDENZA

$= 1 - \prod_{j=1}^{10} P(\text{nella confezione } j \text{ non compare il buono } i)$

APPLICO COMPLEMENTARE

$= 1 - \prod_{j=1}^{10} \left(1 - P(\text{nella confezione } j \text{ compare il buono } i)\right)$

$= 1 - \prod_{j=1}^{10} \left(1 - \frac{1}{20}\right) = 1 - \prod_{j=1}^{10} \frac{19}{20} = 1 - \left(\frac{19}{20}\right)^{10} = E(X_i)$

$$E\left(\sum_{i=1}^{20} X_i\right) = \sum_{i=1}^{20} E(X_i) = 20\left(1 - \left(\frac{19}{20}\right)^{10}\right) \approx 8.025$$

Nota sul valore atteso: immaginiamo data una v.a X, di voler calcolare il valore atteso di una certa quantità

$E((X-c)^2) = E(x^2 \cancel{-2c} + c^2)$ riscriviamo diversamente $\qquad \mu = E(X)$

$E((\underbrace{X-\mu}+\underbrace{\mu-c})^2) = E\left((x-\mu)^2 + 2(x-\mu)(\mu-c) + (\mu-c)^2\right) = \underbrace{E(x-\mu)^2}_{\text{VARIANZA } X} + \cancel{2(\mu-c)\underbrace{E(X-\mu)}_{E(X)-\mu = 0}} + (\mu-c)^2$

$E((X-c)^2) = Var(X) + \underbrace{(\mu-c)^2}_{\text{mai negativo}} \geq Var(X)$

così valuto la bontà di tale approssimazione → se io do a c il ruolo di ridurre l'aleatorietà

approssimazione → sto calcolando valore atteso dell'errore che sto facendo

se $c = \mu$ sarà pari alla $Var(X)$ → minimizzo errore che sto facendo

$E(X+X) = E(X) + E(X) = 2E(X)$

$E(\overset{\shortparallel}{2X}) = 2E(X)$

Se ragiono in Termini della $Var(X+X)$? $\quad Var(X+X) = Var(2X) = 4\,Var(X)$ — va elevato al quadrato

$\cancel{Var(X)+Var(X)}$

Introduciamo un indice, COVARIANZA Tra X e Y $\quad$ v.a X e Y $\quad \mu_x = E(X) \quad \mu_y = E(Y)$

$$cov(X,Y) = E((X-\mu_x)(Y-\mu_y))$$

vedere Tendenzialità di due var. aleatorie

- indice simmetrico
- $cov(X,Y) = E(XY - \mu_x Y - \mu_y X + \mu_x\mu_y)$

$$= E(XY) - \underbrace{\mu_x E(Y)}_{\mu_x\mu_y} - \underbrace{\cancel{\mu_y E(X)}}_{\mu_y\mu_x} + \cancel{\mu_x\mu_y} = E(XY) - E(X)E(Y)$$

- $cov(aX,Y) = E(aXY) - E(aX)E(Y) = aE(XY) - aE(X)E(Y) = a(\overbrace{E(XY) - E(X)E(Y)}^{cov(X,Y)})$

$\downarrow$

$cov(aX,Y) = a\,cov(X,Y)$

- $cov(X+Y,Z) = E((X+Y)\cdot Z) - E(X+Y)\cdot E(Z)$

$$= E(XZ+YZ) - (E(X)+E(Y))E(Z)$$
$$= \boxed{E(XZ)} + \boxed{E(YZ)} - \boxed{E(X)E(Z)} - \boxed{E(Y)E(Z)}$$
$$= cov(X,Z) + cov(Y,Z)$$

Anche questo si può estendere a più variabili

$$cov(\Sigma_i X_i, \Sigma_j Y_j) = \Sigma_i\, cov(X_i, \Sigma_j Y_j)$$
$$= \Sigma_i\, cov(\Sigma_j Y_j, X_i)$$
$$= \Sigma_i \Sigma_j (Y_j, X_i) = \Sigma_i \Sigma_j\, cov(X_i, Y_j)$$

- $cov(X,X) = E((X-\mu_x)(X-\mu_x)) = E((X-\mu_x)^2) = Var(X)$

$$= E(X\cdot X) - E(X)\cdot E(X) = E(X^2) - E(X)^2 = Var(X)$$

$Var(X+Y) = E((X+Y)^2) - E(X+Y)^2$

$\downarrow$

Per capire cosa succede ci Torna utile la covarianza Tra due v.a

$$Var(X+Y) = E((X+Y)^2) - E(X+Y)^2$$
$$= E(X^2+2XY+Y^2) - (E(X)+E(Y))^2$$
$$= E(X^2) + 2E(XY) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2$$
$$= E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2 + 2(E(XY) - E(X)E(Y))$$
$$= Var(X) + Var(Y) + 2\,cov(X,Y)$$

ma se X e Y fossero uguali

$$Var(X,X) = Var(X) + Var(X) + 2\,\overset{Var(X)}{\overset{\shortparallel}{cov(X,X)}} = Var(X) + Var(X) + 2Var(X) = 4\,Var(X)$$

Come estendo la Var della somma di variabili aleatorie quando ce ne sono più di 2?

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i\neq j} cov(X_i, X_j)$$

Che succede se le var. aleatorie che sto sommando sono indipendenti?

Teorema: Se due v.a sono indipendenti $\Longrightarrow$ $E(XY) = E(X)\cdot E(Y)$

Dim: $E(XY) = \sum_i \sum_j x_i \cdot y_j \cdot \underline{P(X=x_i, Y=y_j)}$

per ip. di indipendenza

$= \sum_i \sum_j x_i \cdot y_j \cdot P(X=x_i)\cdot P(Y=y_j)$

doppia sommatoria separabile

$= \underline{\sum_i x_i P(X=x_i)}\ \underline{\sum_j y_j P(Y=y_j)}$

suppongo che le sommatorie convergono

$= E(X)\cdot E(Y)$

Corollari:

- Se X, Y indipendenti $\Longrightarrow cov(X,Y) = E(XY) - E(X)E(Y) = E(X)\cancel{E(Y)} - E(X)\cancel{E(Y)} = \emptyset$
- $X_1, \dots, X_n$ indipendenti $\Longrightarrow Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i\neq j} \cancel{cov(X_i, X_j)}$

nullo per corollario precedente

Es: Var della somma di 10 lanci indipendenti di un dado bilanciato

X_i = esito di un dado bilanciato $\quad i = 1, \dots, 10$

$$Var\left(\sum_{i=1}^{10} X_i\right) = \sum_{i=1}^{10} Var(X_i) = \sum_{i=1}^{10} \frac{35}{12} = 10\cdot\frac{35}{12} = \frac{175}{6}$$

Es 2: # Teste che si verificano lanciando 10 volte moneta bilanciata

A = "esce Testa" $\quad I_A = \begin{cases} 1 & \text{prob. } 1/2 \\ 0 & \text{prob } 1/2 \end{cases}$ $\quad E(I_A) = P(A) = \frac{1}{2}$ $\quad Var(I_A) = P(A)\cdot P(\bar{A}) = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$

$$Var\left(\sum_{i=1}^{10} I_A\right) = \sum_{i=1}^{10} Var(I_A) = 10\cdot\frac{1}{4} = \frac{5}{2}$$

Vediamo se la tendenzialità della covarianza si mostra pure con var. aleatorie

Es: $X = I_A \quad Y = I_B \quad A, B \in \mathcal{A} \quad E(X) = ?$

(Se cov=0 indipendenza)

$E(X) = P(X=1) \quad E(Y) = P(Y=1) \quad XY = \begin{cases} 0 & \text{altrimenti} \\ 1 & \text{se A e B si verificano} \end{cases} \quad E(XY) = P(XY=1)$

$cov(X,Y) = E(XY) - E(X)E(Y)$
$= P(XY=1) - P(X=1)P(Y=1)$
$= P(X=1, Y=1) - P(X=1)P(Y=1)$

ci fosse stato il < se Y assume certo valore c'é meno prob che X lo assuma

Sapere che Y ha assunto una certa specificazione, mi dice che diventa più probabile che X abbia assunto quel valore → c'é relazione diretta tra specificazioni che possono assumere le var.

Ipotizziamo di sapere il segno della $cov(X,Y) > 0 \longrightarrow P(X=1, Y=1) > P(X=1)P(Y=1) \longrightarrow \frac{P(X=1, Y=1)}{P(Y=1)} > P(X=1)$

$P(X=1|Y=1) = \frac{P(X=1, Y=1)}{P(Y=1)}$

interses

é la prob condizionata

C'è un problema, cosa succede se calcolo $cov(2X, 2Y)$?

Se X e Y sono in relazione, il fatto di raddoppiarli non cambia la relazione

la forza che c'è in questa relazione deve essere la stessa

non può essere misurata da cov, costanti posso portarle fuori $cov(2X, 2Y) = 4\,cov(X,Y)$

Stessa problematica nella statistica descrittiva, avevamo introdotto coeff. di correlazione

$\rho = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$ COEFF. DI CORRELAZIONE

$Var(2X) = 4 Var(X)$

$\sigma_{2X} = \sqrt{Var(2X)} = 2\sqrt{Var(X)} = 2\sigma_X$

$\rho_{2X,2Y} = \frac{cov(2X,2Y)}{\sigma_{2X}\sigma_{2Y}} = \frac{\cancel{4}\,cov(X,Y)}{\cancel{4}\,\sigma_X \sigma_Y}$

indipendente dalle scalature, cattura la forza d. relazione che c'è tra le due variabili aleatorie

$-1 \leq \rho \leq 1$

Variabili aleatorie continue (hanno supporto continuo)

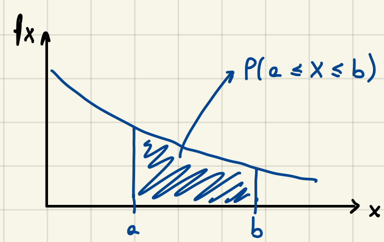
. Funzione di ripartizione è universale, la posso usare per descrivere variabili aleatorie continue

. Funz. di massa di prob. $\longrightarrow$ vale solo per var. aleatorie discrete

sostituita dalla Funzione di densità di probabilità

$f_X : \mathbb{R} \longrightarrow \mathbb{R}^+$ quando voglio calcolare la prob. $X \in B$

$\forall B \subseteq \mathbb{R}$, $P(X \in B) = \int_B f_X(x)\,dx$



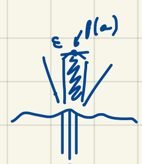
$P(a \leq X \leq b) = \int_a^b f_X(x)\,dx$

$\parallel$

$P(a < X < b)$

$P(X = a) = P(a \leq X \leq a) = \int_a^a f_X(x)\,dx = 0$ PROB. CHE VAR ALEATORIA CONTINUA ASSUMA VALORE FISSO È ∅

Es: $a - \frac{\varepsilon}{2} \leq X \leq a + \frac{\varepsilon}{2}$



per ε molto piccolo e condiz. di continuità di f

$P\left(a - \frac{\varepsilon}{2} \leq X \leq a + \frac{\varepsilon}{2}\right) = \int_{a-\frac{\varepsilon}{2}}^{a+\frac{\varepsilon}{2}} f_X(x)\,dx \approx \varepsilon f(a)$

Quindi cosa ci dice la funzione di densità? Mi dà un'indicazione di quanto sia verosimile che la mia variab. aleatoria assuma specificazioni in un piccolo intorno di a

Cosa succede se calcolo $\int_{-\infty}^{+\infty} f_X(x)\,dx$?

$\int_{-\infty}^{+\infty} f_X(x)\,dx = \int_{\mathbb{R}} f_X(x)\,dx = P(X \in \mathbb{R}) = 1$ (evento Ω)

proprietà che deve sempre soddisfare una funzione da $\mathbb{R}$ in $\mathbb{R}^+$ per essere una funzione di densità di probabilità

$\int_{-\infty}^{a} f_X(x)\,dx = P(X \leq a) = F_X(a)$ funz di ripartizione

ma

la ottengo integrando funz. di densità

se derivo $\frac{d}{dx} F_X(x) = f(x)$

Teor. fond. calcolo integrale

Es: v.a continua X $\quad f_X(x) = \begin{cases} c(4x - 2x^2) & \text{se } 0 < x < 2 \\ 0 & \text{altrimenti} \end{cases} = c(4x - 2x^2)\, I_{(0,2)}(x)$

1) Quanto vale c?

2) Quanto vale $P(X > 1)$

1) $\int_{-\infty}^{+\infty} f_X(x)\,dx = 1 \qquad \int_0^2 c(4x - 2x^2)\,dx = 1 \qquad c\int_0^2 4x - 2x^2\,dx = 1 \qquad c\int_0^2 4x - c\int_0^2 2x^2 = 1 \qquad c\left[2x^2\right]_0^2 - c\left[\frac{2}{3}x^3\right]_0^2 = 1$

$c\left[8 - 0\right] - c\left[\frac{2}{3}8 - 0\right] = 1 \qquad c\,8 - c\frac{16}{3} = 1 \qquad \frac{c24 - 16c}{3} = 1 \qquad \frac{8}{3}c = 1 \qquad c = \frac{3}{8}$

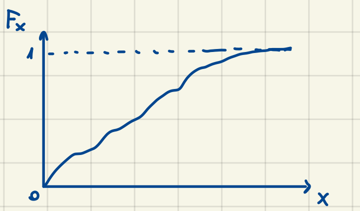
2) $P(X > 1) = \int_1^{+\infty} f_X(x)\,dx = \int_1^2 \frac{3}{8}(4x - 2x^2)\,dx = \frac{3}{8}\int_1^2 4x - 2x^2 = \frac{3}{8}\left[4\frac{x^2}{2} - \frac{2x^3}{3}\right]_1^2 = \frac{3}{8}\left[\frac{4\cdot 4}{2} - \frac{2\cdot 8}{3} - \frac{4}{2} + \frac{2}{3}\right] = \frac{3}{8}\left[8 - \frac{16}{3} - \frac{4}{2} + \frac{2}{3}\right] = \frac{3}{8}\left[\frac{48 - 32 - 12 + 4}{6}\right]$

$= \frac{3}{8}\cdot\frac{8}{6} = \frac{1}{2} \qquad P(X > 1) = \frac{1}{2}$

A parte queste cose diverse il resto vale tutto, indipendenza ecc

$E(x) = \int_{-\infty}^{+\infty} x\, f_X(x)\,dx$

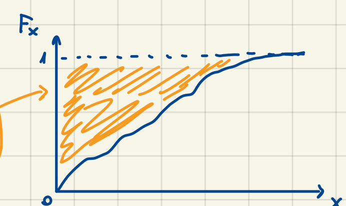
Cosa succede se provo a graficare funz. di ripartizione di una var. aleatoria



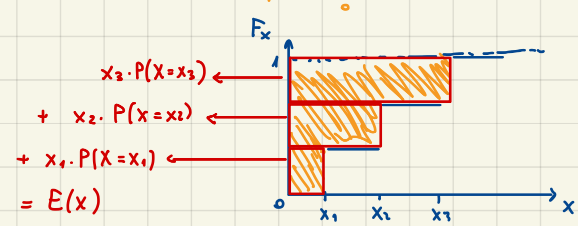
- monotona non decrescente
- continua da dx
- Tende a 1 per arg che tende a ∞

Una funz. di ripartiz. di una v.a CONTINUA È CONTINUA

Variab. aleatorie che non assumono mai specificazioni negative $\quad X \geq 0$



ci interessiamo a calcolare questo $\int_0^{+\infty} 1 - F_X(x)\,dx = E(x)$ si otterrebbe questo (anche per v.a discrete)



Disuguaglianza di Markov

Teorema: se $X \geq 0 \quad \forall a > 0 \quad P(X \geq a) \leq \frac{E(x)}{a}$

Dim: X v.a continua

$E(x) = \int_0^{+\infty} x\, f_X(x)\,dx = \underbrace{\int_0^a \overset{\geq 0}{x}\, \overset{> 0}{f_X(x)}\,dx}_{\geq 0} + \int_a^{+\infty} x\, f_X(x)\,dx \geq \int_a^{+\infty} x f_X(x)\,dx \geq \int_a^{+\infty} a\, f_X(x)\,dx = a\underbrace{\boxed{\int_a^{+\infty} f_X(x)\,dx}}_{P(X \geq a)} = a\,P(X \geq a)$

$E(x) \geq a\,P(X \geq a)$

$\frac{E(x)}{a} \geq P(X \geq a)$

Teorema: se $X \geq 0 \quad \forall a > 0 \quad P(X \geq a) \leq \frac{E(x)}{a}$

Dim: X v.a discreta

$E(x) = \sum_{x \in D_X} x f_X(x) = \underbrace{\sum_{x < a} x f_X(x)}_{\geq 0} + \sum_{x \geq a} x f_X(x) \geq \sum_{x \geq a} x f_X(x) \geq \sum_{x \geq a} a f_X(x) = a \underbrace{\sum_{x \geq a} f_X(x)}_{P(X \geq a)} = a\,P(X \geq a) \qquad E(x) = a\,P(X \geq a)$

$\boxed{P(X < a) = 1 - P(X \geq a) \geq 1 - \frac{E(x)}{a}}$

Disuguaglianza di Tchebyshev

Teorema: v.a X, $E(x) = \mu$, $Var(x) = \sigma^2$

$$\forall r > 0 \quad P(|X-\mu| \geq r) \leq \frac{\sigma^2}{r^2}$$

Qual è la prob. che la mia v.a assuma specificaz. che ha una distanza rispetto al valore atteso $\geq r$

prob. dei due eventi che si complicano è uguale

vale implicaz. inversa perché abbiamo detto $r>0$

Dim: $|X-\mu| \geq r \longleftrightarrow (X-\mu)^2 \geq r^2$

Per Markov

$$P(|X-\mu| \geq r) = P((X-\mu)^2 \geq r^2) = P(Y \geq r^2) \leq \frac{E(Y)}{r^2} = \frac{E((X-\mu)^2)}{r^2} = \frac{\overbrace{Var(X)}^{=\sigma^2}}{r^2}$$

Introduciamo una v.a $Y := (X-\mu)^2$ $\longrightarrow$ valgono ipotesi disug. di Markov per lei

Ci permettono di fare valutaz. crude ma veloci questi due Teo.

Es: $X = \#$pezzi prodotti in 7gg

$E(X) = 50 = \mu$

$P(X \geq 75) \leq \frac{E(x)}{75} = \frac{50}{75} = \frac{2}{3}$ è informativo ciò? Utile per dire se le cose vanno male

Ma se sapessi non solo che $E(x) = 50$ ma anche che $Var(X) = 25$ $\qquad P(|X-\mu| \geq 10) \leq \frac{25}{100} = \frac{1}{4}$

$$P(40 < X < 60) = P(-10 < X-\mu < 10) = P(|X-\mu| < 10) = 1 - P(|X-\mu| \geq 10) \geq 1 - \frac{1}{4} = \frac{3}{4}$$

ho sottratto membro a membro valore atteso

Se prob. alta allora sì in 7gg posso fare tra 40 e 60 pezzi

$$\boxed{P(|X-\mu| \geq r) \leq \frac{\sigma^2}{r^2}} \qquad r = k\sigma \qquad \frac{\cancel{\sigma^2}}{k^2\cancel{\sigma^2}} = \frac{1}{k^2}$$

$k\sigma$

$$\boxed{P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}} \longrightarrow$$ stiamo graduando la prob. di stare oltre 1, 2, 3, ..., dev standard

dev. stand. è una sorta di unità di misura per valutare dispersione $\longrightarrow$ quante dev. standard devo contare per?

$P(|X-\mu| \leq k\sigma)$

lo avevamo fatto quando parlavamo di approx empirica

Modelli $\longrightarrow$ identificano famiglie di esperimenti

rappresenterà una distribuzione general purpose? Es: c'è diff. tra lancio di dado bilanciato, moneta, o tetraedro bilanciato?

No tutti esperim. dove spazio di prob. è equiprobabile

Una sorta di template che parametrizza tutti gli esperimenti in cui io ho uno spazio equiprobabile

Distribuzioni parametrizzate rispetto a esperim. di un certo tipo

parametro: num. esiti

Modello di Bernoulli

descriviamo esperimenti che hanno solo due esiti : SUCCESSO e FALLIMENTO

Come si modella? SUCCESSO 1 X

FALLIMENTO 0 $D_X = \{0,1\}$

e il parametro? $B(p)$ → indica che la v.a che io sto studiando è distribuita secondo un modello di Bernoulli di parametro p

$X \sim B(p)$

$p = P(\text{SUCCESSO})$

X v.a DISCRETA

$$p_X(x) = P(X=x) = \begin{cases} 1-p & x=0 \\ p & x=1 \\ 0 & \text{altrimenti} \end{cases}$$

FUNZ DI MASSA DI PROB

posso riscriverla $= p^x(1-p)^{1-x} I_{\{0,1\}}(x)$

$\boxed{E(X) = p}$ $E(X) = \sum_i x_i P(X=x_i) = 0\,\cancel{P(X=0)} + 1\,\underbrace{P(X=1)}_{=p}$

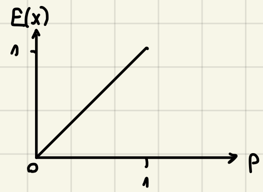
$Var(X) = E((X-p)^2) = \sum_i (x_i-p)^2 P(X=x_i)$

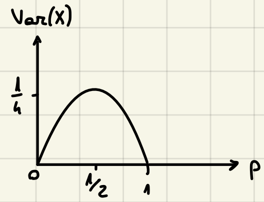
$Var(X) = E(X^2) - E(X)^2$

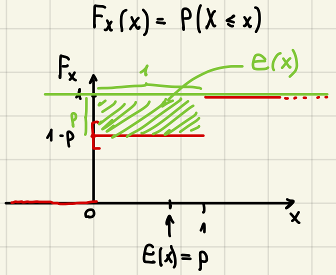
$= E(X) - E(X)^2 = p - p^2 = p(1-p)$

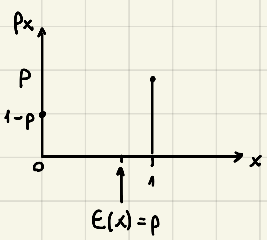
var. bernoulliana $0^2=0 \quad 1^2=1 \quad X^2=X$

$F_X(x) = P(X \leq x)$









Ipotizziamo ora di avere un esperimento Bernoulliano, e lo ripetiamo un num. infinito di volte garantendo indipendenza tra le esecuzioni

↓

Mi chiedo quante delle n volte in cui ho ripetuto l'esperimento, ho avuto un successo?

n ripetiz indipendenti di un esperimento Bernoulliano di parametro p

quanti successi → modello binomiale $B(n,p)$ $n \in \mathbb{N},\ p \in [0,1]$

(Tale conteggio mi porta a questo nuovo modello)

Immaginiamo una var. aleatoria $X \sim B(n,p)$ $D_X = \{0,1,\dots,n\}$

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

la consideriamo la prob. che i primi i esperimenti esito successo e da i+1 a n FALLIMENTO

↓

Essendo indipendenti posso scrivere

Possono essere distribuiti in modi diversi successi e fallimenti, non sono solo prima tutti successi e poi tutti fallimenti

$\underbrace{|\cdot|\ |\ |\cdot|\cdot|\cdot|\ |\ |\ |\ |\ |\ |}_{n \text{ posti}}$ $\binom{n}{i}$ → c'è una corrisp. biunivoca tra tutti i modi che io ho di generare una qualche sequenza di n valori (i dei quali successi, n-i fallimenti) con i possibili sottoinsiemi di i elementi a partire dal mio insieme di n posti

P(successo al 1° ∩ ... ∩ successo all'i-esimo ∩ fallimento all'i+1 esimo ∩ ... ∩ fallimento all'n-esimo) =

$= \underbrace{P(\text{successo al } 1°)}_{p} \cdot \ldots \cdot \underbrace{P(\text{successo all'}i\text{-esimo})}_{p} \cdot \underbrace{P(\text{fall. all'}i+1\text{ esimo})}_{(1-p)} \cdot \ldots \cdot \underbrace{P(\text{fall all'}n\text{-esimo})}_{(1-p)}$

Vediamo se tutto è corretto, che proprietà devono rispettare le funzioni di massa di probabilità?

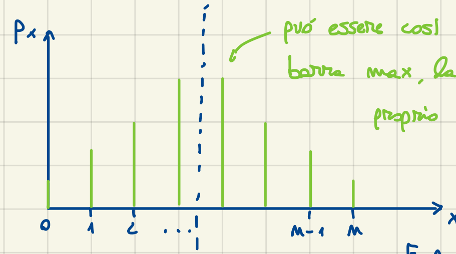
. Assume sempre valori non negativi? Sì facile da vedere

. $\sum_{i=0}^{n} p_X(i) = 1$? $\boxed{\sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i}}$ BINOMIO DI NEWTON

$$\sum_{i=0}^{n} p_X(i) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

$$\sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i} = (a+b)^n$$

$$= (\cancel{p} + 1 - \cancel{p})^n = 1^n = 1$$



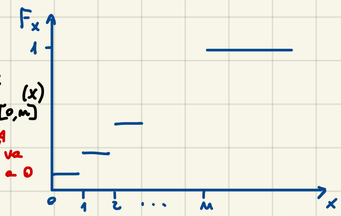
può essere così o con singola barra max, la simmetria sarebbe proprio lo

Casi particolari nel calcolo $p_X(0) = \binom{n}{0} p^0 (1-p)^n = (1-p)^n$

$$\binom{n}{0} = \frac{n!}{0!\,\cancel{n!}} = \frac{1}{0!} = \frac{1}{1} = 1 \qquad p_X(n) = \binom{n}{n} p^n (1-p)^0 = p^n$$

$$F_X(x) = P(X \le x) = \sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} p^i (1-p)^{n-i} \cdot I_{[0,n]}(x) + I_{(n,+\infty)}(x)$$

va 1   se prendo qualcosa > n   va a 0



$$E(X) = \sum_{i=0}^{n} i\, P(X=i) = \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i}$$

$$X = \sum_{i=1}^{n} X_i \qquad X_i = \begin{cases} 1 & \text{successo} \\ 0 & \text{fallimento} \end{cases} \text{ all'i-esima prova}$$

$\forall i \quad X_i \sim B(p)$

$$E(X) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = np$$

↓

valore atteso var aleatoria binomiale

$$Var(X) = Var\left(\sum_{i=1}^{n} X_i\right) \underset{\uparrow}{=} \sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

per indip. delle X_i

Esempio: P(penna difettosa) = 0.01
confezioni da 10 pezzi
rimborso se più di una difettosa

1) % scatole restituite?

↓

0.43 %

P(rimborso) = P(più di una difettosa su 10)

X = # penne difettose per confezione

$X \sim B(10, 0.01)$
↑ n   ↑ p

$$P(\text{rimborso}) = P(X > 1) = \sum_{i=2}^{10} \binom{10}{i} \cdot 0.01^i \cdot 0.99^{10-i}$$

$$= 1 - P(X \le 1) = 1 - P(X=0) - P(X=1) = 1 - \binom{10}{0} 0.01^0 \cdot 0.99^{10} - \binom{10}{1} 0.01^1 \cdot 0.99^9 \approx 0.0043 =: r$$ (prob. restituire una scatola)

(*) 3) Prob. di restituirne almeno una?

$$P(Y \ge 1) = 1 - P(Y=0) = 1 - (1-r)^3 \approx 0.01284$$

2) compro 3 scatole : qual è la prob di restituirne 1?

$Y \sim B(3, r)$

$$P(Y=1) = \binom{3}{1} r^1 (1-r)^2 \approx 0.01278$$

Y = # scatole che restituisco su 3 acquistate → se acquisto indipendente

prob. di restituire 1 scatola

(*)

Due var. aleatorie

$X_1 \sim B(n,p) \qquad X_1 = \sum_{i=1}^{n} X_{1,i} \qquad \left.\begin{array}{l} X_{1,i} \\ X_{2,j} \end{array}\right\} \sim B(p)$

$X_2 \sim B(m,p) \qquad X_2 = \sum_{j=1}^{m} X_{2,j}$

$Y := X_1 + X_2 = \sum_{i=1}^{n} X_{1,i} + \sum_{j=1}^{m} X_{2,j} = \sum_{i=1}^{n+m} Y_i \longrightarrow Y \sim B(n+m, p)$

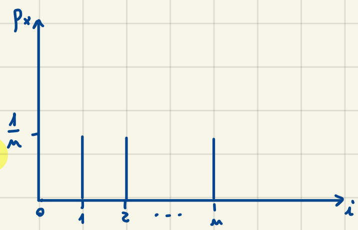
ipotizziamo
indipendenza
delle v.a e tra le v.a

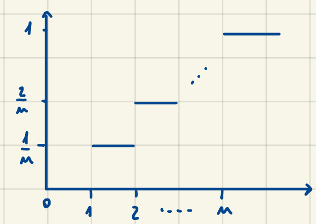
Se ho un esperimento con n ($\in \mathbb{N}$) esiti equiprobabili, numerati da 1 a n

v.a $X :=$ esito $\qquad X \sim U(n)$ MODELLO UNIFORME DISCRETO

$p_X(i) = P(X=i) = \frac{1}{n} I_{\{1,\dots,n\}}(i)$

$F_X(x) = P(X \leq x) = \sum_{i=1}^{\lfloor x \rfloor} P(X=i) = \sum_{i=1}^{\lfloor x \rfloor} \frac{1}{n} = \frac{\lfloor x \rfloor}{n} \cdot I_{[1,n]}(x) + I_{(n,+\infty)}(x)$





$E(X) = \sum_{i=1}^{n} i \boxed{P(X=i)} = \frac{1}{n} \sum_{i=1}^{n} i = \frac{1}{\cancel{n}} \frac{\cancel{n}(n+1)}{2} = \frac{n+1}{2}$

$\boxed{P(X=i)} = 1/n$

$E(X^2) = \sum_{i=1}^{n} i^2 P(X=i) = \frac{1}{n} \sum_{i=1}^{n} i^2 = \frac{1}{\cancel{n}} \frac{\cancel{n}(n+1)(2n+1)}{6}$

$\downarrow$

$Var(X) = E(X^2) - E(X)^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = (n+1)\left(\frac{2n+1}{6} - \frac{n+1}{4}\right) = (n+1) \frac{4n+2-3n-3}{12} = (n+1)\frac{n-1}{12} = \frac{n^2-1}{12}$

$\downarrow$

ci indica quanto la mia distribuz. tende ad assumere valori che sono più accorpati o lontani da $E(X)$

Dispersione scala linearmente, così come $E(X)$, rispetto al num. di esiti

---

A partire da un modello Bernoulliano, invece di conteggiare in una serie di prove ripetute di un esperimento Bernoulliano il numero di successi.

Immaginiamo di ripetere in modo non noto a priori il nostro esperimento Bernoulliano fino a che otteniamo per la prima volta successo

$\downarrow$

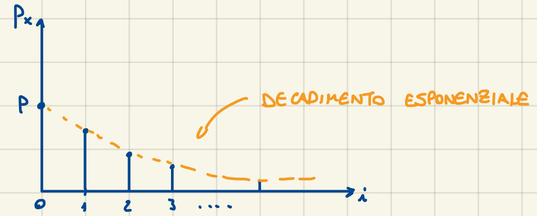
Esperimento consiste nel conteggiare quante volte dobbiamo andare avanti, quando ci fermiamo

Questo conteggio può avere un qualunque risultato all'interno dei num. naturali

Modello Geometrico $\longrightarrow$ ci sono 2 modi di ragionare :
- posso conteggiare num. tot esperimenti
- " " num tot. insuccessi;

$X = \#$ insuccessi prima del primo successo in una sequenza di esperimenti Bernoulliani con lo stesso parametro e tra loro indipendenti

$X \sim G(p) \qquad D_X = \mathbb{N} \cup \{0\} \qquad P(X=i) = (1-p)^i \cdot p \cdot I_{\mathbb{N} \cup \{0\}}(x) \qquad p \in (0,1]$



Se $p = 0$ è come dire che var. al. valore a infinito $\longrightarrow p \in (0,1]$

Se $p = 1$ " " " " " " " a 0 diventa degenere

Verifichiamo che la funz. di massa di prob rispetti le proprietà

1) Non negatività $\rightarrow$ facile

2) $\sum_{i=0}^{+\infty} P(X=i) = \sum_{i=0}^{+\infty} p(1-p)^i = p \boxed{\sum_{i=0}^{+\infty} (1-p)^i} = p \frac{1}{1-(1-p)} = \cancel{p} \frac{1}{\cancel{p}} = 1$

SERIE GEOMETRICA $\sum_{i=0}^{+\infty} \alpha^i = \frac{1}{1-\alpha}$ Nel nostro caso $\alpha = 1-p$

se $-1 \le \alpha \le 1$

$0 < p \le 1$

$0 \le 1-p < 1$ ci va bene così converge sempre

Calcoliamo prima

$$\sum_{i=0}^{+\infty} i\alpha^i = \alpha \sum_{i=0}^{+\infty} i\alpha^{i-1} = \alpha \sum_{i=0}^{+\infty} \frac{d}{d\alpha}\alpha^i = \alpha \frac{d}{d\alpha}\sum_{i=0}^{+\infty}\alpha^i = \alpha \frac{d}{d\alpha}\frac{1}{1-\alpha} = \alpha \frac{1}{(1-\alpha)^2} = \frac{\alpha}{(1-\alpha)^2}$$

$\boxed{\frac{d}{d\alpha}\alpha^i = i\alpha^{i-1}}$

$$E(X) = \sum_{i=0}^{+\infty} i\,P(X=i) = \sum_{i=0}^{+\infty} i\,p(1-p)^i = p\sum_{i=0}^{+\infty} i(1-p)^i = p\frac{(1-p)}{(\cancel{1}-\cancel{1}+p)^2} = \cancel{p}\frac{1-p}{p^{\cancel{2}}} = \frac{1-p}{p}$$

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

$$E(X^2) = \sum_{i=0}^{+\infty} i^2 p(1-p)^i = p(1-p)\sum_{i=0}^{+\infty} i^2(1-p)^{i-1} = p(1-p)\sum_{i=0}^{+\infty} i \cdot \boxed{i(1-p)^{i-1}} = -p(1-p)\sum_{i=0}^{+\infty} i \frac{d}{dp}(1-p)^i = -p(1-p)\frac{d}{dp}\sum_{i=0}^{+\infty} i(1-p)^i =$$

$-\frac{d}{dp}(1-p)^i$

$$= -p(1-p)\frac{d}{dp}\frac{1-p}{p^2} = -p(1-p)\frac{-p^2-(1-p)2p}{p^4} = \cancel{-}(1-p)\frac{\cancel{+}p\cancel{+}2(1-p)}{p^2}$$

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

$$\text{Var}(X) = E(X^2) - E(X)^2 = \frac{(1-p)(2-p)}{p^2} - \left(\frac{1-p}{p}\right)^2 = \frac{1-p}{p^2}(2-\cancel{p}-1+\cancel{p}) = \frac{1-p}{p^2}$$

Prima di vedere funz. di ripartiz.

$$P(X>n) = \sum_{i=n+1}^{+\infty} P(X=i) = \sum_{i=n+1}^{+\infty} p(1-p)^i \frac{(1-p)^{n+1}}{(1-p)^{n+1}} = (1-p)^{n+1} p \sum_{i=n+1}^{+\infty}(1-p)^{i-(n+1)} = p(1-p)^{n+1}\sum_{j=0}^{+\infty}(1-p)^j = \cancel{p}(1-p)^{n+1}\frac{1}{1-(\cancel{1}-p)} = (1-p)^{n+1}$$

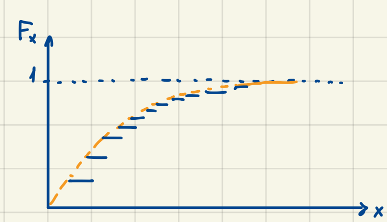
ci siamo riportati alla serie geometrica

nelle σ-algebre posso estendere quanto detto da 3A di Kolmogorov a sommatorie infinite perchè σ-algebre abbiamo chiusura

FACCIAMO CAMBIO VARIABILE $j = i-(n+1)$ con $i = n+1 \rightarrow j = 0$

$i \rightarrow +\infty \quad j \rightarrow +\infty$

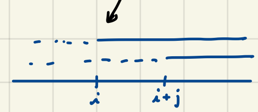
$$F_X(x) = P(X \le x) = 1 - P(X > x) = 1 - (1-p)^{x_i+1}$$



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

$$P(X \ge i+j \mid X \ge i) = \frac{P(X \ge i+j \cap X \ge i)}{P(X \ge i)} = \frac{P(X \ge i+j)}{P(X \ge i)}$$

$$= \frac{(1-p)^{i+j}}{(1-p)^i} = (1-p)^j = P(X \ge j)$$



$P(X \geq i+j \mid X \geq i) = P(X \geq j)$ **PROPRIETÀ DI ASSENZA DI MEMORIA**

Si dice che la v.a geometrica è una v.a che gode della proprietà di assenza di memoria, il fatto di avere un' info parziale sul # esperim andato male fino ad un certo punto, non mi dice nulla

DISTRIBUZIONE GEOMETRICA UNICA CHE GODE DI TALE PROPRIETÀ

Es: X = # quante volte mi ammalo in un anno

MODELLO BINOMIALE $X \sim B(12, p)$ $p = 0.1$

$$P(X=3) = \binom{12}{3} p^3 (1-p)^{12-3} \approx 0.085$$

Vado dal dott. se mi ammalo almeno 3 volte

$$P(X \geq 3) = 1 - P(X=0) - P(X=1) - P(X=2)$$
$$= 1 - \binom{12}{0} p^0 (1-p)^{12-0} - \binom{12}{1} p^1 (1-p)^{12-1} - \binom{12}{2} p^2 (1-p)^{12-2} \approx 0.11 =: p' \text{ (prob. di andare dal medico in 1 anno)}$$

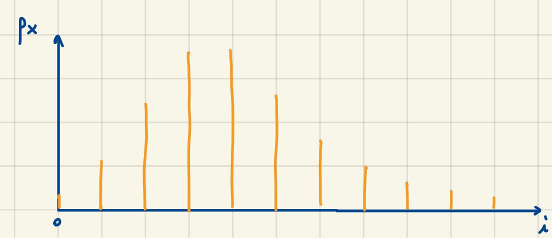
M = Vado dal medico per la prima volta tra 5 anni $\quad P(M) = P(Y=4)$ (Stiamo vedendo insuccessi, non il totale)

ora si posso usare modello geometrico $Y \sim G(p')$ $\quad P(M) = P(Y=4) = p'(1-p')^4 \approx 0.069$

dato che sono andato dal medico per l' ultima volta 2 anni fa, qual è la prob. di non andarci il prox anno?

$$P(Y \geq 3 \mid Y \geq 2) = P(Y \geq 1) = 1 - p'$$

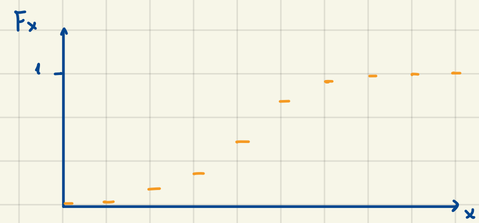
Modello di Poisson $\quad X \sim P(\lambda)$ modello discreto $D_X = \mathbb{N} \cup \{0\}$ $\quad \lambda > 0$



$$p_X(i) = P(X=i) = e^{-\lambda} \frac{\lambda^i}{i!} I_{\mathbb{N}\cup\{0\}}(i)$$

man mano che i cresce, fattoriale cresce più dell' esponenziale $\rightarrow$ prob. diventa piccola

. Non negatività c'è



$$\sum_{i=0}^{+\infty} e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} \underbrace{\boxed{\sum_{i=0}^{+\infty} \frac{\lambda^i}{i!}}}_{\text{SVILUPPO IN SERIE DI TAYLOR (MC LAUREN) di } e^{+\lambda}} = e^{-\lambda} \cdot e^{\lambda} = 1$$

PONIAM $j = i-1$

$$E(X) = \sum_{i=0}^{+\infty} i e^{-\lambda} \frac{\lambda^i}{i!} = \sum_{i=1}^{+\infty} i e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} \lambda \sum_{i=1}^{+\infty} \frac{\lambda^{i-1}}{(i-1)!} = e^{-\lambda} \lambda \sum_{j=0}^{+\infty} \frac{\lambda^j}{j!} = \lambda e^{-\lambda} \cdot e^{\lambda} = \lambda$$

$$\boxed{\sum_{i=0}^{+\infty} i e^{-\lambda} \frac{\lambda^i}{i!} = \lambda}$$

$$E(X^2) = \sum_{i=0}^{+\infty} i^2 e^{-\lambda} \frac{\lambda^i}{i!} = \sum_{i=1}^{+\infty} i^2 e^{-\lambda} \frac{\lambda^i}{i!} = \sum_{i=1}^{+\infty} i\, e^{-\lambda} \frac{\lambda^i}{(i-1)!} = \lambda \sum_{i=1}^{+\infty} \underbrace{i}_{[(i-1)+1]} e^{-\lambda} \frac{\lambda^{i-1}}{(i-1)!} = \lambda \sum_{i=1}^{+\infty} \left((i-1) e^{-\lambda} \frac{\lambda^{i-1}}{(i-1)!} + e^{-\lambda} \frac{\lambda^{i-1}}{(i-1)!} \right)$$

$j = i-1$

$$= \lambda \sum_{i=1}^{+\infty} (i-1) e^{-\lambda} \frac{\lambda^{i-1}}{(i-1)!} + \lambda \sum_{i=1}^{+\infty} e^{-\lambda} \frac{\lambda^{i-1}}{(i-1)!}$$

$$= \lambda \underbrace{\boxed{\sum_{j=0}^{+\infty} j e^{-\lambda} \frac{\lambda^j}{j!}}}_{\lambda} + \lambda \underbrace{\boxed{\sum_{j=0}^{+\infty} e^{-\lambda} \frac{\lambda^j}{j!}}}_{=1} = \lambda^2 + \lambda$$

$E(X) = \lambda \quad E(X^2) = \lambda^2 + \lambda$

$$Var(X) = E(X^2) - E(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Modello di Poisson legato a modello binomiale

l' aumento di n è bilanciato diminuendo p

ipotizzo di avere una famiglia di modelli binomiali, dove man mano aumento parametro n

se n Tende a infinito otteniamo comportamento particolare? $\longrightarrow$ Modello di Poisson

$n \cdot p = \lambda$

si bilanciano

$X \sim B(n,p)$ $\quad$ n è "grande" $\quad$ p è "piccolo" $\quad$ $np = \lambda$ $\quad$ $p = \frac{\lambda}{n}$ $\quad$ $\binom{n}{i} = \frac{n!}{i!(n-i)!}$

$$P(X=i) = \binom{n}{i} p^i (1-p)^{n-i} = \binom{n}{i}\left(\frac{\lambda}{n}\right)^i \left(1-\frac{\lambda}{n}\right)^{n-i} = \frac{n(n-1)\dots(n-i+1)}{i!}\left(\frac{\lambda}{n}\right)^i\left(1-\frac{\lambda}{n}\right)^{n-i}$$

$$= \underbrace{\frac{n}{n}}_{\to 1} \cdot \underbrace{\frac{n-1}{n}}_{\to 1} \cdot \dots \cdot \underbrace{\frac{n-i+1}{n}}_{\to 1} \; \frac{\lambda^i}{i!} \; \frac{\left(1-\frac{\lambda}{n}\right)^n \to e^{-\lambda}}{\left(1-\frac{\lambda}{n}\right)^i \to 1} \underset{n\to+\infty}{=} \frac{\lambda^i}{i!} e^{-\lambda}$$

con $n \to +\infty$

A che serve ciò? Quando n è grande potremmo dover calcolare prob. con tanti addendi, cosa che col modello di Poisson non succede

Es: In media 5 richieste/giorno $\to$ è il $E(X)$ ma $E(X) = \lambda$

In che fraz delle giornate arrivano meno di 3 richieste?

$X = \#$ richieste/giorno $\qquad X \sim P(5)$

$$P(X<3) = P(X=0) + P(X=1) + P(X=2)$$
$$= e^{-5}\left(\frac{5^0}{0!} + \frac{5^1}{1!} + \frac{5^2}{2!}\right) \approx 0.1247$$

P(in 3 gg su 5 arrivano 4 richieste)

$$P(\text{in un gg arrivano 4 richieste}) = P(X=4) = e^{-5}\frac{5^4}{4!} = 0.1755 := p$$

$$Y \sim B(5,p) \qquad P(Y=3) = \binom{5}{3} p^3 (1-p)^2 \approx 0.0367$$

Es: P(pezzo difettoso) = 0.1 $\qquad$ 10 pezzi $\qquad$ P(al più uno difettoso)

$$X \sim B(10, 0.1) \qquad P(X \le 1) = P(X=0) + P(X=1) = \binom{10}{0} 0.1^0 \cdot 0.9^{10} + \binom{10}{1} 0.1^1 \cdot 0.9^9 \approx 0.7361$$

$$X' \sim P(\lambda) \qquad \lambda = 10 \cdot 0.1 = 1 \qquad P(X' \le 1) = P(X'=0) + P(X'=1) = e^{-1}\left(\frac{1^0}{0!} + \frac{1^1}{1!}\right) = \frac{2}{e} = 0.7358$$

Errore a partire da 3ª cifra significativa

Modello ipergeometrico

descrive Tutte quelle situazioni, che possono essere modellate come l'estraz. da un'urna senza reimmissione

q.tà fisse

Urna con M oggetti "funzionanti" + N oggetti "difettosi"

$n = \#$ estrazioni senza reimmissione $\longrightarrow$ ogni estraz. modifica lo spazio di prob. non siamo più in condizione di indipendenza

$M = 5 \quad N = 1 \quad n = 2 \quad P(X=0) = 0$

$X = \#$ oggetti funzionanti $\qquad P(X=i) = \dfrac{\binom{M}{i}\cdot\binom{N}{n-i}}{\binom{N+M}{n}} \, I_{\{0,\dots,n\}}(i)$

ordine estraz. conta

APPLICO PRINCIPIO FOND. CALCOLO COMBINATORIO

se metto così può accadere $\binom{N}{n-i} = \binom{1}{2} := 0$ $\longrightarrow$ pertanto posso dire che l'intervallo è $[\max\{0, n-N\}, \min\{n, M\}]$

#oggetti funz.

M oggetti funzionanti + N ogg. guasti m estrazioni

X = # oggetti funzionanti su m estrazioni

$$P(X=i) = \frac{\binom{M}{i}\cdot\binom{N}{m-i}}{\binom{M+N}{m}}$$

casi favorevoli / # casi possibili

$F_X \to$ come binomiale bisogna sommare valori funz. massa di prob.

Es: $m=6 \quad M+N=20 \quad N=5 \quad M=15$

$$P(X\geq 4) = P(X=4 \cup X=5 \cup X=6) = P(X=4)+P(X=5)+P(X=6)$$

$$= \frac{\binom{15}{4}\binom{5}{2}+\binom{15}{5}\binom{5}{1}+\binom{15}{6}\binom{5}{0}}{\binom{20}{6}} \approx 0.8687$$

Per calcolare $E(x)$ facciamo un ragionamento decomposizionale

Introduco nuovo insieme di v.a X_i (indice legato alle estrazioni)

$$X_i = \begin{cases} 1 & \text{se } i\text{-esimo ogg. estratto funziona} \\ 0 & \text{altrimenti} \end{cases} \qquad \underset{E(X_i)}{\overset{\parallel}{P(X_i=1)}} = \frac{M}{N+M} =: p$$

$$X = \sum_{i=1}^{m} X_i \qquad E(X) = E\left(\sum_{i=1}^{m} X_i\right) = \sum_{i=1}^{m} E(X_i) = \sum_{i=1}^{m} \frac{N}{N+M} = m\frac{N}{N+M} = mp$$

ci sono analogie tra modello ipergeometrico e modello binomiale

$$Var(X_i) = E(X_i)(1-E(X_i))$$
$$= \frac{M}{N+M}\cdot\frac{N}{N+M} = \frac{N\cdot M}{(N+M)^2}$$

non posso cambiare sommatoria, non sono indipendenti le var.

$$Var(X) = Var\left(\sum_{i=1}^{m} X_i\right) = \sum_{i=1}^{m} Var(X_i) + \sum_{i\neq j} cov(X_i,X_j)$$

$$cov(X_i,X_j) = E(X_iX_j)-E(X_i)E(X_j) =$$
$$= P(X_i\cdot X_j=1) - \left(\frac{M}{N+M}\right)^2 =$$

se fossero indip. potrei scrivere prodotto come facciamo?

$$= P(X_i=1\cap X_j=1) - \left(\frac{M}{N+M}\right)^2 = P(X_j=1|X_i=1)P(X_i=1) - \left(\frac{M}{N+M}\right)^2 = \frac{M-1}{N+M-1}\,\frac{M}{N+M} - \left(\frac{M}{N+M}\right)^2 = \frac{M}{N+M}\left(\frac{M-1}{N+M-1}-\frac{M}{N+M}\right)$$

$$P(E|F) = \frac{P(E\cap F)}{P(F)}$$

$$= P(X_j=1|X_i=1)\,P(X_i=1)$$

$$= \frac{M}{N+M}\,\frac{(N+M)(M-1)-(N+M-1)M}{(N+M-1)(N+M)}$$

$$= \frac{M}{N+M}\,\frac{\cancel{NM}-N+\cancel{M^2}-\cancel{M}-\cancel{NM}-\cancel{M^2}+\cancel{M}}{(N+M-1)(N+M)} = \frac{-NM}{(N+M-1)(N+M)^2}$$

$$\parallel$$
$$cov(X_i,X_j)$$

$$Var(X) = m\frac{NM}{(N+M)^2} - m(m-1)\frac{NM}{(N+M-1)(N+M)^2} = m\frac{NM}{(N+M)^2}\left(1-(m-1)\frac{1}{N+M-1}\right)$$

$$= m\frac{N}{N+M}\,\frac{M}{N+M}\left(1-\frac{m-1}{N+M-1}\right)$$

$$= m\frac{M}{N+M}\left(1-\frac{M}{N+M}\right)\left(1-\frac{m-1}{N+M-1}\right)$$

$$= mp(1-p)\left(1-\frac{m-1}{N+M-1}\right) \xrightarrow{N+M\to+\infty} mp(1-p)$$

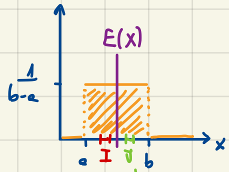
Se ci limitiamo a vedere $E(X)$ e $Var(X)$ se una pollaio molto grande, posso fregarmene che non c'è reimmissione e fare conti con modello binomiale

Modello uniforme continuo

intervallo (due arg. a e b) $\quad a < b$

v.a continua $X \sim U([a,b])$

$$f_X(x) = \frac{1}{b-a} I_{[a,b]}(x)$$

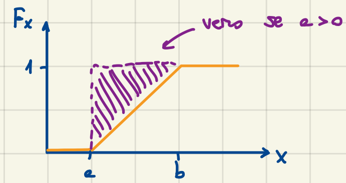


$$I \quad \int dx = \int_{i_1}^{i_2} 1 \cdot dx = x \Big|_{i_1}^{i_2}$$

$$P(X \in I) = \int_I f_X(x)\,dx = \int_I \frac{1}{b-a}\,dx = \frac{1}{b-a}\int_I dx = \frac{|I|}{b-a}$$

dipende da lunghezza dell'intervallo I non dalla sua posizione — se prendo intervallo J non cambia la prob ecco perché modello uniforme

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_a^b \frac{1}{b-a}\,dx = \frac{1}{b-a}\, x\Big|_a^b = \frac{b-a}{b-a} = 1$$



non posso chiamarla stesso modo

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(u)\,du = \int_a^x \frac{1}{b-a}\,du = \frac{1}{b-a}\, u\Big|_a^x = \frac{x-a}{b-a}$$

$$= \frac{x-a}{b-a} I_{[a,b]}(x) + I_{(b,+\infty)}(x)$$

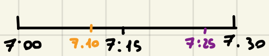
$$E(X) = \int_{-\infty}^{+\infty} x f_X(x)\,dx = \int_a^b \frac{x}{b-a}\,dx = \frac{1}{b-a}\int_a^b x\,dx = \frac{1}{b-a}\frac{x^2}{2}\Big|_a^b = \frac{1}{b-a}\frac{b^2-a^2}{2} = \frac{1}{\cancel{b-a}}\frac{(b+a)\cancel{(b-a)}}{2} = \frac{b+a}{2}$$

→ punto medio dell'intervallo su cui è definita la v.a

$$E(X^2) = \int_a^b x^2 f_X(x)\,dx = \frac{1}{b-a}\int_a^b x^2\,dx = \frac{1}{b-a}\frac{x^3}{3}\Big|_a^b = \frac{b^3-a^3}{(b-a)3} = \frac{\cancel{(b-a)}(a^2+ab+b^2)}{\cancel{(b-a)}3} = \frac{a^2+ab+b^2}{3}$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} = \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4} = \frac{4a^2+4ab+4b^2-3a^2-6ab-3b^2}{12} = \frac{a^2-2ab+b^2}{12} = \frac{(b-a)^2}{12}$$

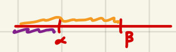
Es:



X = minuto dalle 7:00 nel quale arrivo alla fermata

$X \sim U([0,30])$

$$P(\alpha < X \le \beta) = F(\beta) - F(\alpha)$$



disgiunte

$$P(\text{attendo meno di 5 minuti}) = P(X=0 \cup 10<X\le 15 \cup 25<X\le 30) = P(\cancel{X=0})^{0} + P(10\le X\le 15) + P(25\le X\le 30)$$

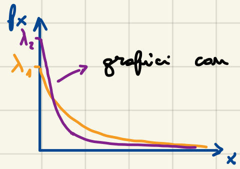
$$= F_X(15) - F_X(10) + F_X(30) - F_X(25)$$

$$\frac{15-0}{30-0} - \frac{10}{30} + \frac{30}{30} - \frac{25}{30} = \boxed{\frac{15-10}{30}} + \boxed{\frac{30-25}{30}} = \frac{1}{3}$$

$P(X \ge 30)$

Modello esponenziale → parametrizzato come gli altri modelli (parametro $\lambda \in \mathbb{R}^+$) $x \in \mathbb{R}^+$

$f_X(x) = \lambda e^{-\lambda x} I_{\mathbb{R}^+}(x)$



grafici con λ diversi non staranno mai completamente sotto o sopra ad altri perché l'integrale di f_X deve fare 1

Cosa succederebbe se io adesso cambiassi λ? $\lambda_2 > \lambda_1$

più λ grande più decresce rapidamente

Per cosa si usa il modello esponenziale?

↓

Usato per modellare il tempo che intercorre tra due eventi nel caso in cui gli eventi si manifestano secondo alcune ipotesi

---

$$\int_0^{+\infty} f_X(x)\,dx = \int_0^{+\infty} \lambda e^{-\lambda x}\,dx = \int_0^{+\infty} e^{-y}\,dy = -e^{-y}\Big|_0^{+\infty} = 0 + e^{-0} = 1$$

sost $y = \lambda x$

$dy = \lambda dx$

$$F_X(x) = \int_0^x f_X(y)\,dy = \int_0^x \lambda e^{-\lambda y} dx = \int_0^{\lambda x} e^{-z}\,dz = -e^{-z}\Big|_0^{\lambda x} = -e^{-\lambda x} + e^0 = 1 - e^{-\lambda x}$$

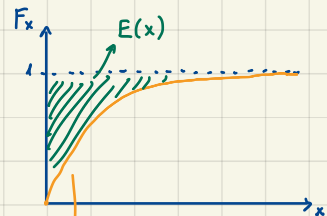
sost $z = \lambda y$

$dz = \lambda dy$

vero solo da 0 a $+\infty$ → per tutti gli argomenti?

↓

$F_X(x) = (1 - e^{-\lambda x}) I_{\mathbb{R}^+}(x)$



se aumento λ va ad 1 più velocemente

Aumentando λ diminuisce l'area sopra la curva → diminuisce valore atteso

$$E(X) = \int_0^{+\infty} x f_X(x)\,dx = \int_0^{+\infty} \underbrace{x}_{f}\,\underbrace{\lambda e^{-\lambda x}}_{g'}\,dx = \cancel{x e^{-\lambda x}}\Big|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} = \int_0^{+\infty} e^{-\lambda x}\,dx = \frac{1}{\lambda}\boxed{\int_0^{+\infty} \lambda e^{-\lambda x}\,dx} = \frac{1}{\lambda}$$

integro per parti; $g(x) = -e^{-\lambda x}$

$g'(x) = \lambda e^{-\lambda x}$

=1 per proprietà della funz. di densità

è lo stesso

$$Var(X) = E(X^2) - E(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

$= 1/\lambda$

$$E(X^2) = \int_0^{+\infty} \underbrace{x^2}_{f}\,\underbrace{\lambda e^{-\lambda x}}_{g'}\,dx = \cancel{x^2 e^{-\lambda x}}\Big|_0^{+\infty} + \int_0^{+\infty} 2x e^{-\lambda x}\,dx = 2\int_0^{+\infty} x e^{-\lambda x}\,dx = \frac{2}{\lambda}\boxed{\int_0^{+\infty} \lambda x e^{-\lambda x}\,dx} = \frac{2}{\lambda^2}$$

Aumentando λ, densità decade più rapidamente a 0 → la mia v.a assume quasi sempre valori vicini a zero

La distribuzione esponenziale e quella geometrica godono di una stessa proprietà (ASSENZA DI MEMORIA)

correlate

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

$P(X > x) = 1 - F_X(x) = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x}$

$P(X > s+t) = e^{-\lambda(s+t)} = e^{-\lambda s} e^{-\lambda t} = P(X > s) \cdot P(X > t)$

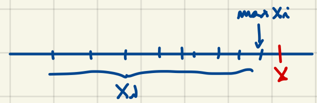
$P(X > s) = \dfrac{P(X > s+t)}{P(X > t)}$

voglio dim. $P(X > s+t \mid X > t) = P(X > s)$ ASSENZA MEMORIA

$= \dfrac{P(X > s+t \cap X > t)}{P(X > t)}$

Immaginiamo di avere n v.a $X_1, \dots, X_n$ indipendenti

$Y := \max_i X_i$ $\longrightarrow$ riusciamo a dire qualcosa sulla distribuzione di Y?



$$F_Y(x) = P(Y \leq x) = P(\max_i X_i \leq x) = P(\forall i \ X_i \leq x) = \boxed{\prod_i^n P(X_i \leq x)} = \prod_i^n F_{X_i}(x)$$

$P(X_i \leq x)$ " $F_{X_i}(x)$

abbiamo ipotesi di indipendenza

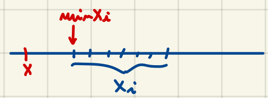
$\max X_i \leq x \longleftrightarrow \forall i \ X_i \leq x$

Se le mie v.a non fossero solo indipendenti, ma anche identicamente distribuite

seguono stesso modello con stesso param $\longrightarrow$ stessa funz. di ripartizione

$(X_1, \dots, X_n$ i.i.d secondo $F \longrightarrow F_Y(x) = \prod_{i=1}^n F(x) = F(x)^n)$

$X_1, \dots, X_n$ indipendenti e basta $\quad Z := \min_i X_i$



$(Z > x) \leftrightarrow \min_i X_i > x \longleftrightarrow \forall i \quad x < X_i$

$$P(Z > x) = P(\forall i \ X_i > x) = \prod_i^n P(X_i > x) = \prod_i^n (1 - P(X_i \leq x)) = \prod_i^n (1 - F_{X_i}(x))$$

$\parallel$

$1 - P(Z \leq x) = 1 - F_Z(x) \qquad F_Z(x) = 1 - \prod_i^n (1 - F_X(x))$

E se le X_i ora fossero i.i.d secondo $F \longrightarrow F_Z(x) = 1 - (1 - F(x))^n$

Immaginiamo v.a indipendenti, seguono modello comune (ma con parametro che cambia)

il modello è quello esponenziale

$X_i \sim \mathcal{E}(\lambda_i)$ + INDIPENDENTI

$F_Z(x) = 1 - \prod_i^n \underbrace{(1 - F_{X_i}(x))}$ $\qquad F_{X_i}(x) = 1 - e^{-\lambda_i x}$ $\qquad \lambda := \sum_{i=1}^n \lambda_i$

$$F_Z(x) = 1 - \prod_i^n e^{-\lambda_i x} = 1 - e^{\sum_{i=1} -\lambda_i x} = 1 - e^{-x \sum_{i=1} \lambda_i} = 1 - e^{-\lambda x}$$ FORMA ANALITICA D. DISTRIBUZ. ESPONENZIALE DI PARAMETRO λ

$\Rightarrow Z \sim \mathcal{E}(\lambda)$ $\qquad X_1, \dots, X_n$ INDIP + $\forall i \ X_i \sim \mathcal{E}(\lambda_i)$ $\longrightarrow$ modelle il minimo tra i tempi dei server

$\min X_i \sim \mathcal{E}(\sum_{i=1}^n \lambda_i)$

$X \sim \mathcal{E}(\lambda) \qquad c \in \mathbb{R}^+ \qquad Y := c \cdot X$

$$F_Y(x) = P(Y \leq x) = P(c \cdot X \leq x) = P(X \leq \tfrac{x}{c}) = F_X(\tfrac{x}{c}) = 1 - e^{-\lambda \frac{x}{c}} = 1 - e^{-\frac{\lambda}{c} x}$$ dimostrato che $Y \sim \mathcal{E}(\frac{\lambda}{c})$

Es: X = n° ore di riparazione $\sim E(1)$

$P(X>2) = 1 - F_X(2) = (1 - e^{-2}) = e^{-2} \approx 0.1353$

$X>3$? $P(X>\underset{1+2}{3} \mid X>2)$ ASS. MEMORIA $\rightarrow P(X>1) = e^{-1} \approx 0.3678$

Es 2:

X = n° migliaia di miglia prima di guastarsi $\sim E(\frac{1}{20})$

$P(X>\underset{20+10}{30} \mid X>10)$ ASS. MEMORIA $\rightarrow P(X>20) = e^{-20/20} = e^{-1} \approx 0.3678$

Come cambia prob. se var descritta modo uniforme continuo

$X \sim U((0,40))$ NON HO PIÙ ASS. DI MEMORIA

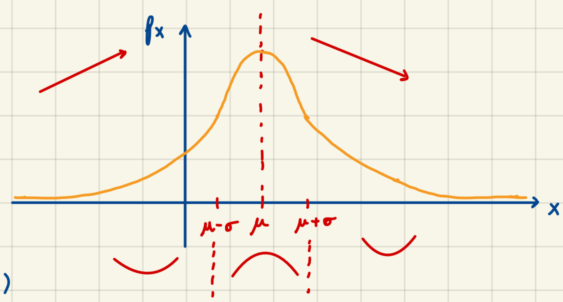
$F_X(x) = \frac{x-0}{40-0}$

$$P(X>30 \mid X>10) = \frac{P(X>30 \cap X>10)}{P(X>10)} = \frac{P(X>30)}{P(X>10)} = \frac{P(30<X<40)}{P(10<X<40)} = \frac{F_X(40)-F_X(30)}{F_X(40)-F_X(10)} = \frac{40-30}{40-10} = \frac{1}{3}$$

Modello Gaussiano/Normale

Specificato da 2 parametri $X \sim G(\mu, \sigma)$ $\mu \in \mathbb{R}$ $\sigma \in \mathbb{R}^+$ $D_X = \mathbb{R}$ NO FUNZIONE INDICATRICE

Funzione di densità: $f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$



$\lim_{x \to \pm\infty} f_X(x) = 0$

$$f'_X(x) = \frac{-1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \frac{1}{\cancel{2}\sigma^2}\cancel{2}(x-\mu) = \frac{1}{\sqrt{2\pi}\,\sigma^3} e^{-\frac{(x-\mu)^2}{2\sigma^2}}(\mu - x)$$

$f'_X(x) \geq 0 \quad \mu - x \geq 0 \quad x \leq \mu$

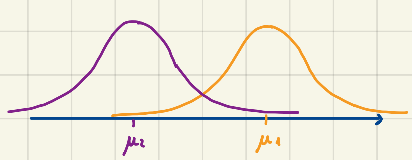
$$f''_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma^3}\left(e^{-\frac{(x-\mu)^2}{2\sigma^2}}\left(-\frac{\cancel{2}(x-\mu)}{\cancel{2}\sigma^2}\right)(\mu-x) - e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) = \frac{1}{\sqrt{2\pi}\,\sigma^3} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\left(+\frac{(x-\mu)^2}{\sigma^2} - 1\right)$$

$f''_X(x) \geq 0 \quad \frac{(x-\mu)^2}{\sigma^2} - 1 \geq 0 \quad \left(\frac{x-\mu}{\sigma}\right)^2 \geq 1$

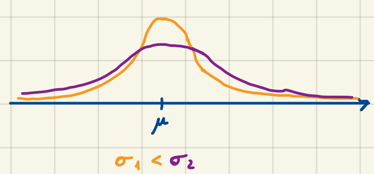
$\rightarrow \frac{x-\mu}{\sigma} \geq 1 \rightarrow x \geq \mu + \sigma$

$\rightarrow \frac{x-\mu}{\sigma} \leq -1 \rightarrow x \leq \mu - \sigma$

Al variare di μ traslo la curva a sx o dx



Al variare di σ varia l'ampiezza



$f_X(\mu) = \frac{1}{\sqrt{2\pi}\,\sigma} \cancel{e^{-\frac{(\mu-\mu)^2}{2\sigma^2}}}$ $\rightarrow$ densità in μ tanto più grande quanto più σ piccola

Devo dimostrare che: $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$ $\qquad$ $F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(u-\mu)^2}{2\sigma^2}} du = 1$

è una densità, non negatività ok

fare questo integrale in tal modo è inutile

$E(X) = \mu$

$Var(X) = \sigma^2$

Proprietà: .) v.a $X \rightarrow aX + b := Y$ se $X \sim G(\mu, \sigma)$

$\downarrow$

$Y \sim G(a\mu + b, |a|\sigma)$

.) $X_1 \sim G(\mu_1, \sigma_1)$ $\qquad X_1 + X_2 \sim G(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$

$X_2 \sim G(\mu_2, \sigma_2)$

INDIPENDENTI

$Y = \frac{X - \mu}{\sigma}$ $\qquad E(Y) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}(X - \mu) = \frac{1}{\sigma}(E(X) - \mu) = 0$

applico trasf. lineare, questa trasf. lineare è la STANDARDIZZAZIONE

$Var(Y) = Var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2} Var(X - \mu) = \frac{1}{\sigma^2} Var(X) = 1$

$\downarrow$

$Y = \frac{X-\mu}{\sigma} = \underset{a}{\frac{1}{\sigma}} x - \underset{b}{\frac{\mu}{\sigma}}$ $\qquad$ se $X \sim G(\mu, \sigma) \longrightarrow Y \sim G(0,1)$

Altra notazione per

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

GAUSSIANA STANDARD/NORMALE STANDARD $\rightarrow Z := \frac{X-\mu}{\sigma} \sim G(0,1)$

$X_1, \dots, X_n$ indipendenti $\forall i \sim G(\mu_i, \sigma_i)$ $\quad Y = \sum_{i=1}^{n} X_i$ $\quad E(Y) = \sum_{i=1}^{n} \mu_i$ $\quad Var(Y) = \sum_{i=1}^{n} \sigma_i^2$ $\quad Y \sim G\left(\sum_{i=1}^{n} \mu_i, \sqrt{\sum_{i=1}^{n} \sigma_i^2}\right)$

solo se indipendenti

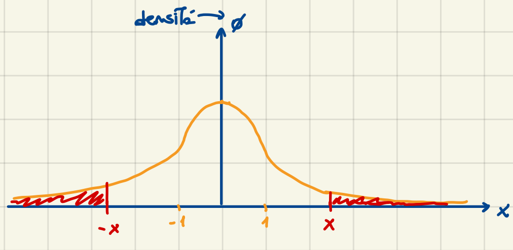
$\downarrow$

RIPRODUCIBILITÀ: sommando insieme v.a che hanno un modello comune si ottiene una v.a che fa ancora parte di tale modello (BINOMIALE)

$F_X(x) = P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = F_Z\left(\frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

$\Phi(0) = \frac{1}{2}$ $\qquad \Phi(-x) = P(Z > x) = 1 - P(Z < x) = 1 - \Phi(x)$

è simmetrica



Qual è la probabilità che un'osservazione della mia v.a ricada entro una deviazione standard?

NORMALE STANDARD

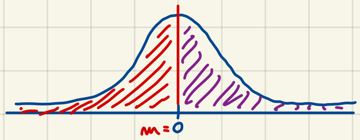
$P(-1 \leq Z \leq 1) = P(|Z| \leq 1) = \Phi(1) - \Phi(-1)$

GENERICA V.A

$P(|X - \mu| \leq n\sigma) = P(|X| \leq n) = \Phi(n) - \Phi(-n) = \Phi(n) - (1 - \Phi(n)) = 2\Phi(n) - 1$

Moda si può applicare non solo come indice di centralità per i campioni, ma si può applicare anche per delle v.a

MEDIANA di una v.a X? valore $m \in \mathbb{R}$ T.c $P(X \leq m) = P(X \geq m) = \frac{1}{2}$

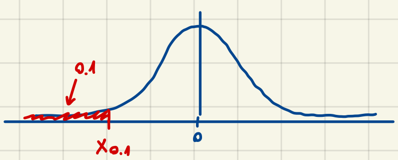


Moda di v.a X? Specificazione che massimizza la densità o la massa di prob.

↓

per una gaussiana ancora 0

Quantile di livello $q \in [0,1]$ di v.a X? Specificazione $x \in \mathbb{R}$ T.c $P(X \leq x) = q$ e $P(X \geq x) = 1-q$



$X \sim G(\mu, \sigma)$

σ è un quantile di X : qual è il suo livello?

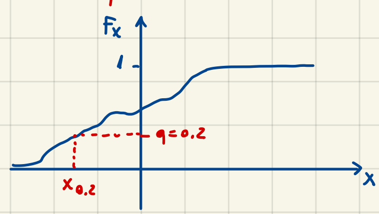
$q = P(X \leq x_q) = F_X(x_q)$

$\boxed{q = F_X(x_q)}$ → dato il livello del quantile capire il valore del quantile? da x_q ottenere q

python ppf

Funz. inversa

$F_X^{-1}(q) = F_X^{-1}(F_X(x_q)) = x_q$ $\qquad x_q = F_X^{-1}(q)$



Non ho problemi per distribuzioni continue, ma ho per quelle discrete

Teorema centrale del limite ← estensione per qualsiasi distribuzione di v.a ← **Teorema di DeMoivre-Laplace**

maggiore n, migliore l'approssimaz. a normale

$X \sim B(n,p) \qquad X \sim G(np, \sqrt{np(1-p)})$

Una binomiale la posso approx come una gaussiana

$X_1, \dots, X_n \sim B(p)$ indip

$Y = \sum_{i=1}^{n} X_i \sim B(n,p) \mathrel{\dot\sim} G(np, \sqrt{np(1-p)})$

↓

Successione di n v.a $X_1, \dots, X_n$ i.i.d $\quad \forall i \quad E(X_i) = \mu \quad Var(X_i) = \sigma^2$

$\sum_{i=1}^{n} X_i \mathrel{\dot\sim} G(n\mu, \sqrt{n}\sigma)$

APPROX. DISTRIBUITO COME

come standardizzo? $\quad \dfrac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \mathrel{\dot\sim} G(0,1)$

aumentando n, migliora l'approssimazione

$$\lim_{n \to +\infty} P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \leq x \right) = \Phi(x)$$

FUNZ. DI RIPARTIZIONE DELLA NORMALE STANDARD

Es: 25000 polizze

X = risarcimento annuo di un cliente

$E(X) = 320 \qquad \sigma_X = 540$

$P($più di $8.3 \cdot 10^6$€ di risarcimento in un anno$)$

X_i = risarcimento annuo del cliente i

Y = risarcimento erogati in un anno $= \sum_{i=1}^{25000} X_i$

Possiamo applicare Teo. centrale del limite

$$Y \mathrel{\dot\sim} G(\overset{8 \cdot 10^6}{\overline{2500 \cdot 320}}, \overset{\approx 8.54 \cdot 10^6}{\overline{\sqrt{25000} \cdot 540}})$$

$$P(Y > 8.3 \cdot 10^6) = P\left(\underbrace{\boxed{\frac{Y - 8 \cdot 10^6}{8.54 \cdot 10^6}}}_{\dot\sim N(0,1)} > \underbrace{\boxed{\frac{8.3 \cdot 10^6 - 8 \cdot 10^6}{8.54 \cdot 10^6}}}_{\approx 3.51}\right) \approx P(\underbrace{Z}_{\sim N(0,1)} > 3.51) = 1 - \Phi(3.51) \approx 2.2 \cdot 10^{-4}$$

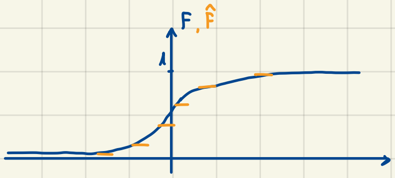
z = st.norm()

z.rvs()

Funzione di ripartizione empirica $\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty, x]}(x_i) \quad \{x_1, \dots, x_n\}$

$\downarrow$

legata alla funzione di ripartizione



$X \sim B(n,p) \quad n$ "grande" $\quad X = \sum_{i=1}^{n} \overset{\sim B(p)}{X_i} \mathrel{\dot\sim} G(np, \sqrt{np(1-p)})$

STANDARDIZZIAMO $\quad \dfrac{X - np}{\sqrt{np(1-p)}} \mathrel{\dot\sim} G(0,1)$

Es:

30% degli iscritti è frequentante $\qquad P($uno studente sia frequentante$) = 0.3$

X = # studenti freq. su 450 $\sim B(450, 0.3)$

$P(X > 150) = \sum_{x=151}^{450} \binom{450}{x} 0.3^x \, 0.7^{450-x}$ TEDIOSO $\rightarrow$ approssimiamo $X \mathrel{\dot\sim} G(\overset{135}{\overline{450 \cdot 0.3}}, \overset{\approx 9.72}{\overline{\sqrt{450 \cdot 0.3 \cdot 0.7}}})$

$\downarrow$

$$= P(X > 150) = P\left(\underbrace{\frac{X - 135}{9.72}}_{\sim G(0,1)} > \overset{\approx 1.54}{\overline{\frac{150 - 135}{9.72}}}\right) \approx P(\overset{\sim G(0,1)}{Z} > 1.54) \approx 0.06$$

Es: $\mu = 12.08$ pollici $\quad \sigma = 3.1$ pollici

X_{2022} = precip del 2022, misurate in pollici a L.A

X_{2023} = " " 2023, " " " " "

$X_{2022}, X_{2023} \sim G(\mu, \sigma) \qquad Y = X_{2022} + X_{2023} \sim G(2\mu, \sqrt{2}\sigma)$

$$P(X_{2022} + X_{2023} > 25) = P(Y > 25) = P\left(\frac{Y - 2\cdot 12.08}{\sqrt{2}\cdot 3.1} > \frac{25 - 2\cdot 12.08}{\sqrt{2}\cdot 3.1}\right) = P(Z > 0.1916) \approx 0.4240$$

approx

$$P(X_{2022} > X_{2023} + 3) = P(X_{2022} - X_{2023} > 3) = P\left(\frac{X_{2022} - X_{2023}}{\sqrt{2}\sigma} > \frac{3}{\sqrt{2}\sigma}\right) = P(Z > 0.6843) = ?$$

$T = X_{2022} - X_{2023} \qquad E(T) = 0 \qquad Var(T) = Var(X_{2022} - X_{2023}) = Var(X_{2022}) + Var(X_{2023}) = 2\sigma^2$

Statistica inferenziale

↓

cerchiamo di fare un' estrazione di informazione in senso induttivo

POPOLAZIONE = v.a $X \sim F$

CAMPIONE: successione di n v.a i.i.d $X_1, \dots, X_n$ (come X)

$X_1, \dots, X_n$ campione $x_1, \dots, x_n$

Perché usiamo le v.a?

Perché ci manca conoscenza di F

STATISTICA/STIMATORE ⟶ mi fornisce qualcosa di F

↓

funzione del campione $\tau : D_X^n \longrightarrow \mathbb{R} \qquad \tau(X_1, \dots, X_n)$

INPUT: campione

OUTPUT: stima di qualcosa che non conosco

Se F completamente sconosciuta **STATISTICA NON PARAMETRICA**

F "parzialmente" conosciuta ⟶ diciamo che la va che descrive la popolazione segue un certo modello

ma

non conosciamo tutti i parametri di quel modello

STATISTICA INFERENZIALE PARAMETRICA

$X \sim F(\theta)$

parametro ignoto

Voglio stimare $\tau(\theta)$ un qualcosa che dipende da θ

$T = \tau(X_1, \dots, X_n)$

$\hat{\tau} = \tau(x_1, \dots, x_n)$ **STIMA**

$\hat{\tau} \approx \tau(\theta)$

Popolaz. $X \sim B(p)$ $\theta = p$ $\tau(\theta) = \tau(p) = p$

Campione $x_1, \dots, x_n$

statistica $T(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$ $\bar{x} \approx p$

$$T(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X}$$

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n} \sum_{i=1}^{n} E(X) = \frac{1}{n} \sum_{i=1}^{n} p = \frac{1}{\cancel{n}} \cdot \cancel{n} p = p$$

Se ho un parametro θ $\tau(\theta) = \theta$

$T(X_1, \dots, X_n) = T$ $\longrightarrow$ STIMATORE NON DISTORTO/NON DEVIATO di $\tau(\theta)$

$E(T) = \tau(\theta)$

Es: X: # minuti attesi in coda Popolazione $X \sim \mathcal{E}(\lambda)$

$\theta = \lambda$ voglio stimare il tempo medio di attesa $E(\lambda) = \frac{1}{\lambda} = \tau(\lambda)$

$T(X_1, \dots, X_n) = \bar{X}$

$$E(\bar{x}) = E\left(\frac{1}{n} \sum_i X_i\right) = \frac{1}{n} \sum_i E(X_i) = \frac{1}{n} \sum_i E(X) = E(X)$$

$\bar{X}$ sempre non deviato rispetto a $E(X)$

---

popolazione $X \sim B(m, p)$ conosco m, ignoro p

$\theta = p$ $\tau(\theta) = \tau(p) = p$ $E(\bar{X}) = E(X) = mp$ $\longrightarrow$ non è uno stimatore non distorto

$$X_1, \dots, X_n \rightsquigarrow \overbrace{\boxed{\underbrace{X_{1,1}}_{Y_1}, \dots, X_{1,m}}}^{X_1}, \dots, X_{n,1}, \dots, \underbrace{X_{n,m}}_{Y_{n \cdot m}}$$

$$\bar{X} = \bar{Y} = \frac{1}{n \cdot m} \sum_{i=1}^{n \cdot m} Y_i = \frac{1}{n \cdot m} \sum_{i=1}^{n} X_i = \frac{1}{m} \underbrace{\boxed{\frac{1}{n} \sum_{i=1}^{n} X_i}}_{\bar{X}} = \frac{1}{m} \bar{X}$$

OPPURE

$T = \frac{1}{m} \bar{X}$ è non deviato rispetto a p

$E(\bar{X}) = mp$

$E\left(\frac{1}{m} \bar{X}\right) = \frac{1}{m} E(\bar{X}) = p$

$X \sim \mathcal{E}(\lambda)$ $\theta = \lambda$ $\tau(\theta) = \tau(\lambda) = \lambda$

$\bar{X}$ stimatore non deviato di $E(X) = \frac{1}{\lambda}$

$E(\bar{X}) = \frac{1}{\lambda}$ $\lambda = \frac{1}{E(\bar{X})}$ $\longrightarrow$ da qui non riesco ad andare avanti

$$Var(\bar{X}) = Var\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} \sum_{i=1}^{n} \underbrace{Var(X)}_{\sigma^2} = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}$$

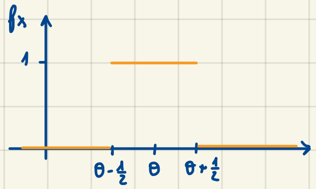
le oscillazioni che la mia stima fa quando uso media campionaria attorno al suo valore centrale

$\downarrow$

crescono al crescere della varianza d. popolazione e diminuiscono al crescere degli oggetti nel mio campione

Ci sono casi in cui si può risolvere problema della stima in modo esatto

$\theta \in \mathbb{N} \quad X \sim U\left(\left(\theta-\frac{1}{2}, \theta+\frac{1}{2}\right)\right)$



$x_1 = 3.1$

$\theta = 3$ l'intero più vicino

MSE (Mean Square Error / Errore quadratico medio)

mi dice quanto è buono uno stimatore per stimare una quantità numerica

$$MSE_{\tau(\theta)}(T) = E\left((T-\tau(\theta))^2\right) \qquad T = t(X_1, \dots, X_n)$$

Relazione tra MSE e il fatto che lo stimatore sia non deviato

$$MSE_{\tau(\theta)}(T) = E\left((T-\tau(\theta))^2\right) = E\left((\underbrace{T-E(T)} + \underbrace{E(T)-\tau(\theta)})^2\right)$$

$$= E\left((T-E(T))^2 + 2(T-E(T))(E(T)-\tau(\theta)) + (E(T)-\tau(\theta))^2\right)$$

$$= \underbrace{E\left((T-E(T))^2\right)}_{= Var(T)} + \cancel{2(E(T)-\tau(\theta))\underbrace{E(T-E(T))}_{=0}} + (E(T)-\tau(\theta))^2$$

$E(T-E(T)) = E(T) - E(E(T)) = E(T) - E(T) = 0$

$$= Var(T) + (E(T)-\tau(\theta))^2 \longrightarrow \text{Introduciamo Bias } b_{\tau(\theta)}(T) = E(T)-\tau(\theta)$$

$$= Var(T) + b_{\tau(\theta)}(T)^2$$

T stimatore non deviato per $\tau(\theta) \longrightarrow b_{\tau(\theta)}(T) = 0$

$E(T) = \tau(\theta)$

$MSE_{\tau(\theta)}(T) = Var(T)$

Calcolo dell'MSE ci dice se uno stimatore oscilla tanto o poco attorno al valore da stimare

MSE ci permette di definire un'altra proprietà desiderabile per uno stimatore

CONSISTENZA IN MEDIA QUADRATICA: uno stimatore T è consistente in media quadratica per $\tau(\theta)$

sse $\boxed{\lim_{n\to+\infty} MSE_{\tau(\theta)}(T) = 0}$

delicato cambiando n $\quad T = t(X_1, \dots, X_n)$ cambiando n cambia #argomenti di t

Consideriamo una famiglia di stimatori $\{T_n\}$ indice taglia del campione

per ogni T_n posso calcolare l'MSE

$$\forall T_n \longrightarrow MSE_{\tau(\theta)}(T_n) \xrightarrow{n\to+\infty} 0$$

CONSISTENZA DEBOLE: Uno stimatore $T_n = T(X_1, \dots, X_n)$ é debolmente consistente rispetto ad una quantità ignota $\tau(\theta)$

sse $\tau(\theta) - \varepsilon \leq T_n \leq \tau(\theta) + \varepsilon$

↓

misura l'incertezza che io ho

andiamo a vedere la prob. di questo evento $P(\tau(\theta) - \varepsilon \leq T_n \leq \tau(\theta) + \varepsilon)$

$$\forall \varepsilon > 0 \quad \lim_{n \to +\infty} P(\tau(\theta) - \varepsilon \leq T_n \leq \tau(\theta) + \varepsilon) = 1$$

$\lim MSE_{\tau(\theta)}(T) = 0$

Se T é consistente in media quadratica per $\tau(\theta)$ $\Rightarrow$ T é anche debolmente consistente per $\tau(\theta)$

$$P(\tau(\theta) - \varepsilon \leq T_n \leq \tau(\theta) + \varepsilon) = P(-\varepsilon \leq T_n - \tau(\theta) \leq \varepsilon) = P(|T_n - \tau(\theta)| \leq \varepsilon) = P((T_n - \tau(\theta))^2 \leq \varepsilon^2)$$

$$= \underbrace{1 - P((T_n - \tau(\theta))^2 > \varepsilon^2) > 1 - \frac{\overbrace{E((T_n - \tau(\theta))^2)}^{= MSE_{\tau(\theta)}(T_n)}}{\varepsilon^2}}_{\text{DISUGUAGLIANZA MARKOV}}$$

$$= 1 - \frac{MSE_{\tau(\theta)}(T_n)}{\varepsilon^2} \xrightarrow{n \to +\infty} 1 \quad \text{indipendentemente dalla scelta di } \varepsilon$$

Es:

$X \sim B(p) \qquad \theta = p \qquad \tau(\theta) = \tau(p) = p \qquad T = T(X_1, X_2, X_3) = \frac{X_1 + X_2 + X_3}{5}$ ⟶ questo stimatore é corretto per p?

$E(T) = E\left(\frac{X_1 + X_2 + X_3}{5}\right) = \frac{1}{5} E(X_1 + X_2 + X_3) = \frac{3}{5} p$ é distorto per p Quanto é il suo bias/distorsione?

$$b_p(T) = E(T) - p = \frac{3}{5}p - p = -\frac{2}{5}p$$

$$MSE_p(T) = Var(T) + b_p(T)^2 = Var\left(\frac{1}{5}(X_1 + X_2 + X_3)\right) + \frac{4}{25}p^2 = \frac{3}{25} Var(X) + \frac{4}{25}p^2 = \frac{3}{25} p(1-p) + \frac{4}{25}p^2$$

Es 2:

$X \sim U((0, \theta)) \qquad \tau(\theta) = \theta$

Proviamo a usare come stimatore la media campionaria

$E(\bar{X}) = E(X) = \frac{\theta}{2}$ $\quad \bar{X}$ é deviato per θ

$2E(\bar{X}) = \theta$ $\quad 2\bar{X}$ non é deviato per θ

$T = 2\bar{X}$ $\quad T$ é non deviato per θ $\quad MSE_\theta(T) = Var(T) = Var(2\bar{X}) = 4\,Var(\bar{X}) = \frac{4}{n} Var(X) = \frac{4}{n} \frac{\theta^2 - 0^2}{12} = \frac{\theta^2}{3n} \xrightarrow{n \to +\infty} 0$

T é consistente in media quadratica per θ

E se volessi considerare come stimatore $T' = \max_{1 \leq i \leq n} X_i$

$T' = \max_{1 \leq i \leq n} X_i \qquad F_{T'}(x) = P(T' \leq x) = P(\max X_i \leq x) = P(\forall i \; X_i \leq x) = \prod_{i=1}^{n} P(X_i \leq x) = \prod_{i=1}^{n} F_X(x)$

$$= F_X(x)^n = \left(\frac{x}{\theta}\right)^n \rightarrow f_{T'}(x) = \frac{d}{dx} F_{T'}(x) = \frac{d}{dx}\left(\frac{x}{\theta}\right)^n = n\left(\frac{x}{\theta}\right)^{n-1} \frac{1}{\theta} \qquad f_{T'}(x) = n \frac{x^{n-1}}{\theta^n}$$

$$E(T') = \int_0^\theta x f_{T'}(x)\,dx = \int_0^\theta x\, n \frac{x^{n-1}}{\theta^n}\,dx = \frac{n}{\theta^n} \int_0^\theta x^n\,dx = \frac{n}{\theta^n} \left.\frac{x^{n+1}}{n+1}\right|_0^\theta = \frac{n}{\theta^n} \frac{\theta^{n+1}}{n+1} = \frac{n}{n+1}\theta$$

T' é deviato per θ

$$T'' = \frac{n+1}{n} T' \quad E(T'') = \frac{n+1}{n} E(T') = \cancel{\frac{n+1}{n}} \cancel{\frac{n}{n+1}} \theta = \theta$$

$MSE(T'') = Var(T'') = E(T''^2) - E(T'')^2$ $\qquad$ $E(T''^2) = E\left(\left(\frac{n}{n+1}T\right)^2\right) = \left(\frac{n}{n+1}\right)^2 E(T^2) = \frac{(n+1)^2}{n^{\cancel{2}}} \frac{\cancel{n}}{n+2}\theta^2$

$$E(T^2) = \int_0^\theta x^2 f_T(x)\,dx = \int_0^\theta x^2 n \frac{x^{n-1}}{\theta^n}dx = \frac{n}{\theta^n}\int_0^\theta x^{n+1}dx = \frac{n}{\theta^n}\frac{x^{n+2}}{n+2}\Big|_0^\theta = \frac{n}{\theta^n}\frac{\theta^{n+2}}{n+2} = \frac{n}{n+2}\theta^2$$

$$E(T''^2) = \frac{(n+1)^2}{n(n+2)}\theta^2$$

$$Var(T'') = E(T''^2) - E(T'')^2 = \frac{(n+1)^2}{n(n+2)}\theta^2 - \theta^2 = \theta^2\left(\frac{(n+1)^2}{n(n+2)} - 1\right) = \frac{\theta^2}{n(n+2)} = MSE_\theta(T'')$$

$\left.\begin{array}{l} T' = 2\bar{X} \\ T'' = \frac{n+1}{n}\max X_i \end{array}\right\}$ non deviati $\qquad$ $E(T') = E(T'') = \theta$ $\qquad$ $MSE_\theta(T') = \frac{\theta^2}{3n}$

$MSE_\theta(T'') = \frac{\theta^2}{n(n+2)}$ $\longrightarrow$ questo MSE va a 0 più rapidamente per n grandi

ma

magari ci sono dei valori di n per cui T' più piccolo di T''

$$\underset{T''}{\frac{\cancel{\theta^2}}{n(n+2)}} \leq \underset{T'}{\frac{\cancel{\theta^2}}{3n}}$$

$\cancel{n}(n+2) \geq 3\cancel{n}$

$n \geq 1$ $\longrightarrow$ MSE di T'' sempre più basso di quello di T'

Es:

$X \sim G(\mu, \sigma)$ $\qquad$ X_1, X_2

$\theta = \sigma$ $\qquad$ $\tau(\theta) = \tau(\sigma) = \sigma^2$ $\qquad$ $T = X_1^2 - X_1 X_2$

prima spezzo il prodotto

$$E(T) = E(X_1^2 - X_1X_2) = E(X_1^2) - E(X_1X_2) = E(X_1^2) - E(X_1)\cdot E(X_2)$$

li considero come v.a X

$$= E(X^2) - E(X)\cdot E(X) = E(X^2) - E(X)^2 = \sigma^2$$

Es:

$X_1, \dots, X_n$ campione estratto da popolazione $X \sim F$ $\qquad$ $E(F) = \mu$, $Var(F) = \sigma^2$

non esplicito modello e parametro che non conosco

<u>Che posso fare?</u>

Considero lo stimatore $T_n = \frac{1}{3}\sum_{i=1}^n X_i$ è non deviato per μ? se $n = 3$ sicuramente non è deviato, è la media campionaria

ma

lo stimatore non è unico, ce n'è un altro per $n \neq 3$?

$$E(T_n) = E\left(\frac{1}{3}\sum_{i=1}^n X_i\right) = \frac{1}{3}\sum_{i=1}^n E(X_i) = \frac{n}{3}\mu = \mu \iff n = 3$$

è consistente in media quadratica?

$\downarrow$

non basta calcolare la Var perché lo stimatore è deviato

$MSE_\mu(T_n) = Var(T_n) + b_\mu(T_n)^2 = \frac{n}{9}\sigma^2 + \left(1 - \frac{n}{3}\right)^2\mu^2$ $\longrightarrow$ NON È CONSISTENTE PIÙ METTO ELEM NEL CAMPIONE PIÙ L'MSE AUMENTA QUADRATICAMENTE

$$Var(T_n) = Var\left(\frac{1}{3}\sum_{i=1}^n X_i\right) = \frac{1}{9}\sum_{i=1}^n \underbrace{Var(X_i)}_{\sigma^2} = \frac{n}{9}\sigma^2$$

$$b_\mu(T_n) = \mu - E(T_n) = \mu - \frac{n}{3}\mu = \left(1 - \frac{n}{3}\right)\mu$$

In merito all'unicità d. stimatore

Es:

$X_1,\dots,X_n \qquad X\sim G(\mu,\sigma) \qquad \tau(\theta)=E(X)=\mu \qquad \theta=\mu$

$T_1=\bar{X} \to$ non deviato $\qquad MSE_\mu(T_1)=Var(T_1)=\frac{\sigma^2}{n}$

$T_2=X_3 \quad E(T_2)=E(X_3)=E(X)=\mu \qquad MSE_\mu(T_2)=Var(T_2)=\sigma^2$

$T_3=\sum_{i=1}^{n}\lambda_i X_i \quad (\forall i\ \lambda_i>0,\ \sum_i \lambda_i=1) \qquad MSE_\mu(T_3)=Var(T_3)=Var\left(\sum_{i=1}^{n}\lambda_i X_i\right)=\sum_{i=1}^{n}Var(X_i\lambda_i)=\sum_i \lambda_i^2 \underbrace{Var(X_i)}_{\sigma^2}=\sigma^2\sum_i\lambda_i^2$

$E(T_3)=\sum_{i=1}^{n}\lambda_i \underbrace{E(X_i)}_{=\mu}=\mu\underbrace{\boxed{\sum_{i=1}^{n}\lambda_i}}_{=1}=\mu$

$x_1,x_2 \qquad T_1=\frac{1}{2}(X_1+X_2) \qquad T_3=\lambda_1X_1+(1-\lambda_1)X_2$

$\min_{\lambda_1}\ \lambda_1^2+(1-\lambda_1)^2 = 2\lambda_1-2(1-\lambda_1)$

$=2\lambda_1-2+2\lambda_1=0$

$2\lambda_1=1 \qquad \lambda_1=\frac{1}{2}=\lambda_2$

Stimatori general purpose

Media campionaria $\to$ é sempre uno stimatore non deviato per il valore atteso della popolazione ed é sempre consistente in media quadratica

Teo. centrale limite $\quad X_1,\dots,X_n$ i.i.d $\quad \sum_{i=1}^{n}X_i \,\dot\sim\, G(n\mu,\sqrt{n}\sigma)$
Tutte con
$E(X)=\mu$ e $Var(X)=\sigma^2$

cosa succede se queste sono le v.a che descrivono un campione?

$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \,\dot\sim\, G(\mu_{\bar{x}},\sigma_{\bar{x}})=G\left(\mu,\frac{\sigma}{\sqrt{n}}\right)$

$E(\bar{X})=\mu$

$Var(\bar{X})=\frac{\sigma^2}{n}$

se un campione ha un num. di elementi abbastanza elevato, la v.a che descrive la media campionaria che ottengo da quel campione é approx normale

Legge dei grandi numeri

FORTE $\quad P\left(\lim_{n\to+\infty}\bar{X}_n=\mu\right)=1$

più num. di elementi nel campione aumenta e tantomeno la media campionaria é una q.tà aleatoria, per $n\to+\infty$ diventa una costante, quella che voglio stimare

DEBOLE

$\lim_{n\to+\infty}P(|\bar{X}_n-\mu|>\varepsilon)=0 \qquad$ da un certo n in poi la prob di fare un certo errore diventa piccola

Altro stimatore general purpose

mi garantisce l'assenza di deviazione

Varianza campionaria : $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \longrightarrow$ non distorto per $Var(X)$

$X_1, \dots, X_n$ pop. X $\quad \tau(\theta) = Var(X)$

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n}(x_i^2 - 2\bar{x}x_i + \bar{x}^2) = \sum_{i=1}^{n}x_i^2 - 2\bar{x}\underbrace{\sum_{i=1}^{n}x_i}_{n\bar{x}} + n\bar{x}^2 = \sum_{i=1}^{n}x_i^2 - n\bar{x}^2$$

$E(x^2) = Var(x) + E(x)^2$

$E(\bar{x}) = \mu \qquad Var(\bar{x}) = \frac{\sigma^2}{n}$

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

$$E((n-1)S^2) = E\left(\sum_{i=1}^{n}x_i^2 - n\bar{x}^2\right) = \sum_{i=1}^{n}E(x_i^2) - nE(\bar{x}^2)$$

$$= \sum_{i=1}^{n}E(x^2) - nE(\bar{x}^2)$$

$$= n\left(E(x^2) - E(\bar{x}^2)\right)$$

$$= n\left(Var(x) + E(x)^2 - Var(\bar{x}) - E(\bar{x})^2\right)$$

$$= n\left(\sigma^2 + \cancel{\mu^2} - \frac{\sigma^2}{n} - \cancel{\mu^2}\right) = n\sigma^2 - \sigma^2 = (n-1)\sigma^2$$

$$\cancel{(n-1)}E(S^2) = \cancel{(n-1)}\sigma^2 \qquad E(S^2) = \sigma^2$$

A volte abbiamo una distribuz. e non possiamo trovare stimatori custom semplicemente

$X \sim E(\lambda) \quad \theta = \lambda \quad \tau(\theta) = \tau(\lambda) = \lambda$

$E(\bar{x}) = \frac{1}{\lambda} \quad \lambda = \frac{1}{E(\bar{x})} \quad \frac{1}{\bar{x}} \quad T = \frac{1}{\bar{x}} = \frac{n}{\sum_i x_i} \quad T = \frac{1}{\bar{x}} \approx \lambda$

$E(T) = E\left(\frac{1}{\bar{x}}\right)$ non possiamo procedere

Come Teo. centrale del limite garantisce che la media campionaria ha una distribuzione approx. normale

Es:

distanza Terra-Stella

$X_1, \dots, X_n \qquad X \sim G(d, 2) \qquad \theta = d \quad \tau(\theta) = d \qquad$ stimatore : $\bar{X}_n$

Voglio fare un ragionamento finito non più asintotico, quanto deve essere grande n perché l'errore che io faccio sulla stima sia piccolo con prob. grande

"Avere almeno 95% di prob. che la stima abbia un errore di al più 0.5 anni luce"

$P(|\bar{x}_n - d| \leq 0.5) \geq 0.95$

$E(\bar{x}_n) = d \qquad Var(\bar{x}_n) = \frac{4}{n} \qquad\qquad Z \sim G(0,1)$

Standardizziamo $\quad P(|\bar{x}_n - d| \leq 0.5) = P\left(\frac{|\bar{x}_n - d|}{2/\sqrt{n}} < \frac{0.5}{2/\sqrt{n}}\right) = P\left(\left|\frac{\bar{x}_n - d}{2/\sqrt{n}}\right| \leq \frac{\sqrt{n}}{4}\right) \approx P\left(|Z| \leq \frac{\sqrt{n}}{4}\right) = P\left(-\frac{\sqrt{n}}{4} \leq Z \leq \frac{\sqrt{n}}{4}\right)$

sto applicando Teo centrale limite

$$= \Phi\left(\frac{\sqrt{n}}{4}\right) - \underbrace{\Phi\left(-\frac{\sqrt{n}}{4}\right)}_{1-\Phi\left(\frac{\sqrt{n}}{4}\right)}$$

$2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1 \geq 0.95$ Trovo valore di n per cui sia vera

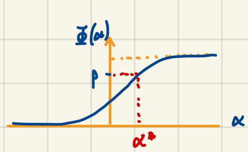
$$= \Phi\left(\frac{\sqrt{n}}{4}\right) - 1 + \Phi\left(\frac{\sqrt{n}}{4}\right) = \underbrace{2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1}_{\text{voglio che sia } \geq 0.95}$$

$\Phi\left(\frac{\sqrt{n}}{4}\right) \geq \frac{1.95}{2} = 0.975 \qquad \Phi^{-1}\left(\Phi\left(\frac{\sqrt{n}}{4}\right)\right) \geq \Phi^{-1}(0.975) \qquad \frac{\sqrt{n}}{4} \geq 1.96 \qquad n \geq (4 \cdot 1.96)^2 \simeq 61.4$

per $n \geq 62$

$P(|\bar{x}_n - d| \leq 0.5) \geq 0.95$

$\Phi(\alpha) \geq p$



$\forall \alpha \geq \alpha^* \quad \Phi(\alpha) \geq p \qquad \alpha^* = \Phi^{-1}(p)$

Rivediamolo in generale

$$P(|\bar{X}_n - \mu| \leq r) \geq 1-\delta$$

APPLICO TCL

$$P\left(\frac{|\bar{X}_n-\mu|}{\sigma/\sqrt{n}} \leq \frac{r}{\sigma/\sqrt{n}}\right) = P\left(\left|\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}\right| \leq \frac{r}{\sigma}\sqrt{n}\right) \approx P\left(|Z| \leq \frac{r}{\sigma}\sqrt{n}\right) = P\left(-\frac{r}{\sigma}\sqrt{n} \leq Z \leq \frac{r}{\sigma}\sqrt{n}\right)$$

$$= \Phi\left(\frac{r}{\sigma}\sqrt{n}\right) - \underbrace{\Phi\left(-\frac{r}{\sigma}\sqrt{n}\right)}_{1-\Phi\left(\frac{r}{\sigma}\sqrt{n}\right)}$$

$$= \Phi\left(\frac{r}{\sigma}\sqrt{n}\right) - \left(1-\Phi\left(\frac{r}{\sigma}\sqrt{n}\right)\right) = 2\Phi\left(\frac{r}{\sigma}\sqrt{n}\right) - 1$$

impongo questo

$$2\Phi\left(\frac{r}{\sigma}\sqrt{n}\right) - 1 \geq 1-\delta$$

$$\Phi\left(\frac{r}{\sigma}\sqrt{n}\right) \geq 1-\frac{\delta}{2} \qquad \frac{r}{\sigma}\sqrt{n} \geq \Phi^{-1}\left(1-\frac{\delta}{2}\right) \rightarrow r = \frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(1-\frac{\delta}{2}\right)$$

$$\frac{\delta}{2} \geq 1-\Phi\left(\frac{r}{\sigma}\sqrt{n}\right) \qquad n \geq \left(\frac{\sigma}{r}\Phi^{-1}\left(1-\frac{\delta}{2}\right)\right)^2$$

Problema, chi mi dice che conosco σ? A meno che venga data la stimiamo come $\sigma \approx \sqrt{s^2}$

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

$$\delta \geq 2\left(1-\Phi\left(\frac{r}{\sigma}\sqrt{n}\right)\right)$$

$$P(|X-\mu| \geq r) \leq \frac{\sigma^2}{r^2} \longrightarrow \text{anche da qui si possono fare i calcoli}$$

TCHEBYSHEV

$$P(|X-\mu| < r) = 1 - P(|X-\mu| \geq r) \geq 1-\frac{\sigma^2}{r^2}$$

impongo

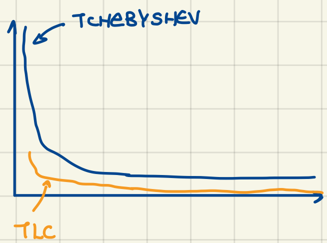
$$P(|\bar{X}-\mu| < r) \geq 1-\frac{\sigma^2}{nr^2} \geq 1-\delta \qquad \frac{\sigma^2}{nr^2} \leq \delta \qquad n \geq \frac{\sigma^2}{\delta r^2}$$

se vero questo $\rightarrow P(|\bar{X}-\mu| < r) \geq 1-\delta$

$$\delta \geq \frac{\sigma^2}{nr^2} \qquad r \geq \sqrt{\frac{\sigma^2}{n\delta}}$$

Confrontiamo Taglia minimale ottenuta con TLC e quella ottenuta con TCHEBYSHEV

$$\frac{\cancel{\sigma^2}}{\cancel{r^2}}\Phi^{-1}\left(1-\frac{\delta}{2}\right)^2 \leq \frac{\cancel{\sigma^2}}{\delta \cancel{r^2}} \qquad \Phi^{-1}\left(1-\frac{\delta}{2}\right)^2 \leq \frac{1}{\delta}$$



$\hat{\sigma} = \sqrt{s^2}$ $\qquad$ s^2 non deviato per σ^2

$\sqrt{s^2}$ è non deviato per σ?

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

stimatore non distorto per la varianza della popolazione σ^2 $\quad E(s^2) = \sigma^2$

?

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

è non deviato per σ? $E(s) = \sigma$?

Se avessimo applicato un'operaz. lineare si ma la radice quadrata non è lineare

$\downarrow$

spoiler è deviato

Consideriamo degli eventi che possono accadere nel tempo

Fissiamo un istante iniziale $t = 0$ $\quad N(t) = \#$ eventi che occorrono tra 0 e t $\;[0,t)$

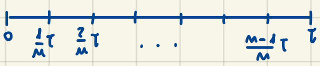
- Per t fissato $N(t)$ è una var. aleatoria

- Se voglio ragionare a valori diversi di t ho una famiglia di var. aleatorie
$\downarrow$
PROCESSI STOCASTICI

Processo di Poisson $\longrightarrow$ se si verificano certe ipotesi

1) $N(0) = 0$
2) indipendenza per intervalli disgiunti
3) $\lim_{h \to 0} \dfrac{P(N(h) = 1)}{h} = \lambda$ $\quad$ (se h "piccolo" $\longrightarrow P(N(h) = 1) \approx \lambda h$)
4) $\lim_{h \to 0} \dfrac{P(N(h) \geq 2)}{h} = 0$ $\quad$ (se h "piccolo" $\longrightarrow P(N(h) \geq 2) \approx 0$)

Allora $N(t) \sim \mathcal{P}(\lambda t)$



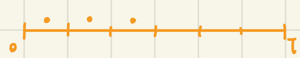
$N(t) = k \in \mathbb{N}$

$m \in \mathbb{N} \qquad m > k$

A: ogni avvenimento cade in un intervallo diverso
B: tutti gli altri modi per cui $N(t) = k$

B:


$A \cup B = \{N(t) = k\}$, $A \cap B = \emptyset$

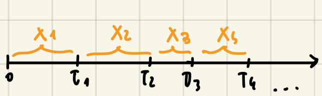


$P(N(t) = k) = P(A) + P(\cancel{B}) = \binom{m}{k}\left(\lambda \frac{t}{m}\right)^k \left(1 - \lambda \frac{t}{m}\right)^{m-k}$

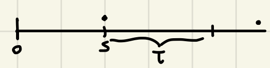
$\binom{m}{k}\left(\lambda \frac{t}{m}\right)^k \left(1 - \lambda \frac{t}{m}\right)^{m-k}$ $\qquad \lambda \frac{t}{m} =: p \qquad B(m,p) \qquad mp = \lambda t$ $\quad$ Approssimazione della binomiale usando il modello di Poisson

$N(t) \sim \mathcal{P}(\lambda t)$

$P(N(t) = k) = e^{-\lambda t} \dfrac{(\lambda t)^k}{k!} I_{\mathbb{N} \cup \{0\}}(k)$



$P(X_1 > t) = P(N(t) = 0) = e^{-\lambda t} \dfrac{(\lambda t)^0}{0!} = e^{-\lambda t}$



$F_{X_1}(t) = P(X_1 \leq t) = 1 - P(X_1 > t) = 1 - e^{-\lambda t} \longrightarrow X_1 \sim \mathcal{E}(\lambda)$

$P(X_2 > t \mid X_1 = s) = P(\text{nessun evento in } [s, s+t) \mid X_1 = s) = P(\text{nessun evento tra } [s, s+t)) = P(N(t) = 0) = e^{-\lambda t} \longrightarrow X_2 \sim \mathcal{E}(\lambda)$

$X_2 \perp X_1$

$\forall i \quad X_i \sim \mathcal{E}(\lambda)$ i.i.d

Stimatore

$X \sim G(p)$ $\quad E(\overline{x}) = E(x) = \frac{1-p}{p}$ $\quad p\,E(\overline{x}) = 1-p$ $\quad p(1+E(\overline{x})) = 1$ $\quad p = \frac{1}{1+E(\overline{x})}$

$T = \frac{1}{1+\overline{x}}$