PROBABILITÀ

qualcosa che non ha un risultato certo → esiti diversi in maniera casuale ↓ $e_1, e_2, \dots, e_n$

SPAZIO CAMPIONE (S o Ω) = insieme di tutti gli esiti di un esperimento → $\{e_1, e_2, \dots, e_N\}$

↳ **EVENTO** → sottoinsieme dello spazio campione → contiene numero di esiti dell'esperimento sottoinsieme di tutti gli esiti possibili

ESEMPIO: lancio di un dado cubico → $e_1, \dots e_6$ esiti se e_k = uscita di k puntini

$S = \{e_1, e_2, e_3, e_4, e_5, e_6\}$ SPAZIO CAMPIONE

evento A : 'uscita di un numero multiplo di 3' → $A = \{e_3, e_6\}$

DEFINIZIONE "CLASSICA" DI PROBABILITÀ → dato un esperimento ed un evento E associato ad esso

$$P(E) = \frac{\text{n° esiti contenuti in } E}{\text{n° esiti possibili (spazio campione)}} = \frac{\text{n° esiti favorevoli}}{\text{n° esiti totali}}$$

→ esiti FINITI EQUIPROBABILI che si possono contare

DEFINIZIONE FREQUENTISTA DI PROBABILITÀ → dato un esperimento ed un evento E associato ad esso, si ripeta l'esperimento in maniera identica e indipendente un numero elevato di volte (N) allora

$$P(E) = \frac{\text{n° di esperimenti in cui si osserva } E}{\text{n° totale di esperimenti eseguiti}}$$

↓ imprecisa se non si fanno infinite prove

↳ esiti finiti ma non equiprobabili

Teorema 2.0.1 (Principio fondamentale del calcolo combinatorio (o di enumerazione))

> *Dati 2 esperimenti per cui il primo ha* **m** *esiti possibili e il secondo* **n** *esiti possibili, si ha che le* ***possibili sequenze ordinate sono***
>
> $$m \cdot n$$
>
> *Generalizzando a N esperimenti, con $n_1, n_2, n_3, \dots, n_N$ esiti la sequenza può variare in $n_1 \cdot n_2 \cdot n_3 \cdot \dots \cdot n_N$ modi possibili.*

→ esempio: targhe auto
ogni targa ha 2 lettere, 3 numeri e altre 2 lettere
quindi esistono
$26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 \cdot 26 \cdot 26$
combinazioni possibili

Disposizioni semplici	Dati **n** oggetti distinti, si dicono disposizioni semplici di n elementi di classe **k** ($k \in N \mid 1 \leq k < n$) gli allineamenti di k elementi diversi presi tra gli oggetti dati	$D_{n,k} = n \cdot (n-1) ... (n-k-1) = \frac{n!}{(n-k)!}$	allineamento di 3 vocali tra le 5 → $D_{5,3} = \frac{5!}{2!} = 5 \cdot 4 \cdot 3 = 60$
Disposizioni con ripetizione	Dati n oggetti distinti, si dicono disposizioni con ripetizione di n elementi di classe k ($k \in N \mid k \geq 1$) gli allineamenti di K elementi non necessariamente diversi che si possono formare a partire dagli n oggetti	$D^R_{n,k} = n \cdot n ... n = n^k$	→ allineamento di 3 vocali con rip. $D^R_{5,3} = 5^3$
Permutazioni semplici	Dati n oggetti distinti, si dicono permutazioni semplici di n elementi tutti gli allineamenti che si possono formare con gli n oggetti	$P_n = D_{n,n} = n \cdot (n-1)(n-2) ... 1 = n!$ ↳ sono disp semplici se k=n	→ allineamenti delle 5 vocali $P_5 = 5! = 120$
Permutazioni con ripetizione	Dati n oggetti non necessariamente distinti, si dicono permutazioni con ripetizione tutti gli allineamenti formati dagli n oggetti dati	$k_1 + k_2 + k_p = n$ $P^R_n = \frac{n!}{k_1!k_2!...k_p!}$ → si usano x gli anagrammi	→ anagrammi di "casa" → $P^R_{\underset{tot}{4} = 1 + \underset{A}{1} + \underset{CeS}{2}} = \frac{4!}{2!} = 12$
Combinazioni semplici	Dati n oggetti distinti, si dicono combinazioni semplici di n elementi di classe k ($k \in N \mid 1 \leq k \leq n$) gli insiemi di k elementi presi tra gli n oggetti dati (senza badare all'ordine)	$C_{n,k} = D_{n,k} \cdot P_k = \frac{n!}{(n-k)!} \cdot \frac{1}{k!} = \frac{n!}{(n-k)!k!} = \binom{n}{k}$	gruppi di 3 vocali (no ordine o rip) → $C_{5,3} = \binom{5}{3} = \frac{5!}{2!\,3!} = \frac{5 \cdot 4 \cdot \cancel{3!}}{2 \cdot \cancel{3!}} = 10$
Combinazioni con ripetizione	Dati n oggetti distinti, si dicno combinazioni con ripetizione di n elementi di classe k ($k \in N \mid \{0\}$) tutti gli insiemi di K elementi anche ripetuti che si possono formare a partire dagli n oggetti dati	$C^R_{n,k} = \frac{(n+k-1)!}{(n-1)!k!} = \binom{n+k-1}{k}$	gruppi di 3 vocali anche ripetute → $C^R_{5,3} = \frac{(5+3-1)!}{(5-1)!\,3!} = \frac{7!}{4!\,3!} = 35$

PARADOSSO DEI COMPLEANNI → in una stanza sono presenti n persone tutte nate in un anno non bisestile, qual è la probabilità che abbiano date di compleanno diverse se $n \leq 365$?

→ A: 'tutti hanno date di compleanno diverse' : $P(A) = \dfrac{n° \text{ casi fav.}}{n° \text{ casi tot.}}$

↓

i casi totali sono disposizioni con ripetizione

↓

$D^R_{365,n} = 365^n$

per i casi favorevoli usiamo le disposizioni semplici

$D_{365,n} = \dfrac{365!}{(365-n)!} = 365 \cdot \ldots \cdot (365-n+1)$

↓

paradosso perché è una probabilità che decresce all'aumentare di n

↓

come calcolare la probabilità che B: 'non tutti hanno date di compleanno diverse'? → $P(B) = 1 - P(A)$

↓

DEFINIZIONE ASSIOMATICA DI PROBABILITÀ → **ASSIOMI DI KOLMOGOROV**

↓

dato un esperimento che prevede più esiti a cui è associato lo spazio campione S e dato un evento $E \subset S$ si definisce $P(E)$ un numero tale che:

A1) $0 \leq P(E) \leq 1$

A2) $P(S) = 1$

A3) dati $E_1, E_2, \ldots E_n \subset S$ mutualmente esclusivi (disgiunti) ovvero tali che $E_i \cap E_k = \emptyset$ se $i \neq k$

$$P\left(\bigcup_{k=1}^{n} E_k\right) = \sum_{k=1}^{n} P(E_k)$$

→ la probabilità dell'unione di eventi disgiunti è la somma delle prob. degli eventi

↓

PROPRIETÀ

↓

P1) Dato $E \subset S$, $P(E^c) = 1 - P(E)$ ⟶

DIMOSTRAZIONE

$E \cup E^c = S$
$E \cap E^c = \emptyset$
(disgiunti)

$1 = P(S) = P(E \cup E^c) = P(E) + P(E^c)$

$1 = P(E) + P(E^c) \longrightarrow P(E^c) = 1 - P(E)$

P1 bis) $P(\emptyset) = 0$ → 0 può significare vuoto o impossibile, o numero di casi infinito/non numerabile → **DIMOSTRAZIONE** → $A_2 \to P(S) = 1$ $S^c = \emptyset \Rightarrow P(S^c) = \emptyset$

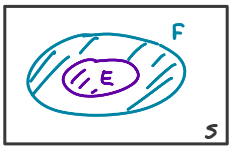
$P(\emptyset) \underset{P_1}{=} 1 - P(S) \underset{A_2}{=} 1 - 1 = 0$

P2) Dati $E, F \subset S$ se $E \subset F$

$P(E) \leq P(F)$ → E è strettamente contenuto in F, quindi non possono coincidere, ma c'è il $\leq$ per quando può succedere che ci siano esiti infiniti e E ed F differiscono di un esito (probabilità 0)

↓

DIMOSTRAZIONE



$F = \left\{ \begin{matrix} \text{esiti di } F \\ \in E \end{matrix} \right\} \cup \left\{ \begin{matrix} \text{esiti di } F \\ \notin E \end{matrix} \right\}$

$F = E \cup (F \cap E^c)$ → disgiunti $E \cap F.. = \emptyset$

$P(F) = P(E \cup (F \cap E^c)) \underset{A3}{=} P(E) + P(F \cap E^c)$

tramite $A_1 \to P(F \cap E^c) \geq 0$

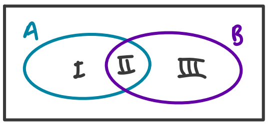
$P(E) + \underbrace{P(F \cap E^c)}_{\geq 0} \geq P(E) \to P(F) \geq P(E) \to P(E) \leq P(F)$

P3) Dati $A, B \subset S$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ → A e B possono non essere disgiunti

↓

DIMOSTRAZIONE



$\text{II} = A \cap B$
$\text{I} = A \setminus (A \cap B) = A \cap B^c$
$\text{III} = B \cap A^c$
I, II, III disgiunti $(\cap = \emptyset)$

$A = \text{I} \cup \text{II}$, $B = \text{II} \cup \text{III}$, $A \cup B = \text{I} \cup \text{II} \cup \text{III}$

$P(A) \underset{A3}{=} P(\text{I}) + P(\text{II})$, $P(B) = P(\text{II}) + P(\text{III})$

$P(A \cup B) = P(\text{I} \cup \text{II} \cup \text{III}) = P(\text{I}) + P(\text{II}) + P(\text{III})$

$P(\text{I}) = P(A) - P(\text{II})$
$P(\text{III}) = P(B) - P(\text{II})$ → $P(A \cup B) = P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B) = P(A) + P(B) - P(A \cap B)$
$P(\text{II}) = P(A \cap B)$

P4) Dati $A, B, C \subset S$

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

PROBABILITÀ CONDIZIONATA → Dati $A, B \subset S$ con $P(B) \neq 0$ si dice probabilità di A condizionata da B

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

→ vuol dire: abbiamo un esperimento, so che si è verificato B, come cambia la probabilità dell'evento A?

↓

esempio: lancio di un dado cubico

B = 'uscita num dispari' = $\{e_1, e_3, e_5\}$

A = 'uscita di 2' = $\{e_2\}$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\emptyset)}{1/2} = 0$$

C = 'uscita multipli di 3' = $\{e_3, e_6\}$

$$P(C|B) = \frac{P(C \cap B)}{P(B)} = \frac{P(e_3)}{1/2} = \frac{1/6}{1/2} = \frac{1}{3}$$

↓

gli eventi che non si condizionano vengono chiamati **eventi indipendenti** → $P(A|B) = P(A)$ ovvero $P(B|A) = P(B)$

DEFINIZIONE CLASSICA DI PROB. RICAVATA IN MANIERA RIGOROSA

↓

Dato un esperimento con esiti in numero finito equiprobabili

$S = \{e_1, e_2, \dots, e_m\}$ m = n° esiti $P(e_1) = P(e_2) = \dots = P(e_m) = p$

↓

gli esiti per definizione sono disgiunti, perché non si possono verificare allo stesso tempo ($e_i \cap e_k = \emptyset$ se $i \neq k$)

→ $S = e_1 \cup e_2 \dots \cup e_m = \bigcup_{j=1}^{m} e_j$

$$1 \underset{A2}{=} P(S) = P\left(\bigcup_{j=1}^{m} e_j\right) \underset{A3}{=} \sum_{j=1}^{m} P(e_j)$$

prob. di ogni esito

$$\sum_{j=1}^{m} P(e_j) = \sum_{j=1}^{m} p = mp \rightarrow mp = 1 \rightarrow p = \frac{1}{m}$$

↓

verifichiamo ora la definizione classica di probabilità

Sia $E \subset S$, $E = \{e_1, \dots e_k\}$ $1 \le k \le m$

consideriamo i primi k esiti → $E = \bigcup_{j=1}^{k} e_j$

$$P(E) = P\left(\bigcup_{j=1}^{k} e_j\right) \underset{A3}{=} \sum_{j=1}^{k} P(e_j) = \sum_{j=1}^{k} \frac{1}{m} = \frac{k}{m}$$

$P(E) = \dfrac{k}{m}$ → k: n° esiti in E (favorevoli); m: n° esiti totali!

→ la probabilità di un evento contenuto in uno spazio campione di esiti equiprobabili è il rapporto tra gli esiti dell'evento e gli esiti del campione

EVENTI DISGIUNTI, INDIPENDENTI → $E, F \subset S$ e $E \cap F = \emptyset$ → non hanno esiti in comune → E ed F non possono verificarsi simultaneamente

↳ indipendenti e disgiunti non sono la stessa cosa!

↳ **DEFINIZIONE RIGOROSA** → $P(E \cap F) = P(E)P(F)$ → gli eventi incompatibili possono essere indipendenti?

↓

$E, F \subset S$ e incompatibili $E \cap F = \emptyset$

se sono indipendenti $P(E \cap F) = P(E)P(F)$ ↗ $P(E \cap F) = 0$

↓

si, se almeno 1 dei due eventi ha prob. 0

in generale, è impossibile che due eventi incompatibili siano indipendenti, perché l'incompatibilità condiziona fortemente

↓

TEOREMA

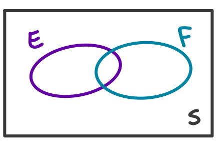
↓

Dati $E, F \subset S$ indipendenti, allora anche E ed F^c, F e E^c, E^c ed F^c sono indipendenti

↳ nel caso di tre indipendenze o +, si ragiona a coppie di eventi

↓

DIMOSTRAZIONE



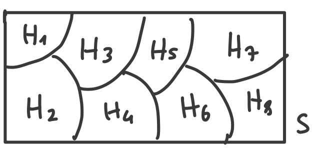
$E = (E \cap F) \cup (E \cap F^c)$ con $(E \cap F) \cap (E \cap F^c) = \emptyset$

$$P(E) = P((E \cap F) \cup (E \cap F^c)) \underset{A3}{=} P(E \cap F) + P(E \cap F^c)$$

$$P(E \cap F^c) = P(E) - P(E \cap F) \underset{ind.}{=} P(E) - P(E)P(F) = P(E)(1 - P(F)) \underset{P_1}{=} P(E)P(F^c)$$

PARTIZIONE DI UNO SPAZIO CAMPIONE → Dato uno spazio campione S si dice partizione di S una suddivisione in eventi $\{H_1, H_2, \dots H_n\}$ tali che $H_i \cap H_j = \emptyset$ se $i \neq j$ e $\bigcup_{j=1}^{n} H_j = S$

↳ si chiamano visualmente ipotesi

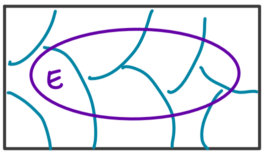


TEOREMA DELLE PROBABILITÁ TOTALI → Data una partizione di S $\{H_1, H_2, \dots, H_n\}$ e un evento $E \subset S$

$$P(E) = \sum_{j=1}^{n} P(E \mid H_j)\, P(H_j)$$

↓

DIMOSTRAZIONE



→ $E = (E \cap H_1) \cup (E \cap H_2) \dots \cup (E \cap H_n)$ → possono esserci insiemi vuoti, ma é corretto

$(E \cap H_i) \cap (E \cap H_j) = 0$ se $i \neq j$ perché $H_i \cap H_j = \emptyset$ → nemmeno le inters. hanno elementi in comune

$$P(E) = P\left(\bigcup_{j=1}^{n} (E \cap H_j)\right) \underset{A3}{=} \sum_{j=1}^{n} P(E \cap H_j) = \sum_{j=1}^{n} P(E \mid H_j)\, P(H_j)$$

esercizio 1 : esame del sangue per individuare una malattia M è efficace al 99% con un 1% di falsi positivi, inoltre ha incidenza nella popolazione dello 0.5%. Qual é la probabilitá che un individuo a caso sia positivo?

H_1 = presentare malattia M $\quad P(H_1) = 0.005 \quad P(E \mid H_1) = 0.99$

$H_2 = H_1^c \quad P(H_2) = 0.995 \quad E$ = 'esame del sangue positivo' $\quad P(E \mid H_2) = 0.01$

$$P(E) = P(E \mid H_1)\, P(H_1) + P(E \mid H_2)\, P(H_2) = 0.99 \times 0.005 + 0.01 \times 0.995 \simeq 1.5\%$$

TEOREMA DI BAYES → Dato uno spazio campione S ed una sua partizione $\{H_1, H_2, \dots H_n\}$ e dato $E \subset S$ $(P(E) \neq 0)$

$$P(H_i \mid E) = \frac{P(E \mid H_i)\, P(H_i)}{\sum_{k=1}^{n} P(E \mid H_k)\, P(H_k)} \quad i = 1, 2, \dots n$$

→ mi domando quale sia la probabilitá di un ipotesi se si é verificato E

↓

DIMOSTRAZIONE

↓

$$P(H_i \cap E) = P(E \mid H_i)\, P(H_i) = P(H_i \mid E)\, P(E)$$

$$P(H_i \mid E) = \frac{P(E \mid H_i)\, P(H_i)}{P(E)} \underset{\substack{\downarrow \\ \text{form. prob. tot.}}}{=} \frac{P(E \mid H_i)\, P(H_i)}{\sum_{k=1}^{n} P(E \mid H_k)\, P(H_k)}$$

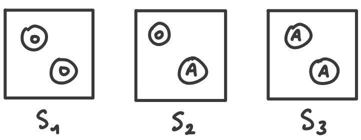
→ **esercizio 1** (esame del sangue), se un individuo risulta positivo, qual é la probabilitá che sia malato? → $P(H_1 \mid E)$

$$P(H_1 \mid E) = \frac{P(E \mid H_1)\, P(H_1)}{\sum_{k=1}^{2} P(E \mid H_k)\, P(H_k)} = \frac{P(E \mid H_1)\, P(H_1)}{P(E)} \simeq 0.33$$

PARADOSSO DELLE SCATOLE DI BERTRAND

↓

3 scatole apparentemente identiche → H_i = 'scelta della scatola S_i' $\quad i = 1, 2, 3 \quad P(H_i) = \frac{1}{3}$



O = 'estrazione di una moneta d'oro'

$$P(O) = P(O \mid H_1)P(H_1) + P(O \mid H_2)P(H_2) + P(O \mid H_3)P(H_3)$$

$$P(O) = 1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} = \frac{1}{3} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$

$P(O \mid H_1) = 1$

$P(O \mid H_2) = \frac{1}{2}$

$P(O \mid H_3) = 0$

se é stata estratta una moneta d'oro, qual é la probabilitá che sia della prima scatola?

$$P(H_1 \mid O) \underset{\substack{\downarrow \\ \text{Bayes}}}{=} \frac{P(O \mid H_1)\, P(H_1)}{P(O)} = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

PROBLEMA DELLA ROVINA DEL GIOCATORE

↓ ↳ LEZIONE 4

A e B giocano lanciando una moneta, se T (esce testa) B da ad A una moneta, se C (esce croce) A da a B una moneta. A possiede k monete $(0 < k < n)$ e B possiede $(n-k)$ monete. Vince il giocatore che ottiene tutte le monete. Prob. che A vinca?

↓

applichiamo la formula delle prob. totali al primo lancio → A = 'A vince' $\quad T$: esce testa → $C = T^c$ → $P(T) = p$ → $P(C) = 1 - p = q$ → $P(A) = P_k$

↓

servirá almeno un lancio di moneta x la sconfitta o vittoria → $P(A) = P(A \mid T)\, P(T) + P(A \mid C)\, P(C)$

$$p_k = p_{k+1} \cdot p + p_{k-1} \cdot q$$

→ formula ricorsiva: a seconda del valore di k, si lega il valore della probabilitá di vittoria

$p_k(\underbrace{p+q}_{1})$

$$(p+q)\, p_k = p_{k+1}\, p + p_{k-1}\, q$$

$$(p_{k+1} - p_k)\, p = (p_k - p_{k-1})\, q$$

$$p_{k+1} - p_k = (p_k - p_{k-1}) \frac{q}{p} \quad \text{se } p \neq 0$$

$k = 1 \quad p_2 - p_1 = (p_1 - p_0) \frac{q}{p}$

$k = 2 \quad p_3 - p_2 = (p_2 - p_1) \frac{q}{p} = p_1 \left(\frac{q}{p}\right)^2$

$\vdots$

$k = n-1 \quad p_n - p_{n-1} = (p_{n-1} - p_{n-2}) \frac{q}{p} = p_1 \left(\frac{q}{p}\right)^{n-1}$

↓

sommiamo tutte le relazioni $\quad p_n - p_1 = \sum_{k=1}^{n-1} p_1 \left(\frac{q}{p}\right)^k$

N.B. $p_n = 1$

$$1 = p_1 + \sum_{k=1}^{n-1} p_1 \left(\frac{q}{k}\right)^k \rightarrow 1 = p_1 \cdot \sum_{k=0}^{n-1} \left(\frac{q}{p}\right)^k$$

$$p_1 = \frac{1}{\sum_{k=0}^{n-1} \left(\frac{q}{p}\right)^k}$$

→ probabilitá che A vinca se ha una sola moneta intersezione

DISPOSITIVI IN SERIE → $D_1, D_2 \dots D_n$ n dispositivi che formano un sistema in serie. Il dispositivo k-esimo $(k = 1, \dots n)$ funziona con probabilità p_k indipendentemente dagli altri. Qual è la probabilità che il sistema funzioni?

↓

il sistema funziona solo se tutti gli n dispositivi funzionano

D_k = 'il dispositivo k-esimo funziona' → $P(D_k) = p_k$ T = 'il sistema funziona'

$$P(T) = P(D_1 \cap D_2 \dots \cap D_n) = P(D_1)P(D_2)\dots P(D_n) = p_1 p_2 \cdots p_n = \prod_{k=1}^{n} p_k$$

DISPOSITIVI IN PARALLELO → V = 'il sistema in parallelo funziona' → unione → $V = D_1 \cup D_2 \dots \cup D_n$

↓

il sistema funziona se almeno uno dei dispositivi funziona

anziché fare $P(V) = P(D_1) + P(D_2) + \dots + P(D_n) - P(D_1 \cap D_3)\, P(D_1 \cap D_3) \dots$ → usiamo il complementare

$$P(V) = 1 - P(V^c) = 1 - P((D_1 \cup \dots D_n)^c) = 1 - P(D_1^c \cap \dots D_n^c) = 1 - P(D_1)^c \dots P(D_n)^c =$$

$$= 1 - (1-p_1)\dots(1-p_n) = 1 - \prod_{k=1}^{n}(1-p_k)$$

il valore prima dell'esperimento non è noto

VARIABILI CASUALI (ALEATORIE) → Dato un esperimento che presenta esiti diversi in maniera casuale è possibile associare all'esperimento una variabile casuale i cui valori corrispondono ai diversi esiti dell'esperimento

↓

v.c. valori reali $X \in R$

è importante che siano valori diversi per esiti diversi → una v.c. è discreta quando presenta un numero finito e numerabile di valori

↓

non tutte le v.c. sono discrete, si possono associare ad esse le probabilità del verificarsi degli esiti

es. lancio moneta equilibrata

$X \in \{0, 1\}$ (T C) $P(X=0) = 1/2$ $P(X=1) = 1/2$ ← $X \in \{x_1, x_2 \dots x_n\}$ (e_1, e_2, e_n → esiti) → $P(X = x_k) = P(e_k)$

↓

la funzione che associa i valori della variabile alla probabilità si chiama

FUNZIONE DI MASSA DI PROBABILITÀ

↓

$p: R \to [0,1]$ $p(a) = P(X = a)$ $a \in R$ → è una funzione che dà 0 dappertutto, tranne nei punti dei valori associati all'esp. → $p(a) = \begin{cases} \frac{1}{2} & \text{se } a = 0 \text{ o } a = 1 \\ 0 & \text{se } a \neq 0 \text{ o } a \neq 1 \end{cases}$ (moneta)

↓

proprietà

1) $0 \le p(a) \le 1 \quad \forall a \in R$

2) se $X \in \{x_1, x_2, \dots x_n\}$ → $1 = P(S) = \sum_{k=1}^{n} p(x_k)$ → se faccio la somma della funzione di massa di tutti i valori della variabile casuale, otterrò la probabilità dello spazio campione

FUNZIONE DI DISTRIBUZIONE DI PROBABILITÀ → CASO DISCRETO

↓

X v.c. discreta $X \in \{x_1, x_2, \dots x_n\}$

$F: R \to [0,1]$

$$F(a) \equiv P(X \le a) \quad \forall a \in R$$

↓

la F può soltanto rimanere uguale a se stessa in un tratto dell'asse reale o crescere al crescere del suo argomento, non può decrescere

→ PROPRIETÀ

1) $0 \le F(a) \le 1 \quad \forall a \in R$

2) $F(-\infty) = P(X \le -\infty) = 0$

3) $F(+\infty) = P(X \le +\infty) = 1$

4) se $a < b \Rightarrow F(a) \le F(b)$

5) per v.c. discrete F è una funzione costante a tratti (gradini) → ci aspettiamo delle discontinuità in corrispondenza dei valori della variabile

se la v.c. è minore di a, a maggior ragione è minore di b

evento '$X < a$' $\subset$ '$X < b$' →

↓

se $A \subset B$ $P(A) \le P(B)$

$P(X < a) \le P(X \le b)$ → $F(a) \le F(b)$

VARIABILE CASUALE CONTINUA → si associano ad esperimenti i cui esiti non possono più essere rappresentati da numeri finiti, ma da un intervallo dell'asse reale o tutto l'asse reale

↓

Una v.c. X si dice continua se è associata a X una funzione (densità di probabilità) $f: \mathbb{R} \to \mathbb{R}^+ \cup \{0\}$ tale che $\forall \beta \in \mathbb{R}$

$$P(X \in \beta) = \int_\beta f(s)\,ds$$

ci serve x calcolare la prob. che la v.c. sia in un sottoinsieme qualunque di R

→ PROPRIETÀ

1) $f: \mathbb{R} \to \mathbb{R}^+ \cup \{0\}$ ovvero $f(x) \ge 0 \quad \forall x \in \mathbb{R}$ (f sempre positiva)

2) $1 = P(X \in \mathbb{R}) = \int_R f(s)\,ds = \int_{-\infty}^{+\infty} f(s)\,ds = 1$

è una certezza!

N.B.

La funzione di massa di probabilità esiste per una v.c. continua? → NO, è priva di interesse poiché sempre nulla, non significativa

$\forall a \in R \quad p(a) = P(X=a) = \int_{x=a} f(s)ds = \int_a^a f(s)ds = 0$

↓

N.B. se $X \in [a_1, a_2]$ $a_1, a_2 \in \mathbb{R}$ e $b \in [a_1, a_2]$ → $P(X=b) = 0$ → non vuol dire che non si verifichi, ma che può assumere valori non numerabili (intervalli) e per questo NON è possibile attribuire una prob $\neq 0$ al singolo valore $(=b)$!

↳ ha senso solo con le v.c. discrete

FUNZIONE DI DISTRIBUZIONE (CASO CONTINUO)

↓

$$F(a) \equiv P(X \leq a) = \int_{-\infty}^{a} f(s)\,ds$$

↓

1) $0 \leq F(a) \leq 1 \quad \forall a \in R$

2) $F(-\infty) = 0$

3) $F(+\infty) = P(X \leq +\infty) = \int_{-\infty}^{+\infty} f(s)ds = 1$

4) $\forall a, b \in R$ con $a<b$ → $F(a) \leq F(b)$ → $\int_{-\infty}^{a} f(s)ds \leq \int_{-\infty}^{b} f(s)ds$

5) F funzione continua per le v.c. continue

per il teorema del calcolo integrale $F(a) = \int_{-\infty}^{a} f(s)ds$ quindi derivando $\frac{d}{da}F(a) = \frac{d}{da}\int_{-\infty}^{a} f(s)ds = f(a)$ rimane la funzione integranda calcolata in a quindi $\frac{dF(a)}{da} = f(a)$ → x questo le v.c. discrete non hanno funzioni di densità, non si può derivare nei punti di discontinuità!

COPPIE DI VARIABILI CASUALI (DISCRETE) → si immagina associata ad un esperimento che produce coppie di esiti (lancio 2 dadi)

(X, Y) $\quad X \in \{x_1, .. x_n\}$ $\quad Y \in \{y_1, .. y_n\}$ → FUNZIONE DI MASSA DI PROBABILITÀ CONGIUNTA → $p(a,b) = P(X=a, Y=b)$

↳ ∩ → si verifica sia $X=a$ che $Y=b$

$p: R^2 \longrightarrow [0,1]$

↓

proprietà:

1) $0 \leq p(a,b) \leq 1 \quad \forall a, b \in R$

2) $\sum_{u=1}^{m} \sum_{j=1}^{n} p(x_u, y_j) = 1$

↓

FUNZIONI DI MASSA DI PROB. MARGINALI

$P_X(a) = P(X=a)$

$P_Y(b) = P(Y=b)$

↓

se conosco la funzione congiunta è facile ricavare le funzioni marginali ma NON viceversa

↓

valgono le stesse proprietà della f. massa di una variabile normale

si concentrano solo su una v.c.

$P_X(a) = P(X=a) = P_n(X=a, Y \leq +\infty) = \sum_{j=1}^{n} p(a, y_j)$

↓

sommo la fun. di massa congiunta sulla v.c. che non mi interessa

FUNZIONE DI RIPARTIZIONE DI PROB. CONGIUNTA → $F(a,b) = P(X \leq a, Y \leq b)$ con $F: R^2 \longrightarrow [0,1]$

e MARGINALI → $F_X(a) = P(X \leq a) = P(X \leq a, Y \leq +\infty) = F(a, +\infty)$

$F_Y(b) = P(Y \leq b) = P(X \leq +\infty, Y \leq b) = F(+\infty, b)$

→ proprietà →
1) $0 \leq F(a,b) \leq 1$
2) $F(-\infty,-\infty) = P(X \leq -\infty ..) = 0$
3) $+\infty = 1$
4) F NON DECRESCE
5) F GRADINI, discontinuità

COPPIE DI V.C. CONGIUNTAMENTE CONTINUE → X e Y sono v.c. congiuntamente continue se $\exists f: R^2 \longrightarrow R^+$ tale che $\forall \mathscr{C} \subset R^2$

$$P((X,Y) \in \mathscr{C}) = \iint_{\mathscr{C}} f(s,t)\,ds\,dt$$

↓

proprietà

1) $f(a,b) \geq 0 \quad \forall a, b$

2) $1 = P((X,Y) \in R^2) = \iint_{R^2} f(s,t)\,ds\,dt$

↓

c'è anche la funzione di ripartizione → $F(a,b) = P(X \leq a, Y \leq b)$ $\quad F: R^2 \to [0,1]$ → le proprietà 1) 2) 3) 4) uguali alle coppie di v.c. discrete

5) F è continua e derivabile

e si possono ricavare quelle marginali

↓

$$F_X(a) = P(X \leq a) = P(X \leq a, Y \leq +\infty) = F(a, +\infty) = \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{a} f(s,t)\,ds\right)dt$$

↓

da cui integrando nella variabile che non mi interessa ricavo una funzione di densità di prob. marginale → $f_X(a) = \frac{d}{da}F_X(a) = \frac{d}{da}\left(\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{a} f(s,t)ds\right)dt\right) = \int_{-\infty}^{+\infty} f(a,t)\,dt$

COPPIA DI V.C. INDIPENDENTI → Due v.c. X e Y si dicono indipendenti se $\forall A, B \in \mathbb{R}$

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

↓

per verificarlo, esistono 3 teoremi

- → TEOREMA 1 → Due v.c. X e Y sono indip. se e solo se $F(a,b) = F_X(a) F_Y(b)$ $\forall a,b \in \mathbb{R}$ → vale sia x discrete che continue ma difficile app.
- → TEOREMA 2 → discrete → $p(a,b) = p_X(a) p_Y(b)$ → + facile con le funzioni di massa
- → TEOREMA 3 → continue → $f(a,b) = f_X(a) f_Y(b)$ $\forall a,b \in \mathbb{R}$ → funzioni di densità

VARIABILI CASUALI → DISCRETE / CONTINUE

- DISCRETE → MASSA → $p_X(x)$ → restituisce la prob. di osservare un valore specifico
- DISTRIBUZIONE → $F_X(x) = P(X \leq x)$ → prob. che la v.c. risulti $\leq$ di un certo valore x
- CONTINUE → DENSITÀ → $f_X(x)$ → probabilità che la v.c. cada in un certo intervallo

VALOR MEDIO → Data una v.c. X (discreta o continua) si definisce valore atteso o valor medio, se esiste → il valor medio non esiste se l'integrale o la serie divergono

$$E[X] \equiv \begin{cases} \sum_{k=1}^{m} x_k p(x_k) & \text{se } X \text{ discreta con } X \in \{x_1, \dots x_n\} \\ \int_{-\infty}^{+\infty} x f(x)\, dx & \text{se } X \text{ continua con } f(x) \text{ funz. di densità} \end{cases}$$

↓

Bernoulliana

$X \sim Be(p)$

↓

$X \in \{0,1\}$ → $X=1$ se succede A; $X=0$ se A^c

↓

$p(1) = P(X=1) = P(A) = p$

$p(0) = P(X=0) = P(A^c) = 1-p = q$

↓

$$E[X] = \sum_{k=1}^{m} x_k p(x_k) = 0\,p(0) + 1\,p(1) = p$$

vale come la probabilità che si verifichi A

↓

in generale non viene né 0 né 1, può avere valori che la v.c. non assume (baricentro)

↳ contatori

↓

PROPRIETÀ DI $E[X]$

1) Data una v.c. X e definita una seconda v.c. $Y = g(X)$

$$E[Y] = E[g(X)] = \begin{cases} \sum_{k=1}^{m} g(x_k) p(x_k) \\ \int_{-\infty}^{+\infty} g(x) f(x)\, dx \end{cases}$$

2) Conseguenza di 1) nel caso in cui

$Y = \alpha X + \beta$ con $\alpha, \beta \in \mathbb{R}$ parametri → trasf. lineare → $E[Y] = \alpha E[X] + \beta$

DIMOSTRAZIONE (CASO DISCRETO)

$$E[Y] = E[\alpha X + \beta] = E[g(X)] \underset{①}{=} \sum_{k=1}^{m} g(x_k) p(x_k) = \sum_{k=1}^{m} g(\alpha x_k + \beta) p(x_k) =$$

$$= \alpha \sum_{k=1}^{m} x_k p(x_k) + \sum_{k=1}^{m} \beta p(x_k) = \alpha \underbrace{\sum_{k=1}^{m} x_k p(x_k)}_{E[X]} + \beta \underbrace{\sum_{k=1}^{m} p(x_k)}_{1} =$$

$$= \alpha E[X] + \beta$$

N.B. se $\alpha = 0$, $E = \beta$

3) Date due v.c. X e Y e definita $Z = h(X,Y)$, se esiste $E[Z]$ è

$$E[Z] = E[h(X,Y)] = \begin{cases} \sum_{k=1}^{m} \sum_{j=1}^{n} h(x_k, y_j) p(x_k, y_j) \\ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x,y) f(x,y)\, dx\, dy \end{cases}$$

DIM. (CASO CONTINUO)

(X,Y) con $f(x,y)$

$$E[Z] = E[h(X,Y)] = E[X+Y] =$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x,y) f(x,y)\, dx\, dy =$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x+y) f(x,y)\, dx\, dy =$$

$$= \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} x f(x,y)\, dy\right) dx + \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} y f(x,y)\, dx\right) dy =$$

$$= \int_{-\infty}^{+\infty} x f_X(x)\, dx + \int_{-\infty}^{+\infty} y f_Y(y)\, dy = E[X] + E[Y]$$

4) Date due v.c. X e Y con $E[X]$ e $E[Y]$ e definita $Z = X+Y$

$$E[Z] = E[X] + E[Y]$$

4bis) Date $X_1, X_2 \dots X_n$ v.c.

$$E[X_1 + \dots + X_n] = E[X_1] + \dots + E[X_n]$$

TUTTE QUESTE PROPRIETÀ DIMOSTRANO CHE $E[X]$ È UNA FUNZIONE LINEARE

$$E[\alpha X + \beta] = \alpha E[X] + \beta$$

$$E[X+Y] = E[X] + E[Y]$$

VARIANZA → Data una v.c. con valor medio $E[X]=\mu$ si definisce varianza, se esiste

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

→ le dimensioni della varianza sono al quadrato, quindi ha senso farne la radice

↓

$\sigma = \sqrt{Var(X)}$ → SCARTO QUADRATICO MEDIO (O DEVIAZIONE STANDARD)

↓ quindi

$$Var(X) = \begin{cases} \sum_{u=1}^{m} (x_u-\mu)^2 p(x_u) \\ \int_{-\infty}^{+\infty} (x-\mu)^2 f_x \, dx \end{cases}$$

N.B. $Var(X) \geq 0$

proprietà

↓

momento secondo

1) $Var(X) = E[X^2] - (E[X])^2$ → se si conosce il valor medio basta calcolare il momento di ordine 2

2) $Var(\alpha X + \beta) = \alpha^2 Var(X)$ → la varianza è un modo per misurare quanto i valori si discostano dal valor medio tanto è più grande, tanto si discostano

3) $Var(X+Y) = ?$

↓

COVARIANZA → Data una v.c. X e Y con valori medi $E[X]=\mu_x$ ed $E[Y]=\mu_Y$, si dice covarianza, (se esiste)

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

→ è la media del prodotto di X meno la sua media per Y meno la sua media

↓

serve a capire come X e Y si influenzano

↓

$Cov(X,Y) > 0$ → X cresce o decresce con Y

$Cov(X,Y) < 0$ → X cresce quando Y decresce

$Cov(X,Y) = 0$ → X e Y sono SCORRELATE

↓

N.B. Due v.c. indipendenti sono SEMPRE scorrelate, ma due v.c. scorrelate non è detto che siano indipendenti

↓

la covarianza tra v.c. indipendenti è nulla, ma non vale il contrario (ci sono v.c. dipendenti con covarianza nulla)

↓

TEOREMA Date due v.c. X e Y indipendenti con media $E[X]=\mu_x$ e $E[Y]=\mu_Y$, allora

$$E[XY] = E[X]E[Y] \Rightarrow Cov(X,Y) = 0$$

proprietà

↓

1) $Cov(X,Y) = Cov(Y,X)$

2) $Cov(X,Y) = E[XY] - E[X]E[Y]$

3) $Cov(X,X) = Var(X)$

4) $Cov(\alpha X, Y) = Cov(X, \alpha Y) = \alpha Cov(X,Y) \quad \forall \alpha \in \mathbb{R}$

5) $Cov\left(\sum_{u=1}^{m} X_u, \sum_{j=1}^{n} Y_j\right) = \sum_{u=1}^{m}\sum_{j=1}^{n} Cov(X_u, Y_j)$

6) $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$

6bis) Date $X_1 \dots X_n$ → $Var\left(\sum_{u=1}^{n} X_u\right) = \sum_{u=1}^{n} Var(X_u) + \sum_{u=1}^{n}\sum_{j=1}^{n} Cov(X_u, X_j)$

DIMOSTRAZIONE:

$Z = X+Y \quad E[Z] = E[X+Y] = \mu_x + \mu_Y$

$$\begin{aligned} Var(Z) &= E[(Z-E[Z])^2] = E[(X+Y-\mu_x-\mu_Y)^2] \\ &= E[(X-\mu_x)^2 + (Y-\mu_Y)^2 + 2(X-\mu_x)(Y-\mu_Y)] = \\ &= E[(X-\mu_x)^2] + E[(Y-\mu_Y)^2] + E[2(X-\mu_x)(Y-\mu_Y)] = \\ &= Var(X) + Var(Y) + 2Cov(X,Y) \end{aligned}$$

per confrontare le cov. di coppie di v.c. diverse si introduce il coefficiente di correlazione

$$Corr(X,Y) \equiv \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} \qquad Corr(X,Y) \in [-1,1]$$

COME SI CALCOLA LA VARIANZA DELLA MEDIA CAMPIONARIA?

↓

$\underbrace{(X_1, X_2, \dots, X_n)}_{\text{CAMPIONE}}$ indipendenti e identicamente distribuite

↓

$E[X_1] = E[X_2] = \dots = E[X_n]$

$Var(X_1) = \dots = Var(X_n)$

↓

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

→ nuova v.c. che rappresenta la media aritmetica dei campioni (MEDIA CAMPIONARIA)

$$E[\bar{X}] = E\left[\frac{X_1 + \dots X_n}{N}\right] = \frac{1}{N} E[X_1 + \dots + X_N] =$$

$$= \frac{1}{N}(E[X_1] + \dots + E[X_2]) = \frac{1}{N}(\underbrace{\mu + \mu \dots + \mu}_{N \text{ volte}}) = \frac{N\mu}{N} = \mu$$

↓

$$Var(\bar{X}) = Var\left(\frac{1}{N}(X_1 + \dots X_N)\right) = \frac{1}{N^2} Var(X_1 + \dots + X_N) =$$

$$= \frac{1}{N^2}(Var(X_1) + \dots + Var(X_n))$$

→ stiamo trattando un campione quindi cov=0

$$= \frac{1}{N^2}(\underbrace{\sigma^2 + \dots}_{N \text{ volte}}) = \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N}$$

Cosa vuol dire la media della media campionaria? → consiste nel valore teorico che si vuole misurare, immaginando un errore che può essere positivo o negativo. μ è il valore teorico e σ^2 misura l'errore al quadrato e quanto si allontana dal valor medio.

$\bar{X}$ ha lo stesso valor medio, e può essere usata per stimare μ come possono essere usate le singole misure con il vantaggio di ridurre l'errore (sommando tanti valori insieme)

↳ $E[\bar{X}] = \mu$ e $Var(\bar{X}) = \frac{\sigma^2}{N}$

DISUGUAGLIANZA DI MARKOV → Data una v.c. $X \geq 0$ con $E[X] = \mu$ e una costante $a > 0$ → $P(X > a) \leq \frac{\mu}{a}$

DIM. CASO CONTINUO

$$\mu = E[X] = \int_{-\infty}^{+\infty} x f(x)dx \underset{X>0}{=} \int_0^{+\infty} x f(x)dx = \int_0^a x f(x)dx + \int_a^{+\infty} x f(x)dx \geq \int_a^{+\infty} x f(x)dx =$$

CI INTERESSA SOLO SE $\frac{\mu}{a}$ È UN NUM. MOLTO PICCOLO

$\frac{\mu}{a} < 1$

$$= \int_a^{+\infty} \underset{\geq a}{x} \underset{\geq 0}{f(x)}dx \geq \int_a^{+\infty} a f(x)dx = \mu \geq a P(X \geq a) \longrightarrow P(X \geq a) \leq \frac{\mu}{a}$$

$$\int_a^{+\infty} a f(x)dx = a\int_a^{+\infty} f(x)dx = a P(X \geq a)$$

le disug. si possono usare anche conoscendo solo $E[X]$ o/e $Var(X)$

DISUGUAGLIANZA DI ČEBYČEV → Data una v.c. X con $E[X] = \mu$ e $Var(X) = \sigma^2$ e data $r > 0$

DIMOSTRAZIONE

$$P(|X - E[X]| \geq r) \leq \frac{Var(X)}{r^2}$$

N.B. anche qui stima significativa solo se $\frac{\sigma^2}{r^2} < 1$

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

da Markov $(X-\mu)^2$ v.c. ≥ 0

esempio → una catena di produzione produce 50 pezzi a settimana in media; stimare la prob. che nella prossima settimana

1) il numero di pezzi prodotti non sia inferiore a 75

X: 'n° pezzi prodotti nella prossima settimana' → $P(X \geq 75) \leq \frac{E[X]}{75} = \frac{50}{75} = \frac{2}{3}$

2) → se si suppone che $Var(X) = 25$, cosa si può dire della $P(40 \leq X \leq 60)$?

$$P(40 < X < 60) = P(40-50 < X-50 < 60-50) = P(-10 < X-50 < 10) = P(|X-50| < 10)$$

in questo caso uso cebycev considerando il complementare

$$= 1 - P(|X-50| \geq 10) \geq 1 - \frac{Var(X)}{100} = 1 - \frac{25}{100} = \frac{3}{4}$$

N.B. $P(|X - E[X]| < a) = 1 - P(|X - E[X]| \geq a) \geq 1 - \frac{Var(X)}{a^2}$

LEGGE DEI GRANDI NUMERI → data una successione di v.c. $X_1, X_2, \dots X_n$ i.i.d. con $E[X_k] = \mu$ e $Var(X_k) = \sigma^2$ $(k = 1, \dots N)$, la loro media (campionaria) aritmetica converge in probabilità al valor medio μ, ovvero $\forall\, \varepsilon > 0$

$$P(|\bar{X} - \mu| \geq \varepsilon) \xrightarrow[N \to +\infty]{} 0 \quad \text{con } \bar{X} = \frac{\sum_{k=1}^{N} X_k}{N}$$

raro

la prob. che il valore assoluto della media campionaria meno la media teorica sia $\geq \varepsilon$ va a 0 se considero infinite variabili

DIMOSTRAZIONE

$E[\bar{X}] = \mu$ e $Var(\bar{X}) = \frac{\sigma^2}{N}$ $\quad |\bar{X} - E[\bar{X}]| = |\bar{X} - \mu|$

$$P(|\bar{X} - \mu| \geq \varepsilon) = P(|\bar{X} - E[\bar{X}]| \geq \varepsilon) \underset{\text{cebycev}}{\leq} \frac{Var(\bar{X})}{\varepsilon^2} = \frac{\sigma^2}{N\varepsilon^2}$$

$$P(|\bar{X} - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{N\varepsilon^2} \xrightarrow[N \to +\infty]{} 0$$

quindi

↳ essenzialmente $P(|\bar{X} - \mu| \geq \varepsilon) \leq 0$ se $N \to +\infty$, ma la probabilità ≥ 0

COROLLARIO DI BERNOULLI (legge dei grandi num.)

→ Al crescere indefinito del numero di esperimenti la frequenza relativa dell'evento A converge in probabilità alla probabilità teorica di A ($P(A)$)

se il numero di esperimenti fatti è molto alto, la frequenza relativa converge in probabilità alla probabilità teorica

$$f_A = \frac{\text{n° esperimenti in cui A si presenta}}{\text{n° esperimenti svolti} = N} \qquad \forall \varepsilon > 0 \quad P(|f_A - P(A)| \geq \varepsilon) \xrightarrow[N \to +\infty]{} 0$$

perché è un corollario della legge dei grandi numeri? →

giustificazione teorica della definizione frequentista di probabilità

DIMOSTRAZIONE: $x_1, x_2, \dots x_N$ v.c. di Bernoulli i.i.d. $X_k \sim Be(p(A)) = \begin{cases} 1 & \text{se A si verifica k-esimo esp.} \\ 0 & \text{altrimenti} \end{cases}$

$E[X_k] = P(A) \qquad Var(X_k) = P(A)(1 - P(A))$

$\bar{X} = f_A = \frac{x_1 \dots x_n}{N}$ = media arit. degli X_k → $E[\bar{x}] = P(A)$ perché $E[\bar{x}] = E[X_k]$

legge dei g.n. $\forall \varepsilon > 0 \quad P(|\bar{x} - E[x]| \geq \varepsilon) \xrightarrow[N \to +\infty]{} 0$

nel nostro caso // $P(|f_A - P(A)| \geq \varepsilon)$ //

$P(|f_A - P(A)| < \varepsilon) \xrightarrow[N \to +\infty]{} 1$

FUNZIONE GENERATRICE DEI MOMENTI

→ Data una v.c. X, si definisce funzione generatrice dei momenti (se esiste)

$$\phi(t) = E[e^{tX}] \quad t \in \mathbb{R} \qquad \text{ovvero} \quad \phi(t) = \begin{cases} \sum_{k=1}^{m} e^{tx_k} p(x_k) & \text{discreta} \\ \int_{-\infty}^{+\infty} e^{tx} f(x)\, dx & \text{continua} \end{cases}$$

proprietà

0) $\phi(0) = E[e^{0X}] = E[1] = 1$

1) $\left.\frac{d\phi(t)}{dt}\right|_{t=0} = \frac{d}{dt} E[e^{tX}]\Big|_{t=0} = E\left[\frac{d}{dt} e^{tX}\right]\Big|_{t=0} =$

$= E[e^{tX} X]\Big|_{t=0} = E[e^{0X} X] = E[X]$

general.

2) $\left.\frac{d^n \phi(t)}{dt^n}\right|_{t=0} = E[X^n]$ → ecco perché si chiama funz. generatrice dei momenti

3) Se X e Y sono v.c. INDIPENDENTI → $Z = X + Y$ → $\phi_Z(t) = \phi_{X+Y}(t) = \phi_X(t)\,\phi_Y(t)$

4) Due v.c. X e Y aventi la stessa funzione generatrice dei momenti sono identicamente distribuite, quindi hanno stessa funzione di ripartizione, stessa funzione di massa se discrete e di densità se continue

MODELLI DI V.C. DISCRETE

① V.C. DI BERNOULLI $\quad X \sim Be(p) \quad X \in \{0, 1\}$ $\begin{cases} p(1) = p \\ p(0) = 1 - p = q \end{cases}$

$E[X] = 0 \cdot q + 1 \cdot p = p \qquad E[X^2] = 1 \cdot p = p \qquad Var(X) = p - p^2 = pq$

$\phi(t) = E[e^{tX}] = e^{t \cdot 0} q + e^{t \cdot 1} p = q + e^t p$

② V.C. BINOMIALE $\quad X \sim B(n, p)$

→ si ripete n volte un esperimento in maniera identica ed indipendente
X = 'n° volte in cui l'evento A si verifica' ('n° di successi') e $p = P(A)$ è la prob. di verificarsi di A nella singola prova.

$E[X] = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k}$

oppure con $Y_k \sim Be$

$E[X] = E[Y_1 + \dots Y_n] = np$

$Var(X) = Var(\sum_{k=1}^{n} Y_k) = npq$

$\phi_X(t) = \phi_{Y_1 + Y_2 \dots}(t) = (\phi_{Y_1}(t))^n = (e^t p + q)^n$

$X \in \{0, 1, \dots, n\}$ → il numero di volte in cui osservo A non può essere $> n$ → $k = 0, 1, \dots n$

x contare i casi coeff. binomiale

$p(k) = P(X = k) = \underbrace{p \cdot p \cdots p}_{k \text{ volte}} \; \underbrace{(1-p)(1-p) \dots}_{n-k \text{ volte}}$

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

↳ k estratti n totali → $C_{n,k}$

PROPRIETÀ DI RIPRODUCIBILITÀ DELLE V.C. BINOMIALI → la somma di due v.c. binomiali è una binomiale se hanno stesso p

Se $X \sim B(n_1, p)$ e $Z \sim B(n_2, p)$ e se X e Z sono indipendenti allora $S = X + Z \sim B(n_1+n_2, p)$

DIMOSTRAZIONE → $\phi_X(t) = (pe^t+q)^{n_1}$, $\phi_Z(t) = (pe^t+q)^{n_2}$

$$\phi_S(t) = \phi_{X+Z}(t) = \phi_X(t)\,\phi_Z(t) = (pe^t+q)^{n_1}(pe^t+q)^{n_2} = (pe^t+q)^{n_1+n_2}$$

③ V.C. GEOMETRICHE

→ si ripete in maniera identica ed indipendente un esperimento fino alla comparsa di A

$X \sim G(p)$

si chiamano geometriche perché c'entrano con le serie geometriche $0 < \alpha < 1$

$$\sum_{k=0}^{+\infty} \alpha^k = \frac{1}{1-\alpha}$$

X = 'n esperimenti fatti fino alla comparsa di A' → $X \in \mathbb{N}$, $k \in \mathbb{N} \setminus \{0\}$

$$p(k) = P(X=k) = P(A_1^c \cap A_2^c \cap .. A_{k-1}^c \cap A_k) = (1-p)^k p = \underbrace{q^{k-1}}_{k-1 \text{ fallimenti}} \underbrace{p}_{\text{ultimo successo}}$$

$$E[X] = \sum_{k=1}^{+\infty} k\,p(k) = \sum_{k=1}^{\infty} k q^{k-1} p = \sum_{k=1}^{+\infty} \overbrace{k q^{k-1}}^{\frac{dq^k}{dq}} p = \sum_{k=0}^{\infty} k q^{k-1} p =$$

$$= p \sum_{k=0}^{\infty} \frac{dq^k}{dq} = p\frac{d}{dq}\left(\sum_{k=0}^{\infty} q^k\right) = p\frac{d}{dq}\left(\frac{1}{1-q}\right) = p\left(-1(1-q)^{-2}(-1)\right) =$$

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

$$E[X^2] = \frac{2}{p^2} - \frac{1}{p} \quad \to \quad Var(X) = E[X^2] - E[X]^2 = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2} = \frac{q}{p^2}$$

④ V.C. BINOMIALI NEGATIVE

→ si ripete un esperimento (i.i.d.) fino a quando si osserva r volte (anche non consecutive l'evento A)

$X \sim NB(r, p)$

negativa perché k può andare a ∞

X = 'n° di esperimenti che occorrono per osservare A r volte'

$X \in \{r, r+1, .. \infty\}$ $\quad r \in \mathbb{N} \setminus \{0\}$ $\quad p = P(A)$

$$p(k) = P(X=k) = P(\underbrace{\text{in } (k-1) \text{ esperimenti A si è verificato } r \text{ volte}}_{B((k-1),p)} \cap \underbrace{\text{nell'ultimo esp. A si verifica}}_{p}) = \binom{(k-1)}{(r-1)} p^{(r-1)} q^{(k-1)-(r-1)} = \binom{k-1}{r-1} p^r q^{k-r}$$

possiamo immaginare di descrivere X con v.c. geometriche → $Y_1, .. Y_r$ con $Y_i \sim G(p)$ → $X = Y_1 + ... Y_r$

$$E[X] = E\left[\sum_{i=1}^{r} Y_i\right] = \sum_{i=1}^{r} E[Y_i] = \sum_{i=1}^{r} \frac{1}{p} = \frac{r}{p}$$

$$Var(X) = Var\left(\sum_{i=1}^{r} Y_i\right) = \sum_{i=1}^{r} Var(Y_i) = \sum_{i=1}^{r} \frac{q}{p^2} = r\frac{q}{p^2}$$

LA SOMMA DI GEOMETRICHE NON È UNA GEOMETRICA MA UNA BINOMIALE NEGATIVA

↳ V.C. GEOMETRICHE NON RIPRODUCIBILI

⑤ V.C. DI POISSON $\quad X \sim P_0(\lambda) \quad \lambda \in \mathbb{R}^+ \quad X \in \mathbb{N} \quad p(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad k = 0,1,2\dots$

↓

approssimazione della binomiale sotto condizioni, compaiono di nuovo le serie x valor medio e varianza

ricordando che $\sum_{k=0}^{+\infty} \frac{\alpha^k}{k!} = e^{\alpha}$ $\quad \phi(t) = E[e^{tx}] = \sum_{k=0}^{+\infty} \frac{e^{tk}\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^t\lambda)^k}{k!} = e^{-\lambda} e^{(e^t\lambda)} = e^{\lambda(e^t-1)}$

$$E[X] = \frac{d}{dt}\phi(t)\Big|_{t=0} = \frac{d}{dt}\left(e^{\lambda(e^t-1)}\right)\Big|_{t=0} = \left(e^{\lambda(e^t-1)}\lambda e^t\right)\Big|_{t=0} = \lambda$$

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

$Var(X) = E[X^2] - E[X]^2 = \lambda^2+\lambda-\lambda^2 = \lambda$ → x LA V.C. DI POISSON $E[X] = Var(X)$

RIPRODUCIBILITÀ → una v.c. di Poisson è riproducibile, infatti $\left.\begin{matrix} X_1 \sim P_0(\lambda_1) \\ X_2 \sim P_0(\lambda_2) \end{matrix}\right\}$ indip. → $X_1 + X_2 \sim ?$

$$\phi_{X_1+X_2}(t) = \phi_{X_1}(t)\,\phi_{X_2}(t) = e^{\lambda_1(e^t-1)}\, e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)} \rightarrow X_1+X_2 \sim P_0(\lambda_1+\lambda_2)$$

N.B. Si dimostra che una v.c. binomiale $B(n,p)$ con $n \gg 1$ $(n \to +\infty)$ e p tale che $np = \lambda \to p = \frac{\lambda}{n} \ll 1$ (p molto piccolo) si può approssimare con una v.c. di Poisson

↓

se considero un numero molto elevato di esperimenti nel quale l'evento si verifica molto raramente, la v.c. binomiale diventa praticamente una poissoniana

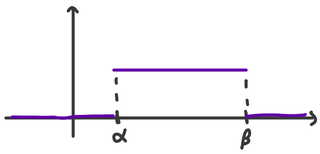
DISTRIBUZIONE DI POISSON = LEGGE DEGLI EVENTI RARI

MODELLI DI VARIABILI CASUALI CONTINUE

① VARIABILE CASUALE UNIFORME → la utilizziamo quando abbiamo a che fare con una v.c. definita in un intervallo che può assumere qualunque valore di quell'intervallo con la stessa prob.

$X \sim U(\alpha,\beta) \quad \alpha < \beta$ e $\alpha,\beta \in \mathbb{R}$

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



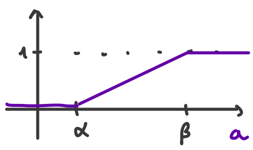
$$1 = \int_{-\infty}^{+\infty} f(x)\,dx = \int_{\alpha}^{\beta} \frac{1}{\beta-\alpha}\,dx = \frac{\beta-\alpha}{\beta-\alpha}$$

$$E[X] = \int_{-\infty}^{+\infty} x f(x)\,dx = \cancel{\int_{-\infty}^{\alpha} 0x\,dx} + \int_{\alpha}^{\beta} \frac{x}{\beta-\alpha}\,dx + \cancel{\int_{\beta}^{+\infty} 0x\,dx} = \frac{1}{\beta-\alpha}\left[\frac{x^2}{2}\right]_{\alpha}^{\beta} = \frac{1}{2}\,\frac{\beta^2-\alpha^2}{\beta-\alpha} = \frac{1}{2}(\alpha+\beta)$$

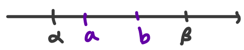
$$E[X^2] = \int_{-\infty}^{+\infty} x^2 f(x)\,dx = \cancel{\int_{-\infty}^{\alpha} x^2 0\,dx} + \int_{\alpha}^{\beta} \frac{x^2}{\beta-\alpha}\,dx + \cancel{\int_{\beta}^{+\infty} x^2 \cdot 0\,dx} = \frac{1}{\beta-\alpha}\left[\frac{x^3}{3}\right]_{\alpha}^{\beta} = \frac{1}{3}\,\frac{\beta^3-\alpha^3}{\beta-\alpha} = \frac{1}{3}(\beta^2+\alpha^2+\alpha\beta)$$

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

$$F(a) = P(X \le a) = \begin{cases} 0 & a < \alpha \\ \int_{\alpha}^{a} \frac{1}{\beta-\alpha}\,dx = \frac{a-\alpha}{\beta-\alpha} & \alpha \le a \le \beta \\ 1 & a > \beta \end{cases} \rightarrow$$



teorema → dati a e $b \in \mathbb{R}$ con $a < b$ e $\alpha < a < b < \beta$

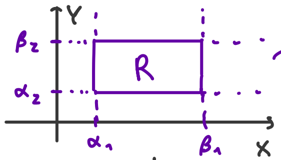


$$P(X \in [a,b]) = \frac{b-a}{\beta-\alpha}$$

↳ DIM: $\int_a^b f(x)\,dx = \int_a^b \frac{1}{\beta-\alpha}\,dx = \frac{1}{\beta-\alpha}[x]_a^b = \frac{b-a}{\beta-\alpha}$

COPPIE DI V.C. UNIFORMI E INDIPENDENTI → $\left.\begin{array}{l} X \sim U(\alpha_1,\beta_1) \\ Y \sim U(\alpha_2,\beta_2) \end{array}\right\}$ indipendenti se $\alpha_1,\alpha_2,\beta_1,\beta_2 \in \mathbb{R}$, $\alpha_1 < \beta_1$, $\alpha_2 < \beta_2$

$$f(x,y) = f_X(x)\, f_Y(y) = \begin{cases} \dfrac{1}{(\beta_1-\alpha_1)(\beta_2-\alpha_2)} & x \in [\alpha_1,\beta_1],\ y \in [\alpha_2,\beta_2] \\ 0 & \text{altrove} \end{cases}$$



→ rettangolo in cui sono definite X e Y
$A_R = (\beta_1-\alpha_1)(\beta_2-\alpha_2)$

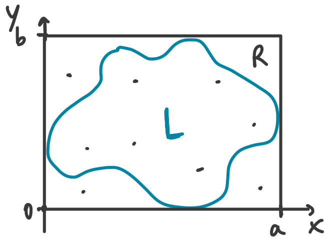
per l'area di una regione che interseca il rettangolo → $\dfrac{\text{AREA}(B \cap R)}{A_R}$

N.B. quando abbiamo due v.c. congiuntamente continue la prob. che siano identiche è nulla $P(X=Y)=0$

METODI MONTE CARLO → metodi che stimano quantità deterministiche tramite probabilità ed eventi casuali

↓

determinazione di un'area → es. lago dalla forma irregolare → come misurare la superficie del lago lanciando palle di cannone all'interno del rettangolo che circoscrive il lago?



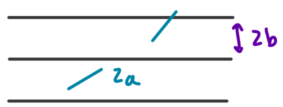
(X,Y) coordinate del punto d'arrivo di una palla $X \sim U(0,a)$, $Y \sim U(0,b)$

N_L = n° palle che finiscono nel lago L, N_{TOT} = n° lanci complessivi

$P((X,Y) \in L) = \dfrac{\text{Area } L}{\text{Area } R}$ → per il corollario di Bernoulli con $n \to +\infty$
$P((X,Y) \in L) = P(A)$

$f_A = \dfrac{n_L}{n_{TOT}}$ per $n_{TOT} \gg 1$ → $f_A \approx P(A)$ ⇒ $\dfrac{n_L}{n_{TOT}} \approx \dfrac{\text{Area } L}{\text{Area } R}$ → $\text{Area } L = \dfrac{n_L \cdot \text{Area } R}{n_{TOT}}$

PROBLEMA DEGLI SPILLI DI BUFFON → immaginiamo di avere una superficie piana suddivisa da rette parallele equidistanti (2b). Si lanciano a caso spilli (con lunghezza $2a < 2b$) su questa superficie, qual è la prob. che uno spillo intersechi una retta?



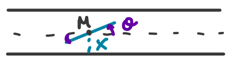
1) Considero il punto medio dello spillo e la sua distanza dalla retta più vicina →

X = 'lunghezza del segmento di perpendicolare che va da M alla retta più vicina'

possiamo supporre che X sia distribuito tra 0 (M sta sulla retta) e b (metà distanza tra una retta e l'altra) → se fosse > b si considera altra retta come + vicina

$X \sim U(0,b)$

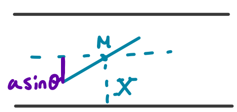
2) Introduco θ angolo formato tra lo spillo e la direzione comune alle rette →



$\theta \sim U(0,\pi)$ → è ragionevole pensare che X e θ siano indipendenti

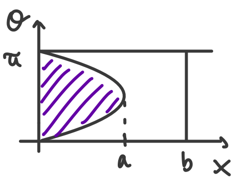
(X,θ) indip. con $f(x,\theta) = \begin{cases} \dfrac{1}{b-0}\,\dfrac{1}{\pi-0} & \text{se } x \in [0,b],\ \theta \in [0,\pi] \\ 0 & \text{altrimenti} \end{cases}$

3) Sotto quali condizioni si osserva l'intersezione tra spillo e retta più vicina? →



$X \le a\sin\theta$

CONDIZIONE DI INTERSEZIONE → $P(\text{uno spillo intersechi retta + vicina}) = P(X \le a\sin\theta)$



$$= \int_0^{\pi}\left(\int_0^{a\sin\theta} \frac{1}{b\pi}\,dx\right)d\theta = \int_0^{\pi} \frac{1}{b\pi}\, a\sin\theta\, d\theta = \frac{a}{b\pi}\int_0^{\pi}\sin\theta\, d\theta = \frac{a}{b\pi}\Big[-\cos\theta\Big]_0^{\pi} = \frac{2a}{b\pi}$$

per il corollario di Bernoulli → A = 'lo spillo interseca la retta più vicina' → $P(A) = \dfrac{2a}{b\pi} \approx f_A = \dfrac{n_A}{n_{TOT}}$

(vale se $n_{TOT} \gg 1$)

$\pi \approx \dfrac{2a}{b}\,\dfrac{n_{TOT}}{n_A}$

↱ x rappresentare tempi

<u>V.C. ESPONENZIALE</u> → una variabile continua che dipende da un parametro $X \sim E(\lambda)$ con $\lambda \in \mathbb{R}^+$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases} \qquad F(a) = P(X \leq a) = \begin{cases} 1 - e^{-\lambda a} & a \geq 0 \\ 0 & a < 0 \end{cases} \qquad P(X > a) = 1 - P(X \leq a) = \begin{cases} e^{-\lambda a} & a \geq 0 \\ 1 & a < 0 \end{cases}$$

$$\phi(t) = E[e^{tX}] = \int_{-\infty}^{+\infty} e^{tx} f(x)\,dx = \int_{-\infty}^{0} e^{tx} \cancel{0}\,dx + \int_{0}^{+\infty} e^{tx} \lambda e^{-\lambda x} dx = \lambda \int_0^{+\infty} e^{-(\lambda - t)x} dx = \lambda \left[-\frac{e^{-(\lambda-t)x}}{\lambda - t} \right]_0^{+\infty} = \frac{\lambda}{\lambda - t}$$

↳ se $\lambda - t > 0$ ovvero $t < \lambda$

$$E[X] = \frac{d\phi(t)}{dt}\Big|_{t=0} = \lambda(\lambda - t)^{-2}\Big|_{t=0} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

$$E[X^2] = \frac{d^2\phi(t)}{dt}\Big|_{t=0} = \frac{d}{dt}\left(\lambda \cdot (\lambda - t)^{-2}\right)\Big|_{t=0} = 2\lambda(\lambda - t)^{-3}\Big|_{t=0} = 2\lambda \cdot \frac{1}{\lambda^3} = \frac{2}{\lambda^2}$$

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

<u>PROPRIETÀ</u>

1) $X \sim E(\lambda)$ $(\lambda \in \mathbb{R}^+)$, $c \in \mathbb{R}^+$ $\rightarrow Y = cX \sim E\left(\frac{\lambda}{c}\right) \xrightarrow{\text{DIM.}} \phi_Y(t) = E[e^{tY}] = E[e^{tcX}] = \phi_X(tc) = \frac{\lambda}{\lambda - tc} = \frac{\lambda/c}{\lambda/c - t}$

2) Sistema formato da n dispositivi in serie, ciascun dispositivo ha una durata di funzionamento esponenziale indip. dagli altri

$D_1, D_2, \dots D_n$ dispositivi $\qquad \underbrace{X_1, X_2, \dots X_n}_{\text{indip.}}$ durata di ciascun disp. $\qquad X_k \sim E(\lambda_k) \quad k = 1, \dots n$

Y = 'durata di funzionamento del sistema' $\rightarrow Y = \min(X_1, X_2, \dots X_n)$

$$F_Y(a) = P(Y \leq a) = P(\min(X_1, \dots X_n) \leq a) = 1 - P(\min(X_1, \dots, X_n) > a) = 1 - P(X_1 > a \cap \dots X_n > a) =$$

$$= 1 - P(X_1 > a) \dots P(X_n > a) = 1 - \begin{cases} 1 & a < 0 \\ e^{-\lambda_1 a} & a \geq 0 \end{cases} \dots \begin{cases} 1 & a < 0 \\ e^{-\lambda_n a} & a \geq 0 \end{cases} = 1 - \begin{cases} 1 & a < 0 \\ e^{-\lambda_1 a} \dots e^{-\lambda_n a} & a \geq 0 \end{cases}$$

se $a \geq 0$

$$= \begin{cases} 0 & a < 0 \\ 1 - e^{-(\lambda_1 + \dots + \lambda_n)a} & a \geq 0 \end{cases} \xrightarrow{\text{quindi}} F_Y(a) = \begin{cases} 0 & a < 0 \\ 1 - e^{-\left(\sum_{k=1}^{n} \lambda_k\right)a} & a \geq 0 \end{cases} \qquad Y \sim E(\lambda_1 + \dots + \lambda_n)$$

3) Allo stesso modo per un sistema in parallelo con Z = 'durata di funzionamento del sistema'

$$F_Z(a) = P(Z \leq a) = P(\max(X_1, \dots X_n) \leq a) = P(X_1 \leq a) \cdot \dots \cdot P(X_n \leq a) = F_{X_1}(a) \cdots F_{X_n}(a) =$$

$$= \begin{cases} 0 & a < 0 \\ 1 - e^{-\lambda_1 a} & a \geq 0 \end{cases} \cdots \begin{cases} 0 & a < 0 \\ 1 - e^{-\lambda_n a} & a \geq 0 \end{cases} = \begin{cases} 0 & a < 0 \\ \prod_{k=1}^{n} (1 - e^{-\lambda_k a}) & a \geq 0 \end{cases}$$

↳ prendiamo il max perché il sistema in parallelo si guasta quando l'ultimo dei dispositivi si guasta

4) <u>ASSENZA DI MEMORIA</u> → la distribuzione futura di una v.c. esp. non dipende dalla sua storia passata (la prob. che una lampadina duri altri 3 anni, se è durata 3 anni, è uguale a quando era nuova → la v.c. si dimentica del tempo trascorso)

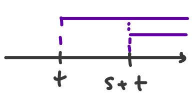
due tempi $s, t > 0$

$X \sim E(\lambda)$ $\lambda \in \mathbb{R}^+$

↓

$P(X > s+t \mid X > t) = P(X > s)$ →

DIMOSTRAZIONE: $P(X > s+t \mid P > t) = \frac{P(X > s+t \cap X > t)}{P(X > t)} =$



→ inters. viene quando $X > s+t$

$$= \frac{P(X > s+t)}{P(X > t)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s}$$

$\shortparallel$

$P(X > s)$

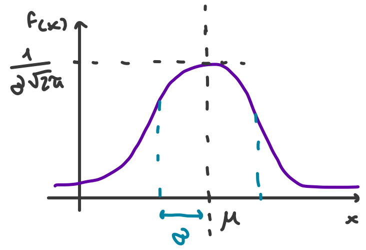
V.C. GAUSSIANA O NORMALE

→ tante v.c. indipendenti identicamente distribuite sommate insieme danno una gaussiana

$$X \sim N(\mu, \sigma^2) \qquad \mu \to E[x], \quad \sigma^2 \to Var(x)$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \rightarrow \begin{matrix} \mu \in \mathbb{R} \\ \sigma^2 \in \mathbb{R}^2 \end{matrix}$$

per $x \to \infty$ la funzione tende a 0 e l'integrale converge



- funzione simmetrica rispetto a $x = \mu$ (picco)
- $+\sigma$ curva si abbassa, $-\sigma$ curva si alza

→ σ distanza tra la media μ e il punto di flesso

cambio di variabili $\frac{x-\mu}{\sigma} = y$ $\qquad dx = \sigma\, dy$, $\quad x = \sigma y + \mu$

$$\phi(t) = E[e^{tx}] = \int_{-\infty}^{+\infty} \frac{e^{tx}}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\cancel{\sigma}\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{t\sigma y + t\mu} e^{-\frac{y^2}{2}} \cancel{\sigma}\, dy = \frac{e^{t\mu}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\left(\frac{y^2}{2} - t\sigma y\right)} dy =$$

$$= \frac{e^{t\mu}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\left(\frac{y^2}{2} - t\sigma y + \underbrace{\frac{\sigma^2 t^2}{2} - \frac{\sigma^2 t^2}{2}}_{\text{aggiungo e tolgo x ottenere un quadrato perfetto}}\right)} dy = \frac{e^{t\mu + \frac{\sigma^2 t^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(y-\sigma t)^2}{2}} dy = \frac{e^{t\mu + \frac{t^2\sigma^2}{2}}}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} dz}_{=\sqrt{2\pi}} =$$

cambio var. $y - \sigma t = z$, $dy = dz$

$$= \frac{\cancel{\sqrt{2\pi}}}{\cancel{\sqrt{2\pi}}} e^{t\mu + \frac{t^2\sigma^2}{2}}$$

$$\phi(t) = e^{t\mu + \frac{t^2}{2}\sigma^2}$$

$$E[x] = \frac{d\phi(t)}{dt}\bigg|_{t=0} = \left(e^{t\mu + \frac{t^2}{2}\sigma^2}(\mu + t\sigma^2)\right)\bigg|_{t=0} = \mu$$

$$\frac{d}{dt}(\mu + t\sigma^2) = \sigma^2$$

$$E[x^2] = \frac{d^2\phi(t)}{dt}\bigg|_{t=0} = \frac{d}{dt}\left(e^{t\mu + \frac{t^2}{2}\sigma^2}(\mu + t\sigma^2)\right)\bigg|_{t=0} = \left[\underbrace{e^{t\mu + \frac{t^2}{2}\sigma^2}(\mu + t\sigma^2)}_{\frac{d e^{t\mu + \frac{t^2}{2}\sigma^2}}{dt}} \cdot (\mu + t\sigma^2) + e^{t\mu + \frac{t^2}{2}\sigma^2}\sigma^2\right]\bigg|_{t=0} = \mu^2 + \sigma^2$$

$$Var(X) = E[x^2] - E[x]^2 = \cancel{\mu^2} + \sigma^2 - \cancel{\mu^2} = \sigma^2$$

PROPRIETÀ:

1) Se $X \sim N(\mu, \sigma^2)$ e $Y = \alpha X + \beta$ con $\beta \in \mathbb{R}$ e $\alpha \in \mathbb{R} \setminus \{0\}$ → $Y \sim N(\alpha\mu + \beta, \alpha^2\sigma^2)$

DIM → $\phi_x(t) = e^{t\mu + \frac{t^2}{2}\sigma^2}$

$$\phi_y(t) = E[e^{t(\alpha X + \beta)}] = E[e^{t\alpha X} e^{t\beta}] = e^{t\beta} E[e^{t\alpha X}] = e^{t\beta} \cdot \phi_x(t\alpha) = e^{t\beta}\left(e^{t\alpha\mu + \frac{t\alpha^2\sigma^2}{2}}\right) = e^{t(\alpha\mu + \beta) + \frac{t^2}{2}(\alpha^2\sigma^2)}$$

APPLICAZIONE DELLA PROPRIETÀ

$X \sim N(\mu, \sigma^2)$ → considero $Z = \frac{1}{\sigma}X - \frac{\mu}{\sigma^2}$ → $Z \sim N\left(\frac{1}{\sigma}\mu - \frac{\mu}{\sigma}, \frac{1}{\sigma^2}\sigma^2\right) = Z \sim N(0, 1)$ → NORMALE STANDARD

↓

a cosa serve?

x calcolare la $F(X)$

$$F_x(b) = P(X \le b) = P(x - \mu \le b - \mu) \longrightarrow P\left(\frac{x-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right) = P\left(Z \le \frac{b-\mu}{\sigma}\right) \leftarrow$$

dividi per σ $\qquad \frac{x-\mu}{\sigma}$ ↳ Z

↓

$$F_z\left(\frac{b-\mu}{\sigma}\right)$$

↓

← TAVOLE PER LA NORMALE STANDARD

- $F_z(a)$ simmetrica rispetto all'asse y, considero a^* e $-a^*$ (punti di flesso) → $F_z(0) = 1/2$ → $1 - F_z(a^*) = F_z(-a^*)$
- $P(Z > 4) \approx 0$ x via delle "code" sempre più vicine a 0

↳ regola del 3 sigma → oltre il 3σ i dati vengono scartati (prob. di osservazione molto piccola) → proprietà 2

2) PROPRIETÀ DELLA GAUSSIANA: RIPRODUCIBILITÀ

$$X_1 \sim N(\mu_1, \sigma_1^2) \text{ indip. } X_2 \sim (\mu_2, \sigma_2^2) \rightarrow Y = X_1 + X_2 \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$$

DIM. $\phi_Y(t) = \phi_{X_1+X_2}(t) = \phi_{X_1}(t)\,\phi_{X_2}(t) = e^{t\mu_1 + \frac{t}{2}\sigma_1^2} \cdot e^{t\mu_2+\frac{t}{2}\sigma_2^2} = e^{t(\mu_1+\mu_2)+\frac{t}{2}(\sigma_1^2+\sigma_2^2)}$ $\left(\phi(t) = e^{t\mu + \frac{t^2}{2}\sigma^2}\right)$

PROCESSO STOCASTICO DI POISSON → FAMIGLIA DI VARIABILI CASUALI DIPENDENTI DA UN PARAMETRO (reale)
(processo stocastico)
tempo

tante poissoniane con parametro diverso (λ costante della famiglia) e t (intervallo di tempo)

di costante λ ($\lambda \in \mathbb{R}^+$) = proc. stoc. tale che
$N(t)$ = n° eventi in $]0,t]$ (t = tempo) con le seguenti 5 proprietà

1) $N(0) = 0$

2) il n° di eventi di due intervalli di tempo disgiunti è indipendente → indipendenza degli incrementi
(n° eventi in $]0,t]$ indipendente dal n° eventi in $]s,w]$)
$N(t)$ = n° eventi in $]0,t]$
$N(w) - N(s)$ = n° eventi in $]s,w]$

3) il n° di eventi dipende dall'ampiezza dell'intervallo, ma non dalla sua posizione sull'asse reale

STAZIONARIETÀ DEGLI INCREMENTI

$t>0, s>0$
e $t<s$

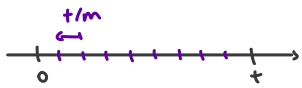
n° eventi in $]0,t] = N(t)$
n° eventi in $]s,s+t] = N(t)$ → perché l'intervallo è di lunghezza t

4) $\lim_{h \to 0} \frac{P(N(h)=1)}{h} = \lambda$ → $P(N(h)=1) \approx \lambda h$ se $h \ll 1$ → la prob. che il numero di eventi in un intervallo molto piccolo sia 1, diviso h è λ.

caratterizza un proc. stoc. rispetto a un altro, cambia la prob. di osservare un evento in un intervallo piccolo
+ è piccolo, + è piccola la possibilità si verifichi

5) $\lim_{h \to 0} \frac{P(N(h) \geq 2)}{h} = 0$ → in un piccolo intervallo ci sarà una piccola prob. di osservare un evento, ma 0 di osservarne + di 1

$N(t)$ come si distribuisce? → si distribuirà come una poissoniana.



dividi $]0,t]$ in m intervallini di lunghezza $\frac{t}{m}$ → se $m \gg 1$, $P(N(\frac{t}{m})=1) = \frac{\lambda t}{m}$ → proprietà 4

$P(N(\frac{t}{m}) \geq 2) \approx 0$ → 5

$P(N(\frac{t}{m})=0) = 1 - \frac{\lambda t}{m}$

se io so che al variare dell'intervallo nell'asse reale si notano differenze, non rientra in Poisson (casse vigilia Natale)

allora $P(N(t)=k)$ se $k \in \mathbb{N}$? → se dobbiamo trovare 3 eventi in m sottointervalli vuol dire che in 3 è capitato 1 evento, negli altri $m-3$, 0 eventi

$N(t) \sim B(m, \frac{\lambda t}{m})$ → ogni intervallino ha prob. $\frac{\lambda t}{m}$ di presentare un evento

è una binomiale perché gli eventi si verificano in maniera identica e indip. → caso limite della poissoniana: legge degli eventi rari → $N(t) \sim B(m, \frac{\lambda t}{m}) \longrightarrow N(t) \sim Po(\cancel{m}\frac{\lambda t}{\cancel{m}}) = Po(\lambda t)$

$m \to +\infty$
$(\frac{\lambda t}{m}) \to 0$

STUDIO DEGLI INTERTEMPI → tempo che intercorre tra un evento e l'altro (sempre in un processo stocastico)

$X_1, X_2, \dots X_k$ → X_1: tempo che intercorre tra l'istante iniziale e il primo evento

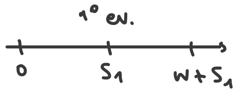
⋮

X_k: tempo che intercorre tra il $(k-1)$ esimo evento e il k-esimo evento

v.c. continue
↓
possono verificarsi in qualunque tempo → $P(X_1 > s) = P(N(s) = 0) = \frac{(\lambda s)^0}{0!} e^{-\lambda s} = e^{-\lambda s}$ → questo significa che se $s > 0$

$P(X_1 \leq s) = 1 - P(X_1 > s) = \underbrace{1 - e^{-\lambda s}}$
↓
Funzione di ripartizione di una v.c. esponenziale di parametro λ
↓
$X_1 \sim E(\lambda)$

Se $w \in \mathbb{R}^+$ $P(X_2 > w)$ vuol dire che è trascorso un certo tempo prima che si verificasse il primo evento:



→ X_2 non misura il tempo dall'istante iniziale ma da quando si é verificato il primo evento

$P(X_2 > w) = P(X_2 > w \mid X_1 = s_1)$ → intervalli di tempo disgiunti non si condizionano

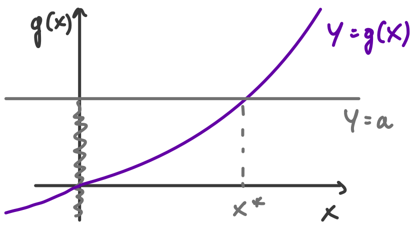
$= P(\text{in }]s_1, s_1+w] \text{ non ci sono eventi}) = P(\text{in }]0, w] \text{ non ci sono eventi})$

$= P(N(w) = 0) = \frac{(\lambda w)^0}{0!} e^{-\lambda w} = e^{-\lambda w}$ → $P(X_2 \leq w) = 1 - P(X_2 > w) = 1 - e^{-\lambda s}$ → $X_2 \sim E(\lambda)$

Per qualunque k → $X_k \sim E(\lambda)$ → ogni volta che osservo un fenomeno che si comporta come una poissoniana, gli intertempi si comportano come un esponenziale

TEORIA DELLE FUNZIONI DI VARIABILI CASUALI → X v.c. , $Y = g(X)$ → concentrandoci su v.c. continue, quanto valgono $F_Y(a)$ e $f_Y(a)$?

1) g monotona crescente

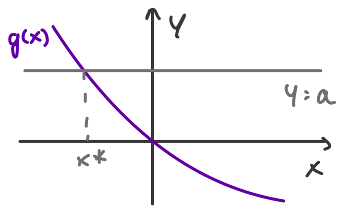


X é nota insieme a $g(X)$ → $F_Y(a) = P(Y \leq a) = P(X \leq x^*)$ $\quad x^*: g(x^*) = a$, $x^* = g^{-1}(a)$ ↓ monotona, g invertibile

$F_Y(a) = P(X \leq x^*) = P(X \leq g^{-1}(a)) = \int_{-\infty}^{g^{-1}(a)} f_X(x)\, dx$

$f_Y(a) = \frac{d}{da} F_Y(a) = \frac{d}{da} \int_{-\infty}^{g^{-1}(a)} f_X(x)\, dx = f_X(g^{-1}(a)) \frac{d g^{-1}(a)}{da}$

2) g monotona decrescente



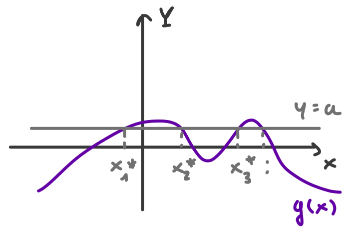
$F_Y(a) = P(Y \leq a) = P(X \geq x^*) = \int_{x^*}^{+\infty} f_X(x)\, dx = \int_{g^{-1}(a)}^{+\infty} f_X(x)\, dx$

$f_Y(a) = \frac{d}{da} F_Y(a) = \frac{d}{da}\left(-\int_{g^{-1}(a)}^{+\infty} f_X(x)\, dx\right) = -f_X(g^{-1}(a)) \cdot \frac{d g^{-1}(a)}{da}$ → derivata negativa
↓
Funzione di densità negativa: impossibile

quindi se g é una funzione monotona:

$f_Y(a) = f_X(g^{-1}(a)) \left| \frac{d g^{-1}(a)}{da} \right|$

3) g non monotona



$F_Y(a) = P(Y \leq a) = P(X \leq x_1^* \cup x_2^* \leq X \leq x_3^* \cup X \geq x_4^*) =$

$= P(X \leq x_1^*) + P(x_2^* \leq X \leq x_3^*) + P(X \geq x_4^*) =$

$= \int_{-\infty}^{x_1^*} f_X(x)\, dx + \int_{x_2^*}^{x_3^*} f_X(x)\, dx + \int_{x_4^*}^{+\infty} f_X(x)\, dx \qquad x_k^* = x_k^*(a)$

<u>VARIABILE CASUALE LOGNORMALE</u> → applicazione della teoria delle funzioni di una v.c.

$$W \sim N(\mu,\sigma^2) \rightarrow X = e^W \rightarrow X \sim \text{Lognormal}(\mu,\sigma^2)$$

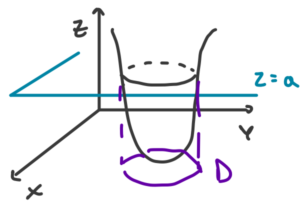
$$F_X(a) = P(X \le a) = \begin{cases} 0 & a \le 0 \\ F_Z\left(\frac{\ln(a)-\mu}{\sigma}\right) & a > 0 \end{cases} \rightarrow f_X(a) = \frac{d}{da}F_X(a) = \begin{cases} 0 & a \le 0 \\ \frac{1}{a\sigma\sqrt{2\pi}}\, e^{\frac{-(\ln(a)-\mu)^2}{2\sigma^2}} & a > 0 \end{cases}$$

$$E[X] = E[e^W] = \phi_W(1) = e^{\mu+\frac{\sigma^2}{2}} \qquad E[X^2] = \phi_W(2) = e^{2\mu+2\sigma^2} \qquad Var(X) = e^{2\mu+\sigma}(e^{\sigma}-1)$$

<u>FUNZIONI DI COPPIE DI V.C.</u> → (X,Y) coppia di v.c. note → $Z = h(X,Y)$

$$E[Z^n] = E[h^n(X,Y)] = \begin{cases} \sum_{i=1}^{m}\sum_{j=1}^{n} h^n(x_i,y_j)\,p(x_i,y_j) \\ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h^n(x,y)\,f(x,y)\,dx\,dy \end{cases}$$

$$F_Z(a) = P(Z \le a) = P(h(X,Y) \le a) = P((X,Y) \in D) = \iint_D f(x,y)\,dx\,dy$$



caso particolare → $Z = X+Y$

$$F_Z(a) = P(X+Y \le a) = P(Y \le -X+a) = \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{-x+a} f(x,y)\,dy\right)dx$$

$$f_Z(a) = \frac{d}{da}F_Z(a) = \int_{-\infty}^{+\infty}\frac{d}{da}\left(\int_{-\infty}^{-x+a} f(x,y)\,dy\right)dx = \int_{-\infty}^{+\infty} f(x,a-x)\,dx$$

→ e se X e Y indip. → $f_Z(a) = \int_{-\infty}^{+\infty} f_X(x) f_Y(y)$

<u>TEOREMA DEL LIMITE CENTRALE</u> → data una successione di v.c. $X_1, \dots X_n$ i.i.d. dove $E[X_u] = \mu$, $Var(X_u) = \sigma^2$

$$Y = \sum_{u=1}^{n} X_u$$ <u>converge in distribuzione a una gaussiana</u> per $n \rightarrow +\infty$

↓

è un teorema limite, n va mandato a ∞, ma μ e $\sigma^2 \rightarrow \infty$ quindi si prende la standardizzazione

$$Y_n = \frac{X_1 + \dots + X_n - n\mu}{\sqrt{2\sigma^2}}$$

→ in questo modo ha $\mu = 0$ e $\sigma^2 = 1$ → $\lim_{n \to +\infty} F_{Y_n}(a) = F_Z(a)$, $\lim_{n \to +\infty} \phi_{Y_n}(t) = \phi_Z(t)$

→ ma quante v.c. servono? almeno 30!

→ se $n \ge 30$ $Y_n \dot\sim N(0,1)$ → si comporta "circa"

↓

se n è finito, anche media e varianza saranno finite $S_n \dot\sim N(n\mu, n\sigma^2)$

in quest'ottica rientra la binomiale (somma di bernoulliane)

<u>APPLICAZIONI DEL TLC</u>:

- $X \sim B(n,p)$ → $X = Y_1 + \dots Y_n$ indip. con $Y_u \sim Be(p)$ se $n \ge 30$ $X \dot\sim N(np, npq)$
- Per qualsiasi $Y_1, \dots Y_n$ v.c. i.i.d. con $n \ge 30$ $X = Y_1 + \dots Y_n \dot\sim N(n\mu, n\sigma^2)$ → non standardizzo perché considero num. finito di v.c.
- $S_n = \frac{\left(\sum_{u=1}^{n} X_u\right) - n\mu}{\sigma\sqrt{n}} \dot\sim N(0,1)$ se $n \ge 30$ ⇒ divido numeratore e denom. per n → $\frac{\frac{\left(\sum_{u=1}^{n} X_u\right)}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} \dot\sim N(0,1)$ → $\bar{X} \dot\sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$\frac{\sum_{u=1}^{n} X_u}{n} = \bar{X}$, $\mu = E[X]$, $\frac{\sigma}{\sqrt{n}} = \sqrt{Var(X)}$

<u>OSSERVAZIONI</u> (necessarie per la dimostrazione)

$$Y_n = \sum_{u=1}^{n} X_u \rightarrow E[Y_n] = E\left[\sum_{u=1}^{n} X_u\right] = \sum_{u=1}^{n} E[X_u] = \sum_{u=1}^{n} \mu = n\mu$$

$$Var(Y_n) = Var\left(\sum_{u=1}^{n} X_u\right) = \sum_{u=1}^{n} Var(X_u) = \sum_{u=1}^{n} \sigma^2 = n\sigma^2$$

$$\frac{Y_n - n\mu}{\sigma\sqrt{n}} = \frac{\left(\sum_{u=1}^{n} X_u\right) - n\mu}{\sigma\sqrt{n}}$$ → standardizzazione della v.c.

$$E\left[\frac{Y_n - n\mu}{\sigma\sqrt{n}}\right] = \frac{E[Y_n] - E[n\mu]}{\sigma\sqrt{n}} = \frac{n\mu - n\mu}{\sigma\sqrt{n}} = 0 \qquad Var\left(\frac{Y_n - n\mu}{\sigma\sqrt{n}}\right) = Var\left(\frac{1}{\sigma\sqrt{n}}Y_n - \frac{n\mu}{\sigma\sqrt{n}}\right) = \frac{1}{(\sigma\sqrt{n})^2}Var(Y_n) = \frac{n\sigma^2}{n\sigma^2} = 1$$

DIMOSTRAZIONE DEL T.L.C. $\rightarrow \lim_{n\to+\infty} P\left(\frac{\left(\sum_{u=1}^{n} X_u\right)-n\mu}{\sigma\sqrt{n}} \le a\right) = F_Z(a)$ se $Z \sim N(0,1)$

1) Se $X_1, \dots X_n$ i.i.d. allora $\phi_{X_1}(t) = \phi_{X_2}(t) = \dots = \phi_{X_n}(t)$ IPOTESI PRELIMINARE: supponiamo $\mu = 0$ e $\sigma^2 = 1$

$$\phi_{X_u}(t) = e^{\ln(\phi_{X_u}(t))} = e^{L_u(t)} \qquad L_u(t) = \ln \phi_{X_u}(t)$$

$$L(t) = L_1(t) = L_2(t) \dots = L_n(t)$$

$$L(0) = \ln(\phi_{X_1}(t))\Big|_{t=0} = \ln(\phi_{X_1}(0)) = \ln(1) = 0 \longrightarrow \phi_Y(0) = 1 \ \forall Y \text{ v.c.}$$

$$L'(t)\Big|_{t=0} = \frac{d}{dt}L(t)\Big|_{t=0} = \frac{d}{dt}\left(\ln(\phi_{X_1}(t))\right)\Big|_{t=0} = \left(\frac{1}{\phi_{X_1}(t)}\frac{d}{dt}\phi_{X_1}(t)\right)\Big|_{t=0} = \frac{1}{\phi_{X_1}(0)} E[X_1] = 1 \cdot \mu = 0 \longrightarrow \frac{d}{dt}\phi_Y(t)\Big|_{t=0} = E[Y]$$

$$L''(t)\Big|_{t=0} = \frac{d^2L(t)}{dt}\Big|_{t=0} = \frac{d}{dt}\left(\frac{1}{\phi_{X_1}(t)}\frac{d\phi_X(t)}{dt}\right)\Big|_{t=0} = \left[-\frac{1}{(\phi_{X_1}(t))^2}\frac{d\phi_{X_1}}{dt}\cdot\frac{d\phi_{X_1}(t)}{dt} + \frac{1}{\phi_{X_1}(t)}\frac{d^2\phi_{X_1}(t)}{dt^2}\right]\Big|_{t=0} =$$

$$= \left[-\frac{1}{(\phi_{X_1}(0))^2}E^2[X_1] + \frac{1}{\phi_{X_1}(0)}E[X_1]^2\right] = \left[-1\cdot\mu^2 + 1\cdot(\sigma^2+\mu^2)\right] = 1$$

$\hookrightarrow E[X_1]^2 = \sigma^2 + \mu^2$ $\qquad \mu = 0 \ \sigma^2 = 1$

2) Riassumendo: $L(t)|_{t=0} = L(0) = 0$, $L'(t)|_{t=0} = L'(0) = 0$, $L''(t)|_{t=0} = L''(0) = 1$

3) Usiamo ora la v.c. S_n (che dalle osservazioni di prima $E[S_n] = 0$ e $Var(S_n) = 1$) $\rightarrow S_n = \frac{\sum_{u=1}^{n} X_u - n\mu}{\sigma\sqrt{n}}$

4) Si vuole dimostrare che $\forall a \in \mathbb{R} \rightarrow \lim_{n\to+\infty} P(S_n \le a) = F_Z(a)$ ovvero $\lim_{n\to+\infty} F_{S_n}(a) = F_Z(a)$

quindi che $\lim_{n\to+\infty} \ln(\phi_{S_n}(t)) = \ln(\phi_Z(t)) = \ln(e^{\frac{t^2}{2}}) = \frac{t^2}{2}$ $\longrightarrow \phi_{S_n}(t) = \phi_Z(t)$

5) $\phi_{S_n}(t) \underset{\substack{\mu=0\\ \sigma^2=1}}{=} \phi_{\frac{\sum_{u=1}^{n} X_u}{\sqrt{n}}}(t) \overset{\text{ind.}}{=} \phi_{\frac{X_1}{\sqrt{n}}}(t) \cdots \phi_{\frac{X_n}{\sqrt{n}}}(t) = \prod_{u=1}^{n} \phi_{\frac{X_u}{\sqrt{n}}}(t) = \prod_{u=1}^{n} E\left[e^{t\frac{X_u}{\sqrt{n}}}\right] = \prod_{u=1}^{n} \phi_{X_u}\left(\frac{t}{\sqrt{n}}\right) = \prod_{u=1}^{n} e^{L\left(\frac{t}{\sqrt{n}}\right)}$

$= e^{nL\left(\frac{t}{\sqrt{n}}\right)}$

6) Quindi sappiamo che $\ln\phi_{S_n}(t) = e^{nL\left(\frac{t}{\sqrt{n}}\right)}$ e si vuole dimostrare che $\lim_{n\to+\infty} \ln\phi_{S_n}(t) = \frac{t^2}{2}$

$$\lim_{n\to+\infty} nL\left(\frac{t}{\sqrt{n}}\right) = \lim_{n\to+\infty} \frac{L\left(\frac{t}{\sqrt{n}}\right)}{n^{-1}} = \lim_{n\to\infty} \frac{L'\left(\frac{t}{\sqrt{n}}\right)\left(-\frac{1}{2}t\,n^{-3/2}\right)}{-n^{-2}} = \lim_{n\to\infty} \frac{L'\left(\frac{t}{\sqrt{n}}\right)\frac{1}{2}t}{n^{-1/2}}$$

7) Applichiamo de l'Hopital (forma indeterminata)

$$\lim_{n\to\infty} \frac{\frac{1}{2}t\, L''\left(\frac{t}{\sqrt{n}}\right)\left(-\cancel{\frac{1}{2}}\, t\, \cancel{n^{-3/2}}\right)}{-\cancel{\frac{1}{2}}\,\cancel{n^{-3/2}}} = \lim_{n\to\infty} \frac{1}{2}t^2 L''\left(\frac{t}{\sqrt{n}}\right) = \frac{1}{2}t^2 \underset{\substack{\shortparallel\\ 1}}{L''(0)} = \frac{1}{2}t^2$$

INFERENZA STATISTICA → ha come scopo quello di determinare le proprietà di una popolazione partendo da un sottoinsieme di individui scelti a caso (CAMPIONE)

- PROBABILITÀ: esamina caratteristiche campione conoscendo la popolazione
- STATISTICA: parte dal campione x asserire qualcosa sulla popolazione

↓

il campione può essere descritto come un set di variabili casuali i.i.d. X_u e lo scopo è trovare la distribuzione $F(x_u)$ (tutti gli X_u hanno la stessa)

inferenza non parametrica

↓

stabilire la forma di F

inferenza parametrica

↓

F é nota a meno di qualche parametro che va stimato

↓

F_θ è la distribuzione di prob. comune a tutta la popolazione, ha una forma conosciuta ma θ é incognito, lo scopo é dare una stima di θ tramite l'analisi del campione

↓

STIMATORE $\hat{\theta}$ → STATISTICA (funzione delle variabili del campione) t.c.

1) Sia corretto $E[\hat{\theta}]=\theta$
2) Sia efficiente: tra tutti i $\hat{\theta}$ corretti scelgo quello con $Var(\hat{\theta})$ minima
3) Sia consistente: $\lim_{n\to\infty} E[(\hat{\theta}-\theta)^2]=0$ → $\lim_{n\to\infty} Var(\hat{\theta})=0$ → se mando a ∞ la numerosità del campione la Var → 0

↓

ad esempio, per una popolazione gaussiana

- media campionaria → $\hat{\theta}$ di μ → $\bar{X}=\sum_{u=1}^{n}\frac{X_u}{n}$
- varianza campionaria → $\hat{\theta}$ di σ^2 → $S^2=\frac{\sum_{u=1}^{n}(X_u-\bar{X})^2}{n-1}$ → perché n-1? se n=1 $S^2=\frac{0}{0}$ → x stimare la varianza occorrono almeno 2 individui nel campione

VARIABILI CASUALI CHI-QUADRO A N GRADI DI LIBERTÀ → $X \sim \chi^2_N$ → somma di normali standard^2

non può essere negativa

$Z_1 \dots Z_n$ v.c. indip. normali standard → $X=\sum_{u=1}^{N} Z_u^2$

$$E[X]=E\left[\sum_{u=1}^{N} Z_u^2\right]=\sum_{u=1}^{n} E[Z_u^2]=\sum_{k=1}^{n}(\underset{1}{Var(Z_u)}+\underset{0}{E[Z_u]})=\sum_{k=1}^{N}1=N$$

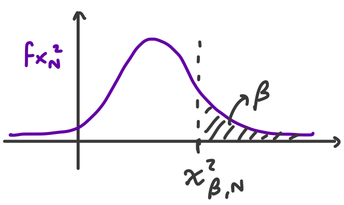
→ numero di gradi di libertà della variabile stessa

↓

grado di libertà = numero di informazioni indipendenti libere di variare nel calcolo di una stima di un parametro

$$E[Z_u^4]=\frac{d^4}{dt^4}\phi_{Z_u}(t)\Big|_{t=0}=\frac{d^4(e^{t^2/2})}{dt^4}\Big|_{t=0}=\dots=3,\quad E[Z_u^2]=1$$

$$Var(Z_u^2)=E[Z_u^4]-E[Z_u^2]^2=3-1=2 \to Var(X)=\sum_{u=1}^{n} Var(Z_u)=2n$$



→ funzione di densità, nelle tabelle della chi-quadro è rappresentato il valore critico $\chi^2_{\beta,N}$ per cui succede

↓

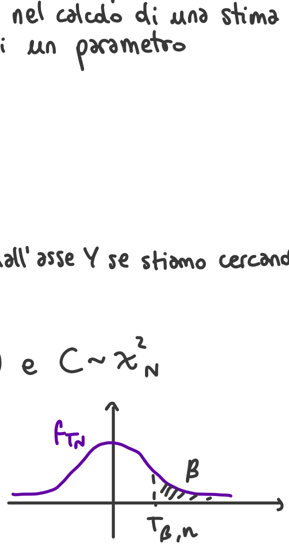
$P(X>\chi^2_{\beta,n})=\beta$ → questi valori sono + lontani dall'asse Y se stiamo cercando β molto piccolo e viceversa

VARIABILI T DI STUDENT A N GRADI DI LIBERTÀ → $Y \sim T_N$ → con $Z\sim N(0,1)$ e $C\sim\chi^2_N$

↓

la sua funzione di densità é più schiacciata e anch'essa presenta un valore critico, a differenza della χ^2_N é simmetrica

$$Y=\frac{Z}{\sqrt{\frac{C}{n}}}$$



VALORE CRITICO $T_{\beta,N}$: $P(Y>T_{\beta,n})=\beta$

↓

$T_{1-\beta,N}=-T_{\beta,N}$ → $P(Y>-T_{\beta,N})=1-\beta$

χ^2_N e T_N sono entrambe stimatori

- STIMATORE → v.c. trattata teoricamente per arrivare alla stima
- STIMA → valore dello stimatore in un certo esperimento → ripetendo l'esperimento la stima può cambiare

lo stimatore é una v.c

DISTRIBUZIONE DI PROBABILITÀ DI UNA MEDIA CAMPIONARIA → $\bar{X} = \frac{\sum_{u=1}^{n} X_u}{n}$ con X_u i.i.d.

- CASO 1 → Se $X_1, \dots X_2$ provengono da una popolazione gaussiana ($X_u \sim N(\mu, \frac{\sigma^2}{n})$) allora $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ per la proprietà di riproducibilità della gaussiana (per n vicino o > 30, vale il TLC)
- CASO 2 → Se $X_1, \dots, X_2$ provengono da una pop. <u>non</u> gaussiana → se $n \geq 30$ $\bar{X} \dot\sim N(\mu, \frac{\sigma^2}{n})$ per TLC

DISTRIBUZIONE DI PROBABILITÀ DELLA VARIANZA CAMPIONARIA → $S^2 = \frac{\sum_{u=1}^{n} (X_u - \bar{X})^2}{n-1}$

n-1 perché c'è la media campionaria e toglierla abbassa di 1 i gradi di libertà

- CASO POP. GAUSSIANA → $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ → per $n \gg 1$ χ^2_N tende a una gaussiana (TLC)

↓

tutte le volte che non conosciamo σ^2 usiamo S^2 (stimatore) e otteniamo → $\frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} \sim T_{n-1}$ → T di Student a n-1 gradi di libertà

↓

la stima puntuale di θ varia al variare dell'esperimento, quindi si preferisce introdurre un range di valori

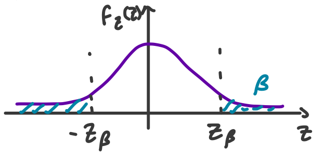
↓

INTERVALLO DI CONFIDENZA

1) INTERVALLO DI CONFIDENZA BILATERALE PER LA MEDIA μ DI UNA POP. GAUSSIANA CON σ^2 NOTA

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \longrightarrow \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

z_β valore critico di $N(0,1)$



$$P(-z_\beta \leq Z \leq z_\beta) = 1 - \beta - \beta = 1 - 2\beta \rightarrow P\left(-z_\beta \leq \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \leq z_\beta\right) = 1 - 2\beta$$

$$P\left(-z_\beta \frac{\sigma}{\sqrt{n}} \leq \bar{X} - \mu \leq z_\beta \frac{\sigma}{\sqrt{n}}\right) = P\left(-\bar{X} - z_\beta \frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{X} + z_\beta \frac{\sigma}{\sqrt{n}}\right) = P\left(\bar{X} - z_\beta \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_\beta \frac{\sigma}{\sqrt{n}}\right) = 1 - 2\beta$$

N.B. Al diminuire di z_β diminuisce l'ampiezza dell'intervallo (vantaggio) ma diminuisce anche la confidenza (svantaggio) quindi occorre aumentare n (campione numeroso per un buon intervallo di confidenza)

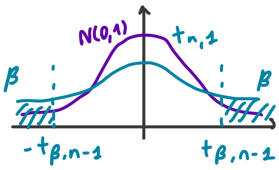
intervallo di confidenza

$\mu \in \left[\bar{X} - z_\beta \frac{\sigma}{\sqrt{n}}, \bar{X} + z_\beta \frac{\sigma}{\sqrt{n}}\right]$ con confidenza $1 - 2\beta$

2) INTERVALLO DI CONFIDENZA BILATERALE PER LA MEDIA μ DI UNA POP. GAUSSIANA CON σ^2 INCOGNITO

$\alpha = 2\beta$

$\frac{\bar{X} - \mu}{\frac{\sigma^2}{\sqrt{n}}} \sim T_{n-1}$ → la t di Student ha una curva + schiacciata di quella gaussiana



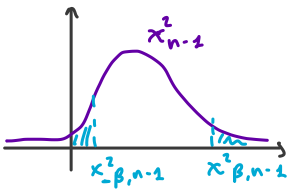
$P(-T_{\beta,n-1} \leq T \leq t_{\beta,n-1}) =$
$= 1 - P(T > T_{\beta,n-1}) + P(T < -T_{\beta,n-1})$
$= 1 - (\beta + \beta) = 1 - 2\beta$ o $1 - \alpha$

↓

poi stessi passaggi di sopra → solo che cerco nelle tavole di Student, no normale

OSS. Dato che le code della normale standard decrescono + velocemente di quelle della T_{n-1}, per β fissato $z_\beta < T_{\beta,n-1}$ → l'intervallo di confidenza con σ^2 incognito sarà più ampio di quello con σ^2 noto

3) INTERVALLO DI CONFIDENZA BILATERALE PER LA VARIANZA σ^2 DI UNA POP. GAUSSIANA



$(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$ $\qquad C \sim \chi^2_{n-1}$

$$P(\chi^2_{1-\beta,n-1} \leq C \leq \chi^2_{\beta,n-1}) = 1 - P(C > \chi^2_{\beta,n-1}) + P(C < \chi^2_{1-\beta,n-1}) =$$

$$= 1 - (\beta + \beta) = 1 - 2\beta$$

$$P\left(\chi^2_{1-\beta,n-1} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\beta,n-1}\right) = P\left(\frac{1}{\chi^2_{\beta,n-1}} \leq \frac{\sigma^2}{(n-1)S^2} \leq \frac{1}{\chi^2_{1-\beta,n-1}}\right) =$$

$$P\left(\frac{(n-1)S^2}{\chi^2_{\beta,n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\beta,n-1}}\right) = 1 - 2\beta$$

↳ non è simmetrico