



Dipartimento  
di Matematica  
Università di Pisa

# APPUNTI DEL CORSO DI **ELEMENTI DI PROBABILITÀ E STATISTICA**

A cura di Chiara Di Sano  
[c.disano1@studenti.unipi.it](mailto:c.disano1@studenti.unipi.it)

Rielaborazione delle lezioni dei prof.  
M. Romito  
G. Di Gesù  
A.A. 2020-2021

## Come nasce la PROBABILITÀ?

La probabilità è una materia giovane, abbiamo sentito parlare di lei la prima volta nel XIX secolo.

Essendo una disciplina giovane, ci sono ancora discussioni filosofiche su di essa.

Oggi la probabilità ha i suoi fondamenti nel sistema assiomatico di Kolmogorov, resta solo da capire quali modelli scegliere.

ESEMPIO: LANCIO DI UNA MONETA ← esperimento completamente aleatorio →

Il problema nel sistema di Kolmogorov è fondato, resta solo da capire quanto valga la probabilità che il lancio di una moneta restituisca T o C

Perché?  
il lancio è influenzato da piccoli fattori che hanno un risultato imprevedibile

Ci sono vari modelli differenti:

### - DESCRIZIONE FREQUENTISTA (statistica):

Si basa sulla frequenza dell'accadere di un evento su un numero finito di esperimenti.

↑  
mette in relazione il nostro modello con quello reale, ecco cosa fa la statistica

↑ Questo modello non può funzionare sempre, ad esempio se volessi calcolare la probabilità che un vulcano esploda non posso farlo perché non posso "ripetere l'esperimento". In questi casi usiamo paradigmi differenti.

## • Che cos'è il SISTEMA ASSIOMATICO?

### IL SISTEMA ASSIOMATICO DI KOLMOGOROV

Il sistema assiomatico si basa sull'idea che la probabilità sia la misura di qualcosa (la facilità con cui qualcosa accade).  
Dunque gli assiomi della probabilità, detti anche assiomi di Kolmogorov dal nome del matematico russo che li definisce per primo.

L'idea di fondo di un problema che noi esaminiamo con strumenti probabilistici è legato all'idea che noi stiamo esaminando un problema che ha delle componenti aleatorie, cioè alcuni degli esentudismi che regolano il fenomeno che noi stiamo osservando non dipendono da leggi, ma sono completamente casuali.

Dunque possiamo immaginare i nostri esperimenti completamente aleatori.

La Probabilità non calcola le cose, il suo compito non è quello di provare a prevedere il singolo risultato, ma prevede il comportamento statistico del problema, cioè le frequenze di uscita e le quantifica.

(Stimiamo in maniera più o meno precisa l'INCIRCA)

### DEFINIZIONI (del sistema assiomatico):

- $\Omega$ : insieme che contiene tutti gli esiti possibili (esiti elementari) di un loro dato evento.

$\sigma$ -algebra: è un insieme di parti di  $\Omega$

Def: Dato un insieme  $\Omega$ , si definisce  $\sigma$ -algebra su  $\Omega$  una famiglia  $\mathcal{F}$  di sottoinsiemi di  $\Omega$  tale che:

- $\emptyset, \Omega \in \mathcal{F}$
- se  $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
- se  $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$

Uno spazio misurabile è una coppia  $(\Omega, \mathcal{F})$  costituita da un insieme non vuoto  $\Omega$  ed una  $\sigma$ -algebra  $\mathcal{F}$  su  $\Omega$ .

Gli elementi di  $\mathcal{F}$  sono detti insiemi numerabili di  $\Omega$ .

Dunque la probabilità è una misura:

Probabilità:  $(\Omega, \mathcal{F}) \quad \mathbb{P}: \mathcal{F} \rightarrow [0, 1]$

è una funzione  
dell'insieme  $\sigma$ -algebra  
all'intervallo  $[0, 1]$   
impossibile  $\rightarrow$  certo

Allora: -  $\mathbb{P}(\Omega) = 1$  = È certo che almeno un elem.  
di  $\Omega$  si verifichi

-  $\mathbb{P}(\emptyset) = 0$  = È impossibile che non si verifichi nulla

- se  $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$  a due a due disgiunti, allora  $\mathbb{P}(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$

Uno spazio di probabilità  
è una terna  $(\Omega, \mathcal{F}, \mathbb{P})$   
cambiamo in base al problema

Questa somma è ben definita,  
perché le probabilità sono,  
per definizione, numeri  $\geq 0$ ,  
quindi per serie positive la  
serie è ben definita

# I RISULTATI PIÙ IMPORTANTI DELLA PROBABILITÀ

I risultati più importanti della probabilità sono i risultati limite.

Il limite di infiniti esperimenti è la formalizzazione matematica del numero grande di esperimenti che uno può fare praticamente.  
[esempio: su un singolo lancio di una moneta non sappiamo dire nulla ma sappiamo dire qualcosa su più lanci]

## IL CASO DI UNO SPAZIO FINITO

Nel caso in cui  $\Omega$  sia un insieme finito e gli eventi elementari sono equiprobabili/simmetrici, si parla di DISTRIBUZIONE UNIFORME DI PROBABILITÀ SU  $\Omega$ .

N.B. Non esiste una distribuzione uniforme di probabilità su un insieme  $\Omega$  numerabile ma infinito.

Ora torniamo al caso di  $\Omega$  finito e di distribuzione uniforme di probabilità e definiamo la nostra funzione probabilità:

$$P(A) = \frac{\#A}{\#\Omega} \quad \text{con } A \subset \Omega$$

definizione classica di probabilità

rapporto tra casi favorevoli e casi possibili

ESEMPIO:

1 dado  $\Omega = \{1, 2, \dots, 6\}$

2 dadi  $\Omega = \{(1, 1), (1, 2), \dots, (6, 6)\}$

## PROPRIETÀ

•  $P(\emptyset) = 0$

•  $P(\Omega) = 1$

•  $0 \leq P(A) \leq 1$  poiché  $A \subset \Omega$ , il numero di A sarà  $\leq \Omega$

• se  $A \subset B$  allora  $P(A) \leq P(B)$  cioè B è più ricco di esiti possibili perciò è più facile che accada, ogni volta che accade A accade B, quindi B può accadere più spesso di quanto accade A, dunque la probabilità di B deve essere per forza maggiore

• se  $A \cap B = \emptyset$ , allora  $P(A \cup B) = P(A) + P(B)$ , per la cardinalità questo è chiaro, la cardinalità dell'unione disgiunta è uguale alla somma delle singole cardinalità, ma è chiaro anche per la probabilità

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

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

- formula di inclusione - esclusione:

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n (-1)^{k+1} \sum_{\substack{J \subset \{1, \dots, n\} \\ \#J = k}} P\left(\bigcap_{i \in J} A_i\right)$$

va dimostrata per induzione

Mostriamo che queste proprietà sono vere se:

- $0 \leq P(\dots) \leq 1$

- $P(\Omega) = 1$

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

- $\Omega \cup \emptyset = \Omega \quad 1 = P(\Omega) = P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset)$

$\Omega \cap \emptyset = \emptyset$  quindi  $P(\emptyset) = 0$

- se  $A \subset B$  allora  $B = A \cup B \setminus A \quad A \cap (B \setminus A) = \emptyset$  quindi  $P(B) = P(A) + \underbrace{P(B \setminus A)}_{\geq 0} \geq P(A)$

- $A \cup A^c = \Omega$

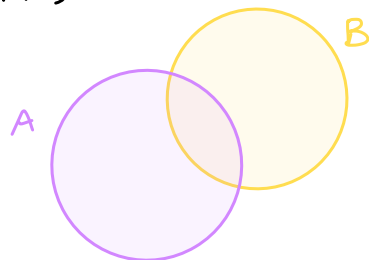
$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$$

$A \cap A^c = \emptyset$

- $A, B \quad B = (B \setminus A) \cup (A \cap B)$

$$A = (A \setminus B) \cup (A \cap B)$$

$$A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$$



a margine osserviamo che se  $(A_i)_{i=1, \dots, n}$  di eventi a due a due disgiunti, allora  $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$

dim per induzione,  $n=2$  vero. Supponiamo la proprietà vera per  $n-1$  eventi.

$$\bigcup_{i=1}^n A_i = \left(\bigcup_{i=1}^{n-1} A_i\right) \cup A_n$$

$$P\left(\bigcup_{i=1}^n A_i\right) = P\left(\bigcup_{i=1}^{n-1} A_i\right) + P(A_n) = \sum_{i=1}^{n-1} P(A_i) + P(A_n) = \dots$$

$$P(A \cup B) = \underbrace{P(A \setminus B)}_{P(A) - P(A \cap B)} + \underbrace{P(B \setminus A)}_{P(B) - P(A \cap B)} + \underbrace{P(A \cap B)}_{P(A \cap B)} = P(A) + P(B) - P(A \cap B)$$

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

Esempi: • lancio di una o più monete/dadi

• estrazione di carte/gettoni

• lancio di una moneta:  $\Omega = \{0, 1\}$  probabilità uniforme  $P(0) = P(1) = 1/2$

• lancio di due monete / lancio di una moneta due volte

$\Omega_0 = \{0,0\}, \{0,1\}, \{1,1\}$   $\hookrightarrow$  Quale dei due è uniforme?

$\Omega = \{(0,0), (0,1), (1,0), (1,1)\}$

Mar + Mer: Romito (Teoria)

Lun: Di Gesù (Esercizi) giacomo.di.gesù@unipi.it  
11-12:30

Richiamo:

Insieme  $\Omega$  ("insieme dei risultati possibili")

$\sigma$ -algebra  $\mathcal{F} \subset \mathcal{P}(\Omega)$  ("insieme degli eventi")

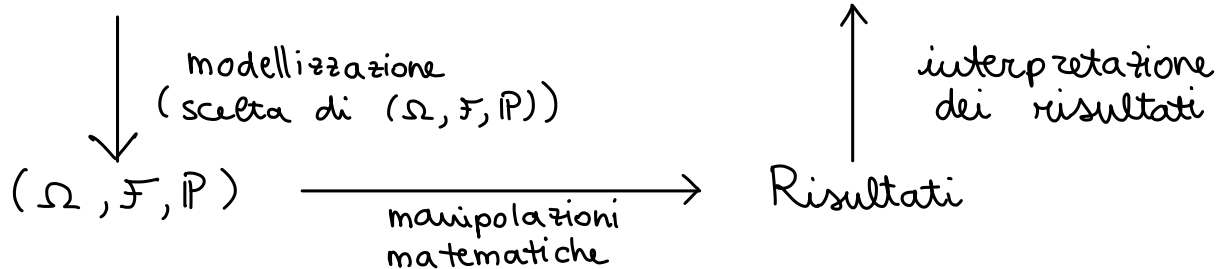
probabilità  $P: \mathcal{F} \rightarrow [0, 1]$  tale che

- $P(\Omega) = 1$

- $P\left(\bigcup_{n=2}^{\infty} A_n\right) = \sum_{n=2}^{\infty} P(A_n)$  se  $A_n \in \mathcal{F} \forall n$  e  $A_k \cap A_n = \emptyset \quad k \neq n$

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

Dato un esperimento aleatorio



① lanciamo un dado due volte

Qual è la probabilità che la differenza  $D$  tra il punteggio massimo e quello minimo sia maggiore o uguale a tre. E quella che sia minore o uguale a uno?

$$P(D \geq 3) = ?$$

$$P(D \leq 1) = ?$$

Soluzione:  $P(D \geq 3) = \frac{1}{3}$

$$P(D \leq 1) = \frac{4}{9}$$

$$\Omega = \{(1,1), (1,2), \dots, (1,6), \\ (2,1), \dots, (2,6) \\ \vdots \\ (6,1), \dots, (6,6)\}$$

Consideriamo la probabilità uniforme su  $\Omega$  per ragioni di simmetria, ovvero  $P(A) = \frac{|A|}{|\Omega|}$

$$|\Omega| = 6^2 = 36$$

$$\begin{array}{l} |\{D \geq 3\}| \longrightarrow \begin{array}{ccc} 1, 4 & 2, 5 & 3, 6 \\ 1, 5 & 2, 6 & \\ 1, 6 & & \end{array} \\ |\{D \leq 1\}| \end{array}$$

$$2 \cdot 3 + 2 \cdot 2 + 2 \cdot 1 = 12$$

$$P(D \geq 3) = \frac{12}{36} = \frac{1}{3}$$

moltiplico per 2 poiché l'ordine dei numeri è indifferente

(2) Siano  $A, B$  due eventi con  $P(A) = \frac{3}{4}$  e  $P(B) = \frac{1}{3}$

• Dimostrare le disuguaglianze  $\frac{1}{12} \leq P(A \cap B) \leq \frac{1}{3}$

• Fare degli esempi per cui vale  $P(A \cap B) = \frac{1}{12}$  e  $P(A \cap B) = \frac{1}{3}$

Soluzione

(\*\*) Usiamo  $F \subset G \Rightarrow P(F) \leq P(G)$

$$A \cap B \subset B \text{ (sempre)} \Rightarrow P(A \cap B) \leq P(B) = \frac{1}{3}$$

$$(*) P(A \cap B) \geq \frac{1}{12}$$

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

FORMULA DI DISINTEGRAZIONE

$$P(A \cap B) = P(A) - P(A \cap B^c)$$

$$\textcircled{>} P(A) - P(B^c) = \frac{3}{4} - \left(1 - \frac{1}{3}\right) = \frac{3}{4} - \frac{2}{3} = \frac{1}{12}$$

Usiamo il fatto che  $P(A \cap B^c) \leq P(B^c)$  perché  $A \cap B^c \subset B^c$   
 $\Rightarrow -P(A \cap B^c) \geq -P(B^c)$

Esempi per cui valgono le uguaglianze

$\Omega = \{1, 2, \dots, 12\}$  con probabilità uniforme

$|A| = 9$  scegliamo  $A = \{1, \dots, 9\}$  e  $B = \{1, \dots, 4\}$

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

invece prendendo  $A = \{1, \dots, 9\}$  e  $B = \{9, \dots, 12\}$

avremmo  $P(A \cap B) = \frac{1}{12}$

(Per casa: Dare delle stime ottimali per  $P(A \cup B)$ )

④  $A, B, C, D$  quattro eventi

Assumiamo (a)  $P(A) = \frac{1}{2}$

(b)  $P(A \cap B \cap D) = \frac{1}{4}$

(c)  $P(A \cap B \cap C \cap D) = \frac{1}{9}$

Calcolare se possibile <sup>Le ipotesi potrebbero non essere sufficienti</sup>

(i)  $P(A \cap (B^c \cup D^c)) = ?$

(ii)  $P(A \cap [(B \cap D)^c \cup C]) = ?$

Domanda preliminare  
Esiste un tale spazio di probabilità? Si verifica costruendolo "a mano".

Svolgimento:

Richiamo di algebra degli insiemi

Leggi di De Morgan:  $\cdot (F \cup G)^c = F^c \cap G^c$

$\cdot (F \cap G)^c = F^c \cup G^c$

Leggi distributive:  $\cdot (F \cap G) \cup H = (F \cup H) \cap (G \cup H)$

$\cdot (F \cup G) \cap H = (F \cap H) \cup (G \cap H)$

$P(A \cap (B^c \cup D^c)) = P(A \cap (B \cap D)^c) \stackrel{(*)}{=} P(A) - P(A \cap B \cap D) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$

(\*) Per la formula di disintegrazione

Sia  $F = (B \cap D)^c \Rightarrow F^c = B \cap D$

$P(A) = P(A \cap F) + P(A \cap F^c) \Rightarrow P(A \cap F) = P(A) - P(A \cap F^c)$

$\Rightarrow P(A \cap (B \cap D)^c) = P(A) - P(A \cap B \cap D) = \frac{1}{2} - \frac{1}{4} = \frac{2-1}{4} = \frac{1}{4}$

(ii)  $P(A \cap [(B \cap D)^c \cup C]) = ?$

Per casa

④ Viene lanciato 6 volte un dado a 5 facce. Qual è la probabilità di ottenere almeno una volta il 3, il 4 e il 5? (ad esempio: (4, 4, 3, 2, 5, 2))

Soluzione

$\Omega = \{1, \dots, 5\}^6$  Suppongo la probabilità uniforme su  $\Omega$   
quindi vale  $P(A) = \frac{|A|}{|\Omega|} = \frac{\text{casi favorevoli}}{\text{casi possibili}}$   
 $|\Omega| = 5^6$   $|A| = ?$

Idea 1: Passaggio al complementare:  $P(A) = 1 - P(A^c)$

$A = A_3 \cap A_4 \cap A_5$  con  $A_k = \text{"esce almeno una volta } k"$   
e quindi  $A_k^c = \text{"non esce mai } k" = B_k$

Idea 2: Formula di Inclusione - Esclusione

$$P(B_3 \cup B_4 \cup B_5) = P(B_3) + P(B_4) + P(B_5) - P(B_3 \cap B_4) - P(B_3 \cap B_5) - P(B_5 \cap B_4) + P(B_3 \cap B_4 \cap B_5)$$

$|B_k| = 4^6$  ho 6 possibilità per  $5-1=4$  numeri

$|B_k \cap B_j| = 3^6$   $k \neq j$

$|B_k \cap B_j \cap B_i| = 2^6$   $k \neq j, j \neq i, i \neq k$

$$\text{Quindi } P(A) = 1 - 3\left(\frac{4}{5}\right)^6 + 3\left(\frac{3}{5}\right)^6 - \left(\frac{2}{5}\right)^6 \approx 0.34944$$

Domanda di combinatoria

⑤ Quanti sono gli interi positivi  $\leq 1000$  e divisibili per almeno uno tra 2, 3 e 5?

OPPURE

Qual'è la probabilità di estrarre un  $n$  divisibile per 2, 3 o 5 tra gli interi  $\leq 1000$ ?

Domanda di probabilità

Soluzione:

$A = \{n \in \mathbb{N}_{>0} \mid 2|n \text{ o } 3|n \text{ o } 5|n\} = A_2 \cup A_3 \cup A_5$  con  $A_k = \{n \in \mathbb{N}_{>0} \mid k|n\}$

$$\begin{aligned} |A| &= |A_2 \cup A_3 \cup A_5| = |A_2| + |A_3| + |A_5| - |A_2 \cap A_3| - |A_3 \cap A_5| - |A_2 \cap A_5| + |A_2 \cap A_3 \cap A_5| \\ &= \frac{1000}{2} + \left\lfloor \frac{1000}{3} \right\rfloor + \frac{1000}{5} - \overset{A_6}{\left\lfloor \frac{1000}{6} \right\rfloor} - \overset{A_{15}}{\frac{1000}{15}} - \overset{A_{10}}{\frac{1000}{10}} + \overset{A_{30}}{\left\lfloor \frac{1000}{30} \right\rfloor} = \\ &= 500 + 333 + 200 - 166 - 66 - 100 + 33 = 734 \end{aligned}$$

# Modello generale (Kolmogorov)

Lezione 2  
02/03/2021  
(Rovito)

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

$0 \leq P \leq 1$

$\sigma$ -algebra

$P(\Omega) = 1$  si verifica almeno un evento possibile

$P(\cup A_n) = \sum P(A_n)$  con  $(A_n)_{n \in \mathbb{N}}$  a due a due disgiunti

la prob di più eventi disgiunti è la somma delle singole probabilità

## Probabilità uniforme

$\Omega$ : insieme finito  $P(A) = \frac{|A|}{|\Omega|} = \frac{\text{casi favorevoli}}{\text{casi possibili}}$

## Probabilità finite

$\Omega$ : insieme finito insieme delle parti

Una probabilità:  $P: \mathcal{P}(\Omega) \rightarrow \mathbb{R}$

- $P(\Omega) = 1$
- $0 \leq P(A) \leq 1 \quad \forall A \subset \Omega$
- Se  $A, B$  sono incompatibili ( $A \cap B = \emptyset$ ):  $P(A \cup B) = P(A) + P(B)$

monete lanciate  $n$  volte

$\Omega = \{ (0,0), (0,1), (1,0), (1,1) \}$

probabilità uniforme

$\frac{1}{4} \quad \frac{1}{4} \quad \frac{1}{4} \quad \frac{1}{4}$

Numero di teste

$\Omega = \{ 0, 1, 2 \}$

$\frac{1}{4} \quad \frac{1}{2} \quad \frac{1}{4}$

Osservazione Le proprietà delle probabilità che abbiamo dimostrato, discenderanno tutte dalla finite addittività

- $A \subset B \rightarrow P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- inclusione-esclusione ...

Costruzione Introdurre densità (discrete) ovvero

$$p: \Omega \rightarrow [0, +\infty)$$

e definiamo

$$(*) \quad P(A) = \frac{\sum_{s \in A} p(s)}{\sum_{s \in \Omega} p(s)}$$

1) (\*) definisce una probabilità su  $\Omega$

2) ad una probabilità  $P$  su un insieme finito  $\Omega$  è possibile associare una densità (discreta)

Per (1) mostriamo l'additività:  $A \cap B = \emptyset \quad A, B \subset \Omega$

$$P(A \cup B) = \frac{\sum_{s \in A \cup B} p(s)}{\sum_{s \in \Omega} p(s)} = \frac{\sum_{s \in A} p(s) + \sum_{s \in B} p(s)}{\sum_{s \in \Omega} p(s)} = P(A) + P(B)$$

Viceversa, per (2), se  $s \in \Omega \quad p(s) = P(\{s\})$

Densità (discreta) Dato  $\Omega$  insieme finito, una densità (discreta) su  $\Omega$  è una funzione  $p: \Omega \rightarrow \mathbb{R}$

•  $p(s) \geq 0$  per ogni  $s \in \Omega$

•  $\sum_{s \in \Omega} p(s) = 1$

(definizione ufficiale)

esempio numero di teste nel lancio di  $n$  monete (codifica 0 croce, 1 testa)

• modello uniforme  $\Omega = \left\{ \begin{array}{l} \text{insieme di tutte le stringhe} \\ \text{di lunghezza } n \text{ composte} \\ \text{da 0 e 1} \end{array} \right\}$

$= \left\{ \begin{array}{l} \text{i numeri da } 0 \text{ a } 2^n - 1 \\ \text{in binario} \end{array} \right\}$

• numero totale di teste  $\Omega = \{0, 1, \dots, n\} \quad K = 0, 1, \dots, n$

probabilità di  $K$  teste  $= \frac{\binom{n}{K}}{2^n}$

$$p(K) = \frac{1}{2^n} \binom{n}{K}$$

•  $p(K) \geq 0$  per ogni  $K$

•  $\sum_{K=0}^n p(K) = 1$

## INDIPENDENZA

$(\Omega, P)$  Due eventi  $A, B \subset \Omega$  sono indipendenti se

$$P(A \cap B) = P(A) P(B)$$

Esempio lancio di una moneta due volte

$$A = \{ \text{testa al 1° lancio} \} = \{10, 11\} \quad P(A) = 1/2$$

$$B = \{ \text{croce al 2° lancio} \} = \{00, 10\} \quad P(B) = 1/2$$

$$A \cap B = \{10\} \quad P(A \cap B) = 1/4 = P(A) P(B)$$

Gli eventi  $A_1, A_2, \dots, A_n$  sono (collettivamente) indipendenti se per ogni  $K = 1, \dots, n$  e per ogni  $i_1, \dots, i_K \in \{1, \dots, n\}$

$$P\left(\bigcap_{j=1}^K A_{i_j}\right) = \prod_{j=1}^K P(A_{i_j})$$

osservazione: ogni sottofamiglia di eventi indipendenti è composta da eventi indipendenti

Nota ① L'indipendenza a 2 a 2 non è indipendenza

$$\Omega = \{1, 2, 3, 4\} \quad A_1 = \{1, 2\} \quad A_2 = \{2, 3\} \quad A_3 = \{1, 3\}$$

prob uniforme

$$P(A_i) = 1/2 \quad i = 1, 2, 3$$

$$P(A_1 \cap A_2) = P(\{2\}) = \frac{1}{4}$$

quindi  $A_1, A_2$  indipendenti (così come  $A_1, A_3$  e  $A_2, A_3$ )

$$P(A_1 \cap A_2 \cap A_3) = P(\emptyset) = 0$$

$A_1, A_2, A_3$  non sono indipendenti

② che la probabilità dell'intersezione di tutti gli eventi fattorizza

non garantisce l'indipendenza:  $\Omega = \{1, \dots, 16\}$  con p. uniforme e

$A_1 = \{1, 2, \dots, 8\}$ ,  $A_2 = \{4, 5, \dots, 11\}$  e  $A_3 = \{7, 8, \dots, 14\}$ , allora si ha

$$P(A_1 \cap A_2 \cap A_3) = P(\{7, 8\}) = 1/8 = P(A_1) \cdot P(A_2) \cdot P(A_3) = \left(\frac{8}{16}\right)^3 = \left(\frac{1}{2}\right)^3 \quad \text{ma}$$

$$P(A_1 \cap A_2) = P(\{4, 5, \dots, 8\}) = \frac{5}{16} \neq \underbrace{P(A_1)}_{\frac{8}{16}} \cdot \underbrace{P(A_2)}_{\frac{8}{16}} = \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

Proposizione  $(\Omega, \mathbb{P})$   $A_1, \dots, A_n$  indipendenti allora  
per ogni  $(d_1, \dots, d_n) \in \{0, 1\}^{2n}$

$A_1^{d_1}, \dots, A_n^{d_n}$  sono indipendenti

dove  $A^0 = A$  e  $A^1 = A^c =$  complementare di  $A$

Lemma (oss preliminare: la nozione di indipendenza è  
invariante per permutazione degli indici che individuano  
la famiglia di eventi indipendenti)

È sufficiente mostrare l'affermazione:

se  $A_1, \dots, A_n$  indipendenti, allora  $A_1^c, A_2, \dots, A_n$  indipendenti

(esempio  $A_1 A_2^c A_3^c A_4$ , usiamo l'affermazione  
 $A_1 A_2 A_3 A_4 \xrightarrow{P} A_2 A_1 A_3 A_4 \xrightarrow{a} A_2^c A_1 A_3 A_4 \xrightarrow{P} A_3 A_2^c A_1 A_4$   
 $\xrightarrow{a} A_3^c A_2^c A_1 A_4 \xrightarrow{P} A_1 A_2^c A_3^c A_4$ )

Fissiamo  $K \geq 1$  e  $i_1, \dots, i_K \in \{1, \dots, n\}$  distinti

• caso 1:  $1 \notin \{i_1, \dots, i_K\}$

$$\mathbb{P}\left(\bigcap_{j=1}^K A_{i_j}\right) = \prod_{j=1}^K \mathbb{P}(A_{i_j}) \quad \text{perché } A_2, \dots, A_n \text{ indep.}$$

• caso 2:  $1 \in \{i_1, \dots, i_K\}$  ad esempio  $i_1 = 1$

$$\begin{aligned} \mathbb{P}\left(A_1^c \cap \bigcap_{j=2}^K A_{i_j}\right) &= \\ &= \mathbb{P}\left(\bigcap_{j=2}^K A_{i_j}\right) - \mathbb{P}\left(\bigcap_{j=1}^K A_{i_j}\right) \\ &= \prod_{j=2}^K \mathbb{P}(A_{i_j}) - \prod_{j=1}^K \mathbb{P}(A_{i_j}) \\ &= \prod_{j=2}^K \mathbb{P}(A_{i_j}) (1 - \mathbb{P}(A_1)) \\ &= \mathbb{P}(A_1^c) \prod_{j=2}^K \mathbb{P}(A_{i_j}) \end{aligned}$$

$$\bigcap_{j=2}^K A_{i_j} = \underbrace{\left(\bigcap_{j=2}^K A_{i_j}\right) \cup \left(A_1^c \cap \bigcap_{j=2}^K A_{i_j}\right)}_{\text{unione disgiunta}}$$

Proposizione Sia  $(\Omega, \mathcal{P})$ , se  $A_1, \dots, A_n$  sono eventi in  $\Omega$ , allora sono equivalenti

- $A_1, \dots, A_n$  indipendenti

- per ogni  $d_1, \dots, d_n \in \{0, 1\}$   $P(\bigcap_{i=1}^n A_i^{d_i}) = \prod_{i=1}^n P(A_i^{d_i})$

## PROBABILITÀ CONDIZIONATA

esempio Supponiamo di avere un sacchetto con 50 monete con due facce "testa" e 50 monete "testa/croce". Qual'è la probabilità di fare testa pescando 1 moneta tra queste 100?  $\frac{3}{4}$

esperimento:

esempio

11 R  
12 B

13  
8 R  
5 B

10  
3 R  
7 B

1) Si sceglie a caso un contenitore

2) Si estrae a caso una pallina dal contenitore scelto

- probabilità che una pallina estratta sia rossa

- probabilità che la pallina estratta provenga dalla seconda scatola, sapendo che è rossa

Le singole palline non sono egualmente probabili

prima soluzione "naive": cambiare il numero di palline in ogni contenitore in modo che

- le palline diventino equiprobabili

- le proporzioni di colori nei due contenitori non cambiano

X CASA

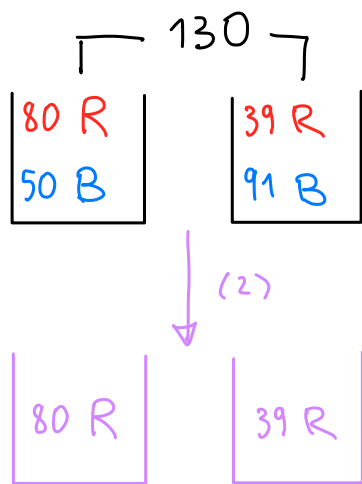
Lezione 3  
03-03-21  
(Rovito)

Soluzione "naive":  $\text{mcm}[13, 10] = 130$

1)  $8 : 13 = x : 130$   $x = \frac{8 \cdot 130}{13} = 80 \rightarrow 50$  palline blu

2)  $3 : 10 = x : 130$   $x = \frac{3 \cdot 130}{10} = 39 \rightarrow 91$  palline blu

Adesso abbiamo:



$$1. \frac{80+39}{260} = \frac{119}{260}$$

$$2. \frac{39}{80+39} = \frac{39}{119}$$

Abbiamo "buttato via"  
il caso che la pallina  
estratta sia blu

Probabilità condizionata:

$(\Omega, P)$   $A, B \subset \Omega$  e supponiamo  $P(B) \neq \emptyset$

probabilità di A  
sapendo che B è  
accaduto

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

$(\Omega, P(\cdot | B))$  nuovo spazio di  
probabilità

$$P(B^c | B) = 0$$

Esempi:

$$P[3 \text{ in un lancio di un dado}] = \frac{1}{6}$$

$$P[\text{dado} = 3 \mid \text{esce un numero dispari}] = \frac{1}{3}$$

$$P[\text{dado} = 3 \mid \text{" " " pari}] = 0$$

osservazione che succede se la probabilità (da non  
condizionata a condizionata) non cambia?

$$\text{se } P(A|B) = P(A) \text{ allora } P(A \cap B) = P(A)P(B)$$

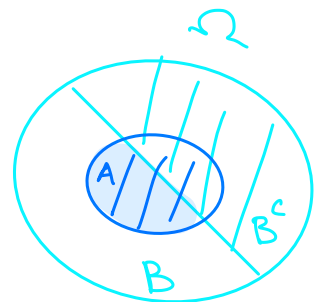
infatti  $A, B$  indipendenti (e non banali)  $\Leftrightarrow P(A|B) = P(A) \Leftrightarrow P(B|A) = P(B)$

2. 

gettoni
1, ..., 90

 due estrazioni

- con reinserimento: indipendenti
- senza reinserimento: dipendenti



CON  $P_c[\text{prima} = 90, \text{seconda} = 90] = \frac{1}{90^2} = P[\text{prima} = 90] P[\text{seconda} = 90]$

$\Omega_c = \{(a, b) : 1 \leq a, b \leq 90\}$  probabilità uniforme  $\# \Omega_c = 90^2$

SENZA  $P_s[\text{prima} = 90, \text{seconda} = 90] = 0$

$\Omega_s = \{(a, b) : 1 \leq a, b \leq 90, a \neq b\}$  probabilità uniforme  $\# \Omega_s = 90 \cdot 89$

$k = 1, \dots, 90$

$P_c[\text{prima estrazione si estrae } k] = 1/90$

$P_s[\text{" " " " "}] = 1/90$

$P_c[\text{seconda estrazione si estrae } k] = 90/90^2 = 1/90$

$P_s[\text{" " " " "}] = \frac{89}{90 \cdot 89} = \frac{1}{90}$

$P_s[\text{gettone } k \text{ estratto alla } 2^{\text{a}} \text{ estrazione} \mid \text{non è stato estratto } 1^{\text{a}}] = \frac{1}{89}$

$= \frac{P_s[\text{gettone } k \text{ alla } 2^{\text{a}} \cap \text{non alla } 1^{\text{a}}]}{P[k \text{ non alla } 1^{\text{a}}]} = \frac{\frac{89}{90 \cdot 89}}{\frac{89}{90}} = \frac{1}{89}$

$P_s[k \text{ alla } 2^{\text{a}} \text{ estrazione}] = P_s[k \text{ alla } 2^{\text{a}} \text{ estr.} \mid \text{non alla } 1^{\text{a}}] P[\text{non alla } 1^{\text{a}}]$

proposizione  $A_1 A_2 \dots A_n$  eventi tali che  $A_1 \cap A_2 \cap \dots \cap A_n$  sia non trascurabile

allora  $P[A_1 \cap A_2 \cap \dots \cap A_n] = P(A_1) P(A_2 \mid A_1) P(A_3 \mid A_1 \cap A_2) \dots P(A_n \mid A_1 \cap \dots \cap A_{n-1})$

dimi per ogni  $j$ ,  $A_1 \cap A_2 \cap \dots \cap A_j$  è non trascurabile, perché

contiene  $A_1 \cap \dots \cap A_n$

$P(A_1) P(A_2 \mid A_1) \dots P(A_n \mid A_1 \cap \dots \cap A_{n-1}) = P(A_1) \frac{P(A_1 \cap A_2)}{P(A_1)} \dots \frac{P(A_1 \cap \dots \cap A_n)}{P(A_1 \cap \dots \cap A_{n-1})}$

□

Riprendiamo il nostro problema:

esperimento:

13	10
8R	3R
5B	7B

1) Si sceglie a caso un contenitore

2) Si estrae a caso una pallina dal contenitore scelto

Calcolare la:

- probabilità che una pallina estratta sia rossa
- probabilità che la pallina estratta provenga dalla seconda scatola, sapendo che è rossa

### Formule di disintegrazione

$(\Omega, \mathbb{P})$   $A, B \subset \Omega$  tali che  $P(B) \neq 0$ ,  $P(B^c) \neq 0$

$A = (A \cap B) \cup (A \cap B^c)$  unione disgiunta

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B^c) & P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= P(A|B)P(B) + P(A|B^c)P(B^c) \end{aligned}$$

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

$A = \{ \text{pallina rossa estratta} \}$

$B = \{ \text{pallina rossa estratta dal 1° contenitore} \}$

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

$$= \frac{8}{13} \cdot \frac{1}{2} + \frac{3}{10} \cdot \frac{1}{2} = \frac{119}{260}$$

### formule di disintegrazione

Sia  $(B_i)_{i=1, \dots, n}$  una partizione di  $\Omega$  in eventi non trascurabili

$$\left. \begin{aligned} &\bullet (B_i)_{i=1, \dots, n} \text{ a due a due disgiunti} \\ &\bullet \bigcup_{i=1}^n B_i = \Omega \\ &\bullet P(B_i) > 0 \text{ per ogni } i=1, \dots, n \end{aligned} \right\} \begin{aligned} &\text{allora per ogni } A \subset \Omega \\ &P(A) = \sum_{i=1}^n P(A|B_i)P(B_i) \end{aligned}$$

dim  $A = A \cap \Omega = A \cap \left( \bigcup_{i=1}^n B_i \right) = \bigcup_{i=1}^n (A \cap B_i)$  unioni disgiunte

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

□

Formule di Bayes  $A, B \subset \Omega$  non trascurabili

$$P[A|B] = P[B|A] \frac{P[A]}{P[B]}$$

$P[\text{palline provengono dal secondo contenitore} \mid \text{estratta rossa}] =$

A
B

$$= P[B|A] \frac{P[A]}{P[B]} = \frac{3}{10} \cdot \frac{1}{2} \cdot \frac{260}{119} = \frac{39}{119}$$

$\downarrow$   $\frac{1}{2}$   $\rightarrow \frac{119}{260}$

In generale, data una partizione  $(B_i)_{i=1, \dots, n}$

$$P(B_i | A) = \frac{P(A | B_i) P(B_i)}{\sum_{j=1}^n P(A | B_j) P(B_j)}$$

Infatti

$$P(B_i | A) = \frac{P(B_i) P(A | B_i)}{P(A)} = \frac{P(A | B_i) P(B_i)}{\sum_{j=1}^n P(A | B_j) P(B_j)}$$

ESPERIMENTO A PROVE RIPETUTE INDIPENDENTI  
(SCHEMA DI BERNOULLI)

- Ripetizione di prove indipendenti
- Ogni prova ha due soli possibili esiti: successo/insuccesso
- Successo avviene con probabilità  $p \in [0, 1]$

Il numero di successi  $S_n \in \{0, \dots, n\}$

$$\{K \text{ successi su } n \text{ prove}\} = \bigcup_{\substack{J \subset \{1, \dots, n\} \\ \#J = K}} \left\{ \begin{array}{l} \text{Successo alle prove con indice} \\ \text{in } J, \text{ insuccesso nelle prove con} \\ \text{indice fuori da } J \end{array} \right\}$$

$$P[K \text{ successi su } n \text{ prove}] = \sum_{\substack{J \subset \{1, \dots, n\} \\ \#J = K}} P[\text{successo in } J, \text{ insuccesso in } J^c]$$

in quanto

$$= \sum_{\substack{J \subset \{1, \dots, n\} \\ \#J = K}} P^K (1-p)^{n-K} = \binom{n}{K} P^K (1-p)^{n-K}$$

$$\{ \text{successi in } J, \text{ insuccessi in } J^c \} = \prod_{j \in J} \{ \text{successo alla } j\text{-esima prova} \} \cdot$$

$$\cdot \prod_{j \in J^c} \{ \text{insuccesso alla } j\text{-esima prova} \}$$

Per indipendenza

$$P(\text{successi in } J, \text{ insuccessi in } J^c) = \prod_{j \in J} P(\text{successo alla } j\text{-esima}) \cdot \prod_{j \in J^c} P(\text{insuccesso alla } j\text{-esima}) \\ = p^{\#J} (1-p)^{\#J^c} = p^k (1-p)^{n-k}$$

$$P(k \text{ successi su } n \text{ prove}) = P_p(k, n)$$

Se  $p = \frac{1}{2}$   $n$  pari

$$P_{1/2}\left(\frac{n}{2}, n\right) = \binom{n}{n/2} \frac{1}{2^n} = \frac{n!}{\left(\frac{n}{2}\right)! \left(\frac{n}{2}\right)!} \frac{1}{2^n}$$

Formula di Stirling  $n! = n^n e^{-n} \sqrt{2\pi n} \left(1 + O\left(\frac{1}{n}\right)\right)$

$$P_{1/2}\left(\frac{n}{2}, n\right) \approx \frac{n^n e^{-n} \sqrt{2\pi n}}{\left(\frac{n}{2}\right)^n e^{-n} \pi n} \frac{1}{2^n} \approx \sqrt{\frac{2}{\pi n}} \rightarrow 0$$

Inoltre per  $k=0, \dots, n$   $P_{1/2}(k, n) \leq P_{1/2}\left(\frac{n}{2}, n\right) \rightarrow 0$

Studieremo  $P_p(k, n)$   $p \in (0, 1)$

### ANCORA SUL PROBLEMA INIZIALE

$$\begin{array}{ll} R_1^1 \dots R_8^1 & \cancel{B_1^1 \dots B_5^1} \\ R_1^2 \dots R_3^2 & \cancel{B_1^2 \dots B_7^2} \end{array} \quad \begin{array}{ll} | P(R_i^1) = \frac{1}{26} & P(B_j^1) = \frac{1}{26} \\ | P(R_i^2) = \frac{1}{20} & P(B_j^2) = \frac{1}{20} \end{array}$$

$$P[\text{rossa}] = P[\{R_1^1 \dots R_8^1, R_1^2 \dots R_3^2\}] = \frac{8}{26} + \frac{3}{20} = \frac{119}{260}$$

$$P[2^\circ \text{ contenitore} | \text{rossa}] = \frac{P[\{R_1^2 \dots R_3^2\}]}{P[R_1^1 \dots R_8^1, R_1^2 \dots R_3^2]} = \frac{3/20}{119/260} = \frac{39}{119}$$

Esercitazione 2  
08-03-2021  
(Di Gesù)

### Problema 1:

Siano  $A, B, C, D$  quattro eventi tali che

- $P(A) = 1/2$
- $P(A \cap B \cap D) = 1/4$
- $P(A \cap B \cap C \cap D) = 1/9$

(i) Dimostrare che le ipotesi fatte non sono in contraddizione

(ii) calcolare, se possibile,  $P(A \cap [(B \cap D)^c \cup C])$

(iii)  $P(A \cap [(B \cap D)^c \cup C]) = P(A) - P(A \cap [(B \cap D)^c \cup C]^c) =$

$$\stackrel{\text{De Morgan}}{=} \frac{13}{36}$$

(i) Cerchiamo  $(\Omega, \mathcal{F}, \mathbb{P})$

$\Omega = \{1, 2, \dots, 36\}$ ,  $\mathbb{P}$  uniforme

$A = \{1, \dots, 18\}$

$D = \{10, \dots, 27\}$

$B = \{1, \dots, 36\}$

$C = \{1, \dots, 4\} \cup \{10, \dots, 13\} \cup \{19, \dots, 22\} \cup \{28, \dots, 31\}$

## Problema 2

$n+2$  persone in fila

insieme delle persone in fila  $\longleftrightarrow \{1, \dots, n+2\}$

$M \rightarrow n+1$

$F \rightarrow n+2$

$\Omega = S_{n+2}$  = permutazioni dell'insieme  $\{1, \dots, n+2\}$

$|\Omega| = (n+2)!$  (è noto ma si dimostra per induzione)

Assumiamo la probabilità uniforme su  $\Omega$

Chiamo  $A_k \subset \Omega$  l'insieme che corrisponde all'evento "M e F sono separati da esattamente k persone"

$$|A_k| = ? \quad \left( \mathbb{P}(A_k) = \frac{|A_k|}{|\Omega|} \right)$$

un elemento  $\omega \in A_k$  è caratterizzato dalle seguenti domande:

1) Chi fra M e F viene prima?  $\rightarrow$  2 possibilità

2) in che posizione è il primo amico? ci sono  $n+1-k$  possibilità

3) in che ordine sono disposte le altre  $n$  persone?  $n!$  possibilità

$$|A_k| = 2(n+1-k)n$$

$$P(A_n) = \frac{2(n+1-k) \cancel{n!}}{(\cancel{n+2})! (n+2)(n+1)} = \frac{2(n+1-k)}{(n+1)(n+2)} \quad \forall k = 0, \dots, n$$

$$\Gamma_{\text{oss}}: A_K = \{ \omega \in S_{n+2} : |\omega(n+1) - \omega(n+2)| = K+1 \}$$

check: è vero che  $\sum_{k=0}^n P(A_k) = 1$ ?

$$(ii) \quad A_3 \cup A_4 \cup \dots \cup A_n = \bigcup_{k=3}^n A_k$$

$$P\left(\bigcup_{k=3}^n A_k\right) = 1 - P(A_0 \cup A_1 \cup A_2) = 1 - (P(A_0) + P(A_1) + P(A_2)) =$$

$$= \dots = \frac{(n-1)(n-2)}{(n+1)(n+2)} = \frac{n^2 - 3n + 2}{n^2 + 3n + 2}$$

osservazione:  $P_n \left( \bigcup_{k=3}^n A_k \right) \xrightarrow{n \rightarrow +\infty} 1$

Problem 3

$(\Omega_n, \mathcal{F}_n, P_n)$

$B_n \subset \mathcal{F}_n \quad \forall n \quad P_n(B_n) \xrightarrow{n \rightarrow \infty} 1$

### Problema 3

Si estraggono due numeri  $a$  e  $b$  da una scatola contenente

$n$  palline numerate da 1 a  $n$ . Calcolare la probabilità

che  $|a-b| = 1$  nell'ipotesi che le estrazioni vengano effettuate:

(i) senza reinserimento

(ii) con reinserimento

(i)  $(n \geq 2)$

$$\Omega = D_{n,2} = \{ f : \{1,2\} \rightarrow \{1,\dots,n\} \text{ injective} \}$$

disposizioni semplici di due elementi  
estratte da un insieme di  $n$  elementi

(  $f(1)$  rappresenta il numero della prima pallina estratta  
 $f(2)$  " " " " seconda " " ecc... )

Assumiamo la probabilità uniforme

$$A = \{ f \in D_{n,2} : |f(1) - f(2)| = 1 \}$$

$$|A| = ?$$

1) Esce prima il numero più piccolo o quello più grande? 2 possibilità

2) Qual'è il numero della prima pallina estratta?  $n-1$  possibilità

$$|A| = 2(n-1)$$

$$\Rightarrow P(A) = \frac{2(n-1)}{n(n-1)} = \frac{2}{n}$$

$$(|D_{n,2}| = \frac{n!}{(n-2)!} = n(n-1))$$

perché posso pescare gli elementi in  $\binom{n}{2}$  modi ma poiché l'ordine di estrazione è influente devo moltiplicare per  $2! \rightarrow \frac{2!n!}{2!(n-2)!}$

$$(ii) \Omega = \{ f : \{1, 2\} \rightarrow \{1, \dots, n\} \} = D_{n,2}^*$$

$$B = \{ f \in D_{n,2}^* : |f(1) - f(2)| = 1 \} = \{ f \in D_{n,2} : |f(1) - f(2)| = 1 \} = A$$

$$|\Omega| = n^2$$

$$P(B) = \frac{|B|}{|\Omega|} = \frac{2(n-1)}{n^2} = \frac{2}{n} - \frac{2}{n^2}$$

#### Problema 4

$\Omega = \{0, 1\}^n \times \{0, 1\}^n$ , probabilità uniforme

giocatore  $(\Omega = \{ \omega_1^{(a)}, \dots, \omega_n^{(a)}, \omega_1^{(b)}, \dots, \omega_n^{(b)} \})$

$G = \text{il giocatore vince} = \{ \omega \in \Omega : \sum_{k=1}^n \omega_k^{(a)} > \sum_{k=1}^n \omega_k^{(b)} \}$

Soluzione  $P(G) = \frac{1}{2} - \frac{1}{2^{2n+2}} \binom{2n}{n} \underset{n=10}{\simeq} 0,4119$

$$\Omega = G \cup B \cup N$$

$$\Rightarrow 1 = P(G) + P(B) + P(N)$$

$$P(G)$$

$$2 P(G) = 1 - P(N) \Rightarrow P(G) = \frac{1}{2} - \frac{1}{2} P(N)$$

### Problema 7

$$P(A) = \frac{1}{11 \cdot 10 \cdot 9 \cdot 7 \cdot 6 \cdot 2} = \frac{3 \cdot 4 \cdot 5 \cdot 8}{11!}$$

### Problema 8

$$\Omega = \{0, \dots, 9\}^n, \quad |\Omega| = 10^n$$

$N_k$  = "non appare la cifra k"

$$(i) \quad P(N_3) = \left(\frac{9}{10}\right)^n$$

$$(ii) \quad P(N_4 \cap N_7) = \left(\frac{8}{10}\right)^n$$

$$(iii) \quad P(N_5^c) = 1 - P(N_5) = 1 - \left(\frac{9}{10}\right)^n$$

(iv)  $C_{3,4}$  = appare il 3 prima del 4

Esempio:  $n=1 \quad P(C_{3,4}) = 0$

$$n=2 \quad C_{3,4} = \{(3,4)\} \Rightarrow P(C_{3,4}) = \frac{1}{100}$$

$$n=3 \quad \begin{array}{ccc} 3 & 4 & \square \end{array} \rightarrow 10 \text{ casi}$$

$$\begin{array}{ccc} 3 & \square & 4 \end{array} \rightarrow 9 \text{ casi}$$

$$\begin{array}{ccc} \square & 3 & 4 \end{array} \rightarrow 8 \text{ casi}$$

---

$$\text{Totale} \quad 27 \text{ casi}$$

$$P(C_{3,4}) = \frac{27}{1000}$$

Caso generale:  $\Omega = C_{3,4} \cup C_{4,3} \cup (N_3 \cup N_4)$

$$\Rightarrow 1 = 2P(C_{3,4}) + \underbrace{P(N_3 \cup N_4)}_{= P(N_3) + P(N_4) - P(N_3 \cap N_4)}$$

$$\Rightarrow 1 = 2P(C_{3,4}) + 2\left(\frac{9}{10}\right)^n - \left(\frac{8}{10}\right)^n$$

$$P(C_{3,4}) = \frac{1}{2} - \left(\frac{9}{10}\right)^n + \frac{1}{2} \left(\frac{8}{10}\right)^n$$

# ESPERIMENTI E PROVE RIPETUTE INDIPENDENTI

Lezione 4  
09-03-2021  
(Romito)

$n$  prove

$p$ : probabilità di successo

$$P[K \text{ successi su } n \text{ prove}] = \binom{n}{k} p^k (1-p)^{n-k}$$

distribuzione  
binomiale

$$P_p(k, n) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, \dots, n$$

$$P_{1/2}\left(\frac{n}{2}, n\right) \sim \frac{1}{\sqrt{n}}$$

proposizione  $p \in (0, 1)$  *parte intera*

1. se  $k \leq [(n+1)p] - 1$  allora

$$P_p(k+1, n) \geq P_p(k, n)$$

2. se  $k \geq [(n+1)p]$  allora

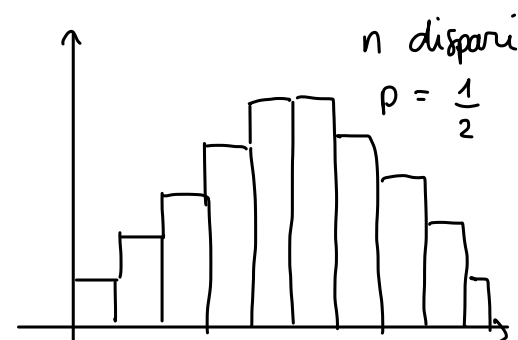
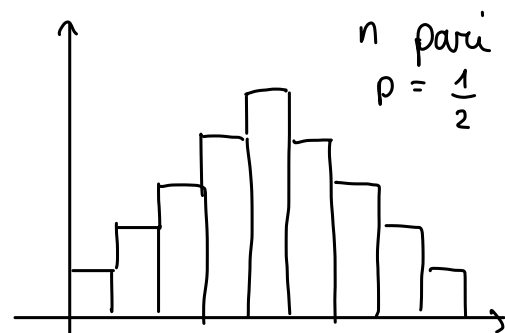
$$P_p(k+1, n) \leq P_p(k, n)$$

3. la probabilità massima è  
assunta per

•  $k = [(n+1)p]$  se  $(n+1)p$  non è intero

•  $k = (n+1)p$ ,  $(n+1)p + 1$  altrimenti

$$4. P_p(k, n) \approx \frac{1}{\sqrt{n}}$$



dimostrazione

$$\frac{P_p(k+1, n)}{P_p(k, n)} = \frac{\binom{n}{k+1} p^{k+1} (1-p)^{n-k-1}}{\binom{n}{k} p^k (1-p)^{n-k}} = \frac{n-k}{k+1} \frac{p}{1-p}$$

$$\frac{P_p(k+1, n)}{P_p(k, n)} \leq 1 \rightarrow (n-k)p \leq (k+1)(1-p) \rightarrow np \leq 1-p+k$$

$$\rightarrow k \geq (n+1)p - 1$$

$$\geq 1 \rightarrow k \leq (n+1)p - 1$$

Per l'ultima

$k_0$ : numero di prove che corrisponde al max

$$P_p(k_0, n) = \binom{n}{k_0} p^{k_0} (1-p)^{n-k_0} = \frac{n!}{k_0! (n-k_0)!}$$

$$\left[ \text{Stirling: } m! \sim m^m e^{-m} \sqrt{2\pi m} = \dots + \left(1 + O\left(\frac{1}{m}\right)\right) \right]$$

$$\frac{n!}{k_0(n-k_0)!} \sim \frac{n^n e^{-n} \sqrt{2\pi n}}{k_0^{k_0} e^{-k_0} \sqrt{2\pi k_0} (n-k_0)^{n-k_0} e^{-(n-k_0)} \sqrt{2\pi(n-k_0)}} \sim \frac{1}{\sqrt{n}}$$

$$\left. \begin{array}{l} k_0 \sim n \\ n-k_0 \sim n \end{array} \right\} \sim \frac{1}{\sqrt{n}} \frac{n^n}{k_0^{k_0} (n-k_0)^{n-k_0}}$$

$$P_P(k_0, n) \sim \frac{1}{\sqrt{n}} n^n \frac{p^{k_0}}{k_0^{k_0}} \frac{(1-p)^{n-k_0}}{(n-k_0)^{n-k_0}}$$

$$\frac{p^{k_0}}{k_0^{k_0}} = \frac{p^{k_0}}{[(n+1)p]^{k_0}} = \left( \frac{p}{(n+1)p} \right)^{k_0}$$

$$\frac{x}{[x]} \rightarrow 1 \text{ se } x \rightarrow \infty$$

$$\sim \left( \frac{1}{n+1} \right)^{k_0}$$

$$\frac{(1-p)^{n-k_0}}{(n-k_0)^{n-k_0}} \sim \left( \frac{1-p}{n-(n+1)p} \right)^{n-k_0} \sim \left( \frac{1}{n+1-\frac{1}{1-p}} \right)^{n-k_0}$$

$\parallel$   
 $(1-p)(n+1)-1$

## Teorema di De Moivre - Laplace

Sia  $p \in (0, 1)$ , detto  $S_n$  il numero di successi in  $n$  prove indipendenti ripetute con probabilità di successo  $p$ , si ha,

per ogni  $a, b \in \mathbb{R}$ ,  $a < b$

$$P\left[ \underline{np} + a \sqrt{np(1-p)} \leq S_n \leq \underline{np} + b \sqrt{np(1-p)} \right] \xrightarrow{n \rightarrow \infty} \int_a^b \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}}_{\text{funzione Gaussiana}} dy$$

Semplificazioni:

dim nel caso  $p = \frac{1}{2}$   $n = 2m$   $(a, b) \rightarrow (0, a)$

$$P[ np + a \dots \leq S_n \leq np + b \dots ] =$$

$$= P[ np + a \dots \leq S_n < np ] + P[ 0 \leq S_n \leq np + b \dots ]$$

faremo vedere che

$$P[ np \leq S_n \leq np + a \sqrt{np(1-p)} ] \rightarrow \int_0^a \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$



premessa:  $1+x = e^{x+R(x)}$   $|R(x)| \leq x^2$   $|x| \leq \frac{1}{2}$

infatti  $\log(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \dots = x + R(x)$

$$|R(x)| \leq \frac{1}{2}x^2 + \frac{1}{3}|x^3| + \frac{1}{4}x^4 + \dots \leq \frac{1}{2} \sum_{k=2}^{\infty} |x|^k$$

$$= \frac{1}{2} \frac{x^2}{1-|x|} \leq x^2$$

Ora

$$P\left[m \leq S_n \leq m + a \frac{\sqrt{n}}{2}\right] = \sum_{k=m}^{m+a \frac{\sqrt{n}}{2}} p_{\frac{1}{2}}(k, n)$$

prendiamo  $0 \leq k \leq a \frac{\sqrt{n}}{2}$

$$p_{\frac{1}{2}}(m+k, n) = \binom{2m}{m+k} \frac{1}{2^{2m}} = \frac{n-(m+k)+1}{m+k} \binom{2m}{m+k-1} \frac{1}{2^m}$$

$$= \frac{m-k+1}{m+k} p_{\frac{1}{2}}(m+(k-1), n) = \dots$$

$$= \frac{(m-k+1)(m-k+2) \dots (m-k+k)}{(m+k)(m+k-1) \dots (m+k-(k-1))} p_{\frac{1}{2}}(m, n)$$

$$= \frac{\left(1 - \frac{k-1}{m}\right) \left(1 - \frac{k-2}{m}\right) \dots \left(1 - \frac{0}{m}\right)}{\left(1 + \frac{k}{m}\right) \left(1 + \frac{k-1}{m}\right) \dots \left(1 + \frac{1}{m}\right)} p_{\frac{1}{2}}(m, n)$$

$$= \frac{\textcircled{1} e^{-\frac{k-1}{m}} \dots e^{-\frac{0}{m}} e^{R(-\frac{k-1}{m}) + \dots + R(\frac{0}{m})}}{\textcircled{2} e^{\frac{k}{m}} \dots e^{\frac{1}{m}} e^{R(\frac{k}{m}) + \dots + R(\frac{1}{m})}} p_{\frac{1}{2}}(m, n)$$

$$\textcircled{1} e^{-\left(\frac{1}{m} + \dots + \frac{k-1}{m}\right)} = e^{-\frac{k(k-1)}{2m}} \quad \textcircled{1} = e^{-\frac{1}{2m}(K^2+K+K^2-K)} = e^{-\frac{K^2}{m}}$$

$$\textcircled{2} e^{\frac{1}{m} + \dots + \frac{k}{m}} = e^{\frac{K(K+1)}{2m}}$$

$$\textcircled{3} e^{R(-\frac{0}{m}) + \dots + R(-\frac{k-1}{m}) - (R(\frac{1}{m}) + \dots + R(\frac{k}{m}))} = e^{E_k}$$

$$|E_k| = \left| R(-\frac{0}{m}) + \dots + R(-\frac{k-1}{m}) - R(\frac{1}{m}) + \dots - R(\frac{k}{m}) \right|$$

$$\frac{K}{m} \leq \frac{a\sqrt{n}}{2n} = \frac{a}{2\sqrt{n}} \leq \frac{1}{2} \text{ per } n \text{ abbastanza grande}$$

$$|E_k| \leq \left(\frac{1}{m}\right)^2 + \dots + \left(\frac{k-1}{m}\right)^2 + \left(\frac{1}{m}\right)^2 + \dots + \left(\frac{k}{m}\right)^2 \leq 2 \sum_{j=1}^K \left(\frac{j}{m}\right)^2$$

$$\leq 2 \frac{k^3}{m^2} \leq 2 \frac{a^3 n^{\frac{3}{2}}}{8 \frac{n^2}{4}} = \frac{a^3}{\sqrt{n}} \quad k \leq \frac{a}{2} \sqrt{n}$$

quindi  $e^{\epsilon_k} \rightarrow 1$

$$\text{Dunque } P_{\frac{1}{2}}(m+k, n) = e^{-\frac{k^2}{m}} e^{\epsilon_k} P_{\frac{1}{2}}(m, n) \quad 0 \leq k \leq \frac{a}{2} \sqrt{n}$$

Dalla formula di Stirling

$$P_{\frac{1}{2}}(m, n) = \binom{2m}{m} \frac{1}{2^{2m}} = \frac{(2m)!}{m! m! 2^{2m}} = \frac{(2m)^{2m} e^{-2m} \sqrt{4\pi m} (1 + O(\frac{1}{m}))}{2^{2m} m^{2m} e^{-2m} (2\pi m) (1 + O(\frac{1}{m}))}$$

$$= \frac{1}{\sqrt{\pi m}} (1 + O(\frac{1}{m}))$$

Ora

$$\sum_{k=0}^{\frac{1}{2}a\sqrt{n}} P_{\frac{1}{2}}(m+k, n) = \frac{1}{\sqrt{\pi m}} (1 + O(\frac{1}{m})) \sum_{k=0}^{\frac{1}{2}a\sqrt{n}} e^{-\frac{k^2}{m}} e^{\epsilon_k}$$

$$\cdot |\epsilon_k| \leq \frac{a^3}{\sqrt{n}}$$

$$\cdot \int_k^{k+1} e^{-\frac{x^2}{m}} dx \leq e^{-\frac{k^2}{m}} \leq \int_{k-1}^k e^{-\frac{x^2}{m}} dx$$

Si ha

$$\sum_{k=0}^{\frac{1}{2}a\sqrt{n}} P_{\frac{1}{2}}(m+k, n) \leq \frac{1}{\sqrt{\pi m}} (1 + O(\frac{1}{m})) e^{\frac{a^3}{\sqrt{n}}} \sum_{k=0}^{\frac{1}{2}a\sqrt{n}} e^{-\frac{k^2}{m}}$$

$$\leq \frac{1}{\sqrt{\pi m}} (1 + O(\frac{1}{m})) e^{\frac{a^3}{\sqrt{n}}} \sum_{k=0}^{\frac{1}{2}a\sqrt{n}} \int_{k-1}^k e^{-\frac{x^2}{m}} dx$$

$$\leq \frac{1}{\sqrt{\pi m}} \underbrace{(1 + O(\frac{1}{m})) e^{\frac{a^3}{\sqrt{n}}}}_{\rightarrow 1} \int_{-1}^{\frac{1}{2}a\sqrt{n}} e^{-\frac{x^2}{m}} dx$$

$$\frac{1}{\sqrt{\pi m}} \int_{-1}^{\frac{1}{2}a\sqrt{n}} e^{-\frac{x^2}{m}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{\frac{2}{m}}}^a e^{-\frac{1}{2}y^2} dy \xrightarrow{n \rightarrow \infty} \int_0^a \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy$$

$$\text{Cambio di variabile } y = x\sqrt{\frac{2}{m}} \quad dx = \sqrt{\frac{m}{2}} dy$$

$$\frac{1}{2} a\sqrt{n} \frac{2}{\sqrt{n}} = a$$

La minoration

$$\sum_{k=0}^{\frac{1}{2}a\sqrt{n}} P_{\frac{1}{2}}(m+k, n) \geq \frac{1}{\sqrt{\pi m}} \left(1 + O\left(\frac{1}{m}\right)\right) e^{-\frac{a^2}{4n}} \sum_{k=0}^{\frac{1}{2}a\sqrt{n}} e^{-\frac{k^2}{m}}$$

$$\sum_{k=0}^{\frac{1}{2}a\sqrt{n}} e^{-\frac{k^2}{m}} \geq \sum_{k=0}^{\frac{1}{2}a\sqrt{n}} \int_k^{k+1} e^{-\frac{x^2}{m}} dx \geq \int_0^{\frac{1}{2}a\sqrt{n}} e^{-\frac{x^2}{m}} dx$$

Come prima

$$\frac{1}{\sqrt{\pi m}} \int_0^{\frac{1}{2}a\sqrt{n}} e^{-\frac{x^2}{m}} dx = \frac{1}{\sqrt{2\pi}} \int_0^a e^{-\frac{1}{2}y^2} dy$$

Esercitazione 3  
15-03-2021  
(Di Gesù)

### Problema 1.

Siano A, B, C tre eventi. Consideriamo le due affermazioni H1: "A, B sono indipendenti" e H2: "A, B sono indipendenti condizionatamente a C" (ovvero  $P(A \cap B | C) = P(A | C)P(B | C)$ ).

(i) Vale l'implicazione  $H1 \Rightarrow H2$  ?

(ii) Vale l'implicazione  $H2 \Rightarrow H1$  ?

(iii) Sotto quali ipotesi su C vale  $H1 \Leftrightarrow H2$ ?

i) non vale l'implicazione, controesempio:

Sia  $\Omega = \{1, 2, 3, 4\}$ . Assumo la probabilità uniforme

Sia  $A = \{1, 3\}$ ,  $B = \{2, 3\}$  e  $C = \{3, 4\}$ .

Osservo che A e B sono indipendenti poiché:

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

Allora dovremmo avere:

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

Ma notiamo che  $\frac{P(A \cap B \cap C)}{P(C)} = \frac{1}{4} \cdot \frac{4}{2} = \frac{1}{2}$  mentre

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

concludere che  $H1 \not\Rightarrow H2$ .

ii) non vale l'implicazione

Siano  $\Omega = \{1, \dots, 9\}$ ,  $A = \{1, 2\}$ ,  $B = \{2, 3, 4\}$ ,  $C = \{2, 3\}$

Suppongo la probabilità uniforme.

Notiamo che:

$$\frac{1}{2} = \frac{1}{9} \cdot \frac{9}{2} = \frac{P(A \cap B \cap C)}{P(C)} = \frac{P(A \cap C)}{P(C)} \frac{P(B \cap C)}{P(C)} = \frac{1}{9} \cdot \frac{9}{2} \cdot \frac{2}{9} \cdot \frac{9}{2} = \frac{1}{2}$$

quindi vale  $H_2$ . Quindi dovremmo avere:

$$P(A \cap B) = P(A)P(B) \text{ ma ho } P(A \cap B) = \frac{1}{9} \text{ e}$$

$$P(A)P(B) = \frac{2}{9} \cdot \frac{3}{9} = \frac{2}{27}.$$

Quindi concludo che  $H_2 \not\Rightarrow H_1$ .

iii)  $(\Leftarrow) H_2$

Se  $C$  è indipendente da  $A$  e  $B$  e da  $A \cap B$  otteniamo

$$\begin{aligned} P(A \cap B | C) &= \frac{P(A \cap B \cap C)}{P(C)} \stackrel{\uparrow \text{associatività}}{=} \frac{P((A \cap B) \cap C)}{P(C)} \stackrel{\uparrow \text{indip.}}{=} \frac{P(A \cap B)P(C)}{P(C)} = \\ &\stackrel{ip}{=} \frac{P(A \cap C)}{P(C)} \frac{P(B \cap C)}{P(C)} \stackrel{\uparrow \text{ind.}}{=} \frac{P(A)P(C)}{P(C)} \frac{P(B)P(C)}{P(C)} = P(A)P(B) \end{aligned}$$

Quindi  $A$  e  $B$  sono indipendenti e vale  $H_1$ .

$(\Rightarrow) H_1$

Supponendo le stesse ipotesi del punto precedente:

$$P(A)P(B) \stackrel{ip}{=} P(A \cap B) \Leftrightarrow P(A)P(B)P(C) = P(A \cap B)P(C)$$

$$\stackrel{\text{ind.}}{\Leftrightarrow} P(A)P(B \cap C) = P(A \cap B \cap C) \Leftrightarrow \frac{P(A)P(C)}{P(C)} \frac{P(B \cap C)}{P(C)} = \frac{P(A \cap B \cap C)}{P(C)}$$

$$\stackrel{\text{ind.}}{\Leftrightarrow} \frac{P(A \cap C)}{P(C)} \frac{P(B \cap C)}{P(C)} = \frac{P(A \cap B \cap C)}{P(C)} \stackrel{\text{def}}{\Leftrightarrow} P(A \cap B | C) = P(A | C)P(B | C).$$

□

# TEOREMA DI DE MOIVRE - LAPLACE

Sia  $p \in (0, 1)$ , detto  $S_n$  il numero di successi in  $n$  prove indipendenti ripetute con probabilità di successo  $p$ , si ha, per ogni  $a, b \in \mathbb{R}$ ,  $a < b$

$$P[ np + a \sqrt{np(1-p)} \leq S_n \leq np + b \sqrt{np(1-p)} ] \xrightarrow{n \rightarrow \infty} \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy$$

esempio lanciamo 1000 monete  $p = \frac{1}{2}$

$P[470 \leq S_{1000} \leq 530]$  la probabilità di ottenere tra le 470 e 530 teste

$n = 1000$

$$470 = 1000 \cdot \frac{1}{2} + a \frac{\sqrt{1000}}{2}$$

$$= 500 + 5a\sqrt{10}$$

scrivo 470 e 530 in modo tale che io possa usare DHL

$$5a\sqrt{10} = -30 \longrightarrow a = -\frac{6}{\sqrt{10}} = -1,9$$

$$530 = 1000 \cdot \frac{1}{2} + b \frac{\sqrt{1000}}{2}$$

$$5b\sqrt{10} = 30$$

$$b = 1,9$$

ricordo  $a$  e  $b$

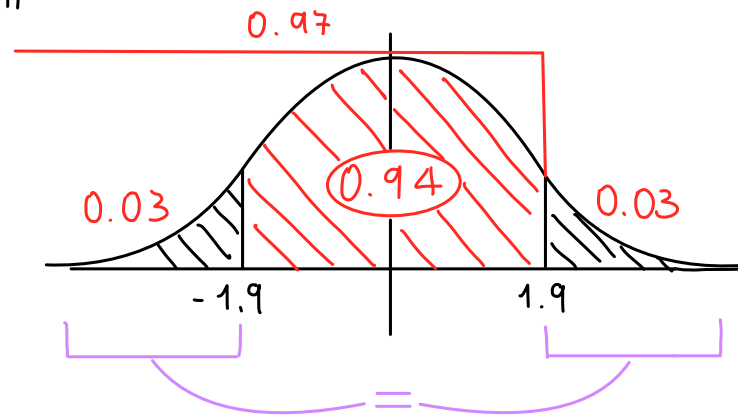
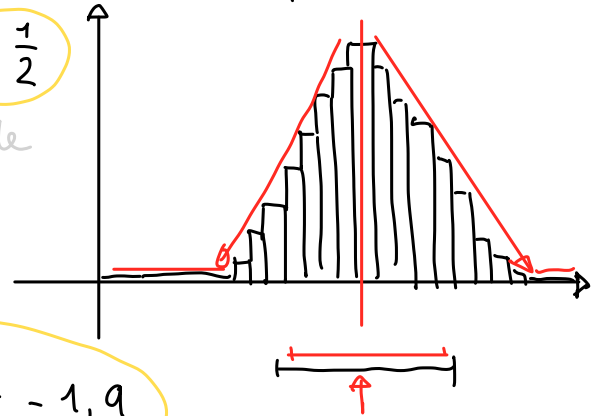
$$P[470 \leq S_{1000} \leq 530] \approx \int_{-1,9}^{1,9} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy = 0.94$$

$$\int_{-\infty}^{1,9} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy = 0.97$$

$$\int_{\mathbb{R}} e^{-\frac{1}{2}y^2} dy = \sqrt{2\pi}$$

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy = 1$$

$$\int_{1,9}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy = 1 - 0.97 = 0.03$$



## LEGGE DEI GRANDI NUMERI (per esperimenti a prove ripetute indipendenti)

Sia  $p \in (0, 1)$ ,  $S_n$  il numero di successi, allora la frequenza dei successi tende a  $p$  per  $n \rightarrow \infty$

$$\text{per ogni } \varepsilon > 0 \quad \mathbb{P} \left[ \left| \frac{S_n}{n} - p \right| \geq \varepsilon \right] \rightarrow 0 \quad \text{per } n \rightarrow \infty$$

(convergenza in probabilità)

dim

utilizziamo De Moivre - Laplace

Per ogni  $c > 0$  e usiamo

DM L con  $a = -c$  e  $b = c$

↳ l'evento: gli esiti per cui il numero di successi è compreso tra  $np - n\varepsilon$  oppure maggiore di  $np + n\varepsilon$

$\Omega = \{0, 1\}^n$  con probabilità

$$p(\omega) = p^{\sum_{i=1}^n w_i} (1-p)^{n - \sum_{i=1}^n w_i}$$

$$\omega \in \Omega \quad \sum_{i=1}^n w_i \quad \left\{ \omega : \left| \frac{\sum_{i=1}^n w_i}{n} - p \right| \geq \varepsilon \right\}$$

$$\mathbb{P} [ np - c\sqrt{np(1-p)} \leq S_n \leq np + c\sqrt{np(1-p)} ] \rightarrow \int_{-c}^c \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy$$

$\varepsilon > 0$  e  $c > 0$ , allora per  $n$  abbastanza grande  $c\sqrt{np(1-p)} < n\varepsilon$

Per questi  $n$

$$\{ np - c\sqrt{np(1-p)} \leq S_n \leq np + c\sqrt{np(1-p)} \} \subset \{ np - n\varepsilon \leq S_n \leq np + n\varepsilon \}$$

allora

$$\mathbb{P} [ np - c\sqrt{np(1-p)} \leq S_n \leq np + c\sqrt{np(1-p)} ] \leq \mathbb{P} [ np - n\varepsilon \leq S_n \leq np + n\varepsilon ]$$

Al limite

$$\int_{-c}^c \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy \leq \liminf_n \mathbb{P} [ np - n\varepsilon \leq S_n \leq np + n\varepsilon ]$$

$$\text{Se ora } c \rightarrow +\infty \quad \int_{-c}^c \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy \rightarrow \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy = 1$$

$$1 \leq \liminf \mathbb{P} [ np - n\varepsilon \leq S_n \leq np + n\varepsilon ]$$

$$\leq \limsup \quad "$$

$$\leq 1$$

$$\text{ovvero } \mathbb{P} [ np - n\varepsilon \leq S_n \leq np + n\varepsilon ] \rightarrow 1$$

Al complementare  $P\left[\left|\frac{S_n}{n} - p\right| > \varepsilon\right] \rightarrow 0$

□

Osservazione la probabilità di non avere neanche un successo dopo  $n$  prove è  $(1-p)^n$  (molto piccola)

In altre parole: se un evento ha probabilità di accadere prima o poi accadrà

———— \* ————

## SPAZI DI PROBABILITÀ DISCRETI

esempio: consideriamo un esperimento a prove ripetute indip. con probabilità di successo  $p \in (0,1)$   
Consideriamo quante prove sono necessarie per avere il primo successo.

$$P[\text{primo successo alla } 47^{\text{a}} \text{ prova}] = (1-p)^{46} p$$

Il modello più naturale per il problema

$$\Omega = \{1, 2, \dots\} = \mathbb{N} \setminus \{0\}$$

$$P[\text{primo successo alla } k\text{-esima prova}] = (1-p)^{k-1} p \quad k \geq 1$$

modello di probabilità discreta

$(\Omega, P)$  dove  $\Omega$  è un insieme (al più) numerabile e

$P$  soddisfa:

- $P(\Omega) = 1$

è importante fissare un modello e verificare che sia coerente

- $0 \leq P(A) \leq 1$  per ogni  $A \subset \Omega$

- se  $(A_n)_{n \in \mathbb{N}}$  sottoinsiemi di  $\Omega$  a due a due disgiunti

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} P(A_n) \quad \sigma\text{-additività}$$

osservazione la additività:  $A \cap B = \emptyset$  allora  $P(A \cup B) = P(A) + P(B)$

•  $\sigma$ -additività  $\rightarrow$  additività:

- $P(\emptyset) = 0$  infatti  $A_n = \emptyset$  per ogni  $n$

$$P(\emptyset) = P\left(\bigcup_{n=0}^{\infty} A_n\right) = \sum_{n=0}^{\infty} P(A_n) = \sum_{n=0}^{\infty} P(\emptyset)$$

quindi  $P(\emptyset) = 0$

- $A \cap B = \emptyset \quad A_0 = A \quad A_1 = B \quad A_n = \emptyset \text{ per } n \geq 2$

$$P(A \cup B) = P\left(\bigcup_{n=0}^{\infty} A_n\right) = \sum_{n=0}^{\infty} P(A_n) < P(A) + P(B)$$

- Esistono funzioni  $P: \mathcal{P}(\Omega) \rightarrow \mathbb{R}$  tali che
  - $P(\Omega) = 1$
  - $0 \leq P(A) \leq 1 \quad A \subset \Omega$
  - se  $A, B \subset \Omega$ ,  $A \cap B = \emptyset$  allora  $P(A \cup B) = P(A) + P(B)$
  - Non vale la  $\sigma$ -additività
- Siccome vale la additività, allora
  - se  $A \subset B$  allora  $P(A) \leq P(B)$
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B) \leftarrow$
  - formula di inclusione-esclusione

### formula di disintegrazione (estensione)

Definiamo una partizione  $(B_n)_{n \geq 1}$  di  $\Omega$  come

- $\bigsqcup_{n=1}^{\infty} B_n = \Omega$

- $(B_n)_{n \geq 1}$  a due a due disgiunti

- $P(B_n) > 0$  per ogni  $n \geq 1$  *tutti gli eventi hanno probabilità di accadere*

Allora per ogni  $A \subset \Omega$

$$P(A) = \sum_{n=1}^{\infty} P(A \mid B_n) P(B_n)$$

esempio  $B_n = \{\text{primo successo alla } n\text{-esima prova}\}$

### INDIPENDENZA

Una famiglia  $(A_j)_{j \in J}$  ( $J$  insieme di indici arbitrario) di eventi

in  $(\Omega, \mathcal{P})$  è costituita da eventi indipendenti se per ogni  $J' \subset J$  finito

$$P\left(\bigcap_{j \in J'} A_j\right) = \prod_{j \in J'} P(A_j)$$

osservazione Ogni sottofamiglia di eventi collettivamente indipendenti è costituita da eventi collettivamente indipendenti

densità discreta  $\Omega$  (al più) numerabile. Una densità discreta è una funzione  $p: \Omega \rightarrow \mathbb{R}^+$ :

- $p(\omega) \geq 0$  per ogni  $\omega \in \Omega$
- $\sum_{\omega \in \Omega} p(\omega) = 1$

osservazione: Potrebbe essere sufficiente chiedere:

- $\tilde{p}(\omega) \geq 0$
- $\sum \hat{p}(\omega) < \infty$

perché in tal caso normalizzando  $p(\omega) = \frac{\tilde{p}(\omega)}{\sum \hat{p}(\omega)}$

teorema Se  $p$  è una densità discreta, allora  $A \subset \Omega$   $P(A) = \sum_{\omega \in A} p(\omega)$  è una prob. (discreta). Viceversa, se  $(\Omega, P)$ , spazio di prob. discreto  $p(\omega) = P[\{\omega\}]$  è una densità discreta.

dim  $P(\Omega) = \sum_{\omega \in \Omega} p(\omega) = 1$  per definizione

$$0 \leq \sum_{\omega \in A} p(\omega) \leq \sum_{\omega \in \Omega} p(\omega) = 1$$

Verifichiamo la  $\sigma$ -additività:  $(A_k)_{k \geq 1}$  a due a due disgiunti,  $A = \bigcup_{k=1}^{\infty} A_k$

$$P(A) = \sum_{\omega \in A} p(\omega) = \sum_{k=1}^{\infty} \left( \sum_{\omega \in A_k} p(\omega) \right) = \sum_{k=1}^{\infty} P(A_k)$$

associatività a blocchi.

Lezione 6  
17-03-2021  
(Romito)

$(\Omega, P)$   $\Omega$  al più numerabile,  $P: \mathcal{P}(\Omega) \rightarrow \mathbb{R}$

- $P(\Omega) = 1$
- $0 \leq P(A) \leq 1$  per ogni  $A \subset \Omega$
- ( $\sigma$ -additività) se  $(A_n)_{n \geq 0} \subset \mathcal{P}(\Omega)$  a due a due disgiunti:  $P\left(\bigcup_{n=0}^{\infty} A_n\right) = \sum P(A_n)$

proposizione (continuità dall'alto e dal basso della probabilità)

Siano dati un insieme numerabile  $\Omega$  e una funzione

$P: \mathcal{P}(\Omega) \rightarrow \mathbb{R}$  t.c.

- $P(\Omega) = 1$

- $0 \leq P(A) \leq 1$  per ogni  $A \subset \Omega$

- $P(A \cup B) = P(A) + P(B)$  se  $A, B \subset \Omega$ ,  $A \cap B = \emptyset$

allora sono equivalenti

1.  $P$  è  $\sigma$ -additiva

← a due a due disgi.

$$(A_n)_{n \geq 0} \subset \mathcal{P}(\Omega) \rightarrow P\left(\bigcup_{n=0}^{\infty} A_n\right) = \sum P(A_n)$$

2. se  $(A_n)_{n \geq 0} \subset \mathcal{P}(\Omega)$ ,  $A_n \subset A_{n+1}$  per ogni  $n \geq 0$  allora

$$P\left(\bigcup_{n=0}^{\infty} A_n\right) = \lim_n P(A_n)$$

3. se  $(A_n)_{n \geq 0} \subset \mathcal{P}(\Omega)$ ,  $A_{n+1} \subset A_n$  per ogni  $n \geq 0$ , allora

$$P\left(\bigcap_{n=0}^{\infty} A_n\right) = \lim_n P(A_n)$$

Commenti: 1. Se vale l'additività finita, allora

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

per  $(A_k)_{1 \leq k \leq n} \subset \mathcal{P}(\Omega)$ , a due a due disgiunti

2. le proprietà 2,3 sono effettivamente un po' più semplici da verificare

dim (1)  $\Rightarrow$  (2) Sia  $(A_n)_{n \geq 0} \subset \mathcal{P}(\Omega)$  crescente per inclusione

$$B_0 = A_0 \quad B_n = A_n \setminus A_{n-1} \quad n \geq 1$$

$(B_n)_{n \geq 0}$  a due a due disgiunti:  $k < n \Rightarrow B_n = A_n \setminus A_{n-1}$

$$B_k = A_k \setminus A_{k-1} \subset A_k \subset A_{n-1} \text{ quindi } A_k \cap A_n = \emptyset$$

Inoltre •  $A_n = \bigcup_{k=0}^n B_k$

•  $\bigcup_{n=0}^{\infty} B_n = \bigcup_{n=0}^{\infty} A_n$

Per la prima  $B_k \subset A_n \subset A_n \rightarrow \bigcup_{k=0}^n B_k \subset A_n$

Viceversa se  $\omega \in A_n$ ,  $\min\{k=0, \dots, n : \omega \in A_k\} = k_0$

Se  $k_0 = 0$   $\omega \in A_0 = B_0$ . Se invece  $k_0 > 0$   $\omega \in A_{k_0}$  ma  $\omega \notin A_{k_0-1} \rightarrow \omega \in B_{k_0}$

In ogni caso  $\omega \in \bigcup_{k=0}^n B_k$

Per la seconda:  $B_n \subset A_n \rightarrow \bigcup_{n=0}^{\infty} B_n \subset \bigcup_{n=0}^{\infty} A_n$

$$A_n = \bigcup_{k=0}^n B_k \subset \bigcup_{k=0}^{\infty} B_k$$

$$P\left(\bigcup_{n=0}^{\infty} A_n\right) = P\left(\bigcup_{n=0}^{\infty} B_n\right) \stackrel{\sigma\text{-add.}}{=} \sum_{n=0}^{\infty} P(B_n) = \lim_{n \rightarrow \infty} \sum_{k=0}^n P(B_k)$$

finita additività  $\Leftrightarrow \lim_{n \rightarrow \infty} P\left(\bigcup_{k=0}^n B_k\right) = \lim_n P(A_n)$

(2)  $\Rightarrow$  (1)

$(A_n)_{n \geq 0} \subset \mathcal{P}(\Omega)$  a due a due disgiunti. Poniamo

$$B_n = \bigcup_{k=0}^n A_k \quad n \geq 0$$

Allora:

- $(B_n)_{n \geq 0}$  crescente per inclusione

- $\bigcup_{n=0}^{\infty} B_n = \bigcup_{n=0}^{\infty} A_n$

$$P\left(\bigcup_{n=0}^{\infty} A_n\right) = P\left(\bigcup_{n=0}^{\infty} B_n\right) \stackrel{\text{ipotesi (2)}}{=} \lim_{n \rightarrow \infty} P(B_n) = \lim_{n \rightarrow \infty} P\left(\bigcup_{k=0}^n A_k\right)$$

finita additività  $\Rightarrow \lim_{n \rightarrow \infty} \sum_{k=0}^n P(A_k) = \sum_{n=0}^{\infty} P(A_n)$

(2)  $\Rightarrow$  (3) La finita additività ci dice che, se  $A \subset \Omega$

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

se  $(A_n)_{n \geq 0} \subset \mathcal{P}(\Omega)$  decrescente per inclusione,  $B_n = A_n^c$

$(B_n)_{n \geq 0}$  crescente per inclusione e  $\left(\bigcap_{n=0}^{\infty} A_n\right)^c = \bigcup_{n=0}^{\infty} B_n$

$$P\left(\bigcap_{n=0}^{\infty} A_n\right) = 1 - P\left(\bigcup_{n=0}^{\infty} B_n\right) \stackrel{(2)}{=} 1 - \lim_n P(B_n) = \lim_n 1 - P(B_n) = \lim_n P(A_n)$$

(3)  $\Rightarrow$  (2) analogo

corollario  $(\Omega, P)$  spazio di probabilità discreto e se  $(A_n)_{n \geq 0} \subset \mathcal{P}(\Omega)$

allora  $P\left(\bigcup_{n=0}^{\infty} A_n\right) \leq \sum_{n=0}^{\infty} P(A_n)$

dice  $B_n = \bigcup_{k=1}^n A_k \quad n \geq 0 \quad (B_n)_{n \geq 0}$  crescente per inclusione e  $\bigcup_{n=0}^{\infty} B_n = \bigcup_{n=0}^{\infty} A_n$

$$P\left(\bigcup_{n=0}^{\infty} A_n\right) = P\left(\bigcup_{n=0}^{\infty} B_n\right) \stackrel{(2)}{=} \lim_n P(B_n) \leq \limsup_n \sum_{k=0}^n P(A_k) = \sum_{n=0}^{\infty} P(A_n)$$

$$P(B_n) = P\left(\bigcup_{k=0}^n A_k\right) \leq \sum_{k=0}^n P(A_k)$$

□

Esempio 1. (distribuzione geometrica)  $\Omega = \mathbb{N} \setminus \{0\}$   $p(k) = p(1-p)^{k-1}$

$$p \in (0,1)$$

$$P \left[ \begin{array}{l} \text{primo successo dopo esattamente} \\ k \text{ prove in un esperimento} \\ \text{a prove ripetute indipendenti} \\ \text{con prob. di successo } p \end{array} \right] = p(k)$$

$p$  densità discreta

- $p(k) \geq 0$  per ogni  $k \geq 1$

- $\sum_{k=1}^{\infty} p(k) = \sum_{k=1}^{\infty} p(1-p)^{k-1} = p \sum_{n=0}^{\infty} (1-p)^n = p \frac{1}{1-(1-p)} = 1$

2. (distribuzione di Poisson)  $\Omega = \mathbb{N}$   $p(k) = \frac{\lambda^k}{k!} e^{-\lambda}$

(esempio: misura la radioattività, persone su un vagone,  
probabilità di infettarsi, ...)

teorema (Legge degli eventi rari)

Consideriamo un esperimento di  $n$  prove ripetute indipendenti  
con probabilità di successo  $p_n$ . Supponiamo che  $\lim_n np_n = \lambda > 0$

Allora per ogni  $k \geq 0$

$$P[K \text{ successi in } n \text{ prove (con prob di successo } p_n)] = \mathcal{P}_{p_n}(k, n) \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}$$

per  $n \rightarrow \infty$

dim Infatti se  $k \geq 0$

$$\mathcal{P}_{p_n}(k, n) = \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{n!}{k!(n-k)!} p_n^k (1-p_n)^{n-k}$$

moltiplico e divido per  $n^k$

$$= \underbrace{\frac{(np_n)^k}{k!}}_{\rightarrow \frac{\lambda^k}{k!}} \underbrace{\frac{n(n-1)\cdots(n-k+1)}{n^k}}_{\rightarrow 1} \underbrace{(1-p_n)^{-k}}_{\rightarrow 1} \underbrace{\left(1 - \frac{np_n}{n}\right)^n}_{\rightarrow e^{-\lambda}} \quad \text{limite notevole}$$

$\leadsto \frac{\lambda^k}{k!} e^{-\lambda}$

□

# VARIABILI ALEATORIE

$(\Omega, \mathbb{P})$  discreto. Una variabile aleatoria è una funzione  $X: \Omega \rightarrow S$  dove  $S$  è un insieme arbitrario (tipicamente si ha  $S = \mathbb{R}, \mathbb{R}^d$ , insieme di simboli)

Ad esempio: prima di lanciare un dado non sappiamo che numero uscirà, lo rappresento allora con una variabile aleatoria:

1. lancio di un dado:  $\Omega = \{1, \dots, 6\}$   $\mathbb{P}$  uniforme  
 $S = \mathbb{R}$   $X: \Omega \rightarrow \mathbb{R}$   $X(\omega) = \omega$

2. lancio di due dadi:  $\Omega = \{1, \dots, 6\}^2$   $\mathbb{P}$  uniforme  
 $S = \mathbb{R}^2$   $X: \Omega \rightarrow \mathbb{R}^2$   $X(\omega) = \omega$

$X$  si potrà scrivere come  $X = (X_1, X_2)$   $\left\{ \begin{array}{l} X_1 \text{ risultato del } 1^\circ \text{ dado} \\ X_2 \text{ risultato del } 2^\circ \text{ dado} \end{array} \right.$

Queste due funzioni sono diverse però "potrebbero essere" la stessa vediamo cosa hanno in comune

## legge (o distribuzione) di una variabile aleatoria

$(\Omega, \mathbb{P})$   $X: \Omega \rightarrow S$ . La legge è la coppia  $(S_X, \mathbb{P}_X)$

dove:

- $S_X = X(\Omega)$  insieme dei valori assunti dalla v.a.

- $\mathbb{P}_X: \mathcal{P}(S_X) \rightarrow \mathbb{R}$  probabilità con cui tali valori sono assunti

$$\mathbb{P}_X(A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) \quad A \subset S_X$$

$$= \mathbb{P}(X^{-1}(A))$$

$$= \mathbb{P}(X \in A)$$

= la controimmagine di  $A$  attraverso  $X$

Le tre variabili di prima hanno la stessa legge  
l'"enfasi" non è sulla funzione ma sulla legge

Osservazione 1.  $(S_X, \mathbb{P}_X)$  spazio di probabilità discreto

2.  $S_X$  può essere implicito, nel senso che possiamo definire

$P_x(A) = P(X^{-1}(A))$   
 $P_x: \mathcal{P}(S) \rightarrow \mathbb{R}$  come  $P_x(A) = P[X^{-1}(A)] = P[X \in A] \quad A \subset \Omega$   
 in quanto  $X^{-1}(A) = X^{-1}(A \cap X(\Omega)) = X^{-1}(A \cap S_x)$

### densità discreta di una variabile aleatoria

$(\Omega, \mathbb{P}) \quad X: \Omega \rightarrow S$ , definiamo la densità discreta  $p_x$  di  $X$   
 come:  $p_x: S_x \rightarrow \mathbb{R} \quad p_x(x) = P[X=x] \quad \left( \begin{aligned} &= P[\{\omega: X(\omega)=x\}] \\ &= P[X^{-1}(\{x\})] \end{aligned} \right)$

1.  $p_x$  è una densità discreta

- $p_x(x) = P[X=x] \geq 0$

- $$\begin{aligned} \sum_{x \in S_x} p_x(x) &= \sum_{x \in S_x} P[X=x] = \sum_{x \in X(\Omega)} P[X^{-1}(\{x\})] \\ &= P\left[\bigcup_{x \in X(\Omega)} X^{-1}(\{x\})\right] = \\ &= P\left[X^{-1}\left(\bigcup_{x \in X(\Omega)} \{x\}\right)\right] = P[X^{-1}(X(\Omega))] \\ &= P(\Omega) = 1 \end{aligned}$$

Ricordiamo:

$\mathcal{P}(S)$  = insieme delle parti di  $S$

2.  $p_x$  è la densità discreta associata alla probabilità  $P_x$

3. Possiamo in realtà pensare  $p_x: S \rightarrow \mathbb{R}$  e semplicemente

$$S_x = \{x \in S: p_x(x) > 0\}$$

### Problema 10.

Un gruppo di  $n$  studenti viene chiamato per un esame orale con una singolare modalità: gli studenti vengono numerati da 1 a  $n$  (ogni studente ha un numero differente). Si procede poi nel modo seguente:

1. si estrae un numero tra gli  $n$  disponibili. Se lo studente corrispondente al numero estratto non ha ancora sostenuto l'esame, allora sarà il candidato successivo;
  2. se invece il numero estratto corrisponde ad uno studente già esaminato, allora l'estrazione si dirà sprecata e si procederà ad estrarre un numero tra i soli numeri non ancora estratti.
- Si ripete poi la modalità illustrata fino a quando tutti gli studenti hanno sostenuto la prova.

Per  $j \in \{1, 2, \dots, n-1\}$ , sia  $A_j$  l'evento

$A_j = \{\text{non si spreca una estrazione per selezionare il } (j+1)\text{-esimo candidato}\}.$

- Calcolare la probabilità di  $A_j$  per ogni  $j=1, 2, \dots, n-1$ .
- Se  $1 \leq j \leq n-2$  dire se  $A_j$  e  $A_{j+1}$  sono indipendenti.
- Se  $1 \leq j \leq n-5$ , dire se  $A_j$  e  $A_{j+4}$  sono indipendenti.
- Determinare tutti i sottoinsiemi  $J \subset \{1, 2, \dots, n-1\}$  tali che gli eventi  $(A_j)$ , con  $j \in J$  siano indipendenti.

$$(i) \quad P(A_j) = \frac{(n-j)}{n} = \frac{\text{"tutti quelli non estratti"}}{\text{"tutti i candidati"}}$$

(ii) + (iii) + (iv) Sono indipendenti

È chiaro che sono indipendenti ma bisogna verificarlo impostando un modello

Possibili modelli:

•  $\Omega = \{1, \dots, n\}^n$  con IP uniforme ✓ (modello ridotto)

•  $\Omega = \{0, 1\}^{n-1}$  con IP condizionata

$$p(\omega) = \left(\frac{n-1}{n}\right)^{\omega_1} \left(\frac{1}{n}\right)^{1-\omega_1} \dots \left(\frac{n-k}{n}\right)^{\omega_k} \left(\frac{k}{n}\right)^{1-\omega_k} \dots \left(\frac{1}{n}\right)^{\omega_{n-1}} \left(\frac{n-1}{n}\right)^{1-\omega_{n-1}}$$

•  $\Omega = S_n \times \{0, 1\}^{n-1}$

"insieme delle permutazioni di  $\{1, \dots, n\}$

$\omega \in \Omega$ ,  $\omega = (\sigma, e)$   $e = (e_1, \dots, e_{n-1})$

$\sigma = (2, 3, 4, 1, 7, \dots)$

**Problema 9.**

Siano dati due esperimenti a prove ripetute indipendenti con probabilità di successo rispettivamente  $p_1$  e  $p_2$  che siano indipendenti tra loro. Qual è la probabilità che il primo successo del primo esperimento avvenga prima del primo successo del secondo?

NB le prove ripetute indipendenti sono infinite

$$\begin{aligned}
 P(A) &= \sum_{k_2=2}^{\infty} \sum_{k_1=1}^{k_2-1} p_1 (1-p_1)^{k_1-1} p_2 (1-p_2)^{k_2-1} = \cancel{p_1} p_2 \sum_{k_2=2}^{\infty} (1-p_2)^{k_2-1} \underbrace{\sum_{k_1=0}^{k_2-2} (1-p_1)^{k_1}}_{\frac{1-(1-p_1)^{k_2-1}}{\cancel{p_1}}} \\
 &= p_2 \sum_{k_2=2}^{\infty} (1-p_2)^{k_2-1} - p_2 \sum_{k_2=2}^{\infty} [(1-p_2)(1-p_1)]^{k_2-2} \\
 &= p_2 \sum_{k_2=1}^{\infty} (1-p_2)^{k_2} - p_2 \sum_{k_2=1}^{\infty} [(1-p_2)(1-p_1)]^{k_2} \\
 &= p_2 \left[ \frac{1}{p_2} - 1 \right] - p_2 \left[ \frac{1}{p_1+p_2-p_1p_2} - 1 \right] \\
 &= 1 - p_2 - \frac{p_2}{p_1+p_2-p_1p_2} + p_2 \\
 &= \frac{p_1 + \cancel{p_2} - p_1p_2 - \cancel{p_2}}{p_1+p_2-p_1p_2} = \frac{p_1(1-p_2)}{p_1+p_2-p_1p_2}
 \end{aligned}$$

**Problema 7** (distribuzione ipergeometrica  $H(k, m, n)$ ).

Consideriamo un esperimento di  $k$  prove ripetute dove ogni prova consiste nel campionare (senza reinserimento) una popolazione di  $n$  oggetti,  $m$  dei quali sono considerati un successo (e  $n-m$  insuccessi). Se  $p_0, p_1$ , etc, sono le probabilità di avere rispettivamente 0 successi, 1 successo, etc,

- Determinare tutti i valori di  $p_i$ , al variare di  $i = 0, 1, \dots, k$ .
- Verificare (direttamente) che
- Calcolare

Confrontare poi i risultati con il caso  $B(k, mn)$ .

$\binom{n}{k}$  è il numero di possibili esiti (non ordinati) dell'esperimento.

Di questi  $\binom{m}{i} \binom{n-m}{k-i}$  presentano  $i$  successi: è il numero di modi di scegliere  $i$  successi da un insieme di  $m$  successi e  $k-i$  insuccessi da un insieme di  $n-m$  insuccessi.

## VARIABILI ALEATORIE

$$(\Omega, \mathbb{P}) \quad S \quad X: \Omega \rightarrow S$$

legge  $S_x = X(\Omega)$

$$P_x(A) = \mathbb{P}[X \in A] (= \mathbb{P}[X^{-1}(A)]) \quad A \subset S$$

$(S_x, P_x)$  spazio di probabilità (discreto)

$$\rightarrow P_x(x) = \mathbb{P}[X = x] \quad x \in S \quad (\text{se } x \in S \setminus S_x \quad P_x(x) = 0)$$

Per ricapitolare:

$$(\Omega, \mathbb{P}) \quad S$$

$$X: \Omega \rightarrow S$$

$$\bullet P_x(x) = \mathbb{P}[X = x] \quad x \in S$$

$$\bullet P_x[A] = \mathbb{P}[X \in A] \quad A \subset S$$

Densità discreta  $p: S \rightarrow \mathbb{R}$

$$p(x) \geq 0$$

$$\sum_{x \in S} p(x) = 1$$

Ad esempio se  $(\Omega, \mathbb{P})$

$$p: \Omega \rightarrow \mathbb{R}$$

$$p(\omega) = \mathbb{P}[\{\omega\}]$$

↳ spazio di probabilità discreta

Problema: Data una "legge" (in termini di probabilità o densità) ad esempio una densità discreta  $p_0: S \rightarrow \mathbb{R}$ , esistono uno spazio di probabilità discreto  $(\Omega, \mathbb{P})$  e su di esso una v.a.  $X: \Omega \rightarrow S$  tale che  $p_0$  è la densità discreta di  $X$ ?

Costruzione canonica:  $p_0: S \rightarrow \mathbb{R}$  densità discreta

Supponiamo  $S$  numerabile (altrimenti  $\{x \in S: p_0(x) > 0\}$  è certamente al più numerabile)

$$\Omega = S$$

$\mathbb{P}$  le misure di probabilità associata a  $p_0$

$$X: \Omega \rightarrow S \quad X(\omega) = \omega \quad \omega \in \Omega$$

$$p_x(x) = P[X=x] = P[\{\omega \in \Omega : X(\omega) = x\}] \\ = P[\{x\}] = p_o(x)$$

esempio lancio di due dadi  $S = \{1, \dots, 6\}^2$   
 $x \in S \quad (x = (x_1, x_2)) \quad p(x) = \frac{1}{36}$

$X$  con legge data da  $p$

$$X: \Omega \rightarrow S \quad X = (X_1, X_2)$$

composizione  $(\Omega, P) \quad X: \Omega \rightarrow S \quad f: S \rightarrow S'$   
 $f(X): \Omega \rightarrow S'$

ad esempio  $X, Y$  v.a. su  $(\Omega, P) \quad X: \Omega \rightarrow S \quad Y: \Omega \rightarrow S'$   
 $(X, Y): \Omega \rightarrow S \times S' \quad f(x, y)$

$$S = S' = \mathbb{R} \quad f(x, y) = x + y \quad X + Y$$

uguaglianza di v.a.  $X: \Omega \rightarrow S \quad Y: \Omega' \rightarrow S \quad (\Omega, P) \text{ spazi } (\Omega', P') \text{ probab.}$

• uguaglianza: se  $\Omega = \Omega', P = P'$  e se  $P[X=Y] = 1$   
 (osservazione: perché non dice  $X(\omega) = Y(\omega) \quad \forall \omega \in \Omega$ ?)

• uguaglianza in legge:  $X, Y$  hanno la stessa legge,  
 ovvero  $(S_X = S_Y) \quad \underline{P_X = P_Y}$  (o equivalentemente  $P_X = P_Y$ )

Osservazioni 1. Se  $X, Y$  sono uguali, allora sono uguali in legge

$$P_X(x) = P[X=x] = P[\{X=x\}] = P[\{X=x\} \cap \{X=Y\} \cup \{X=x\} \cap \{X \neq Y\}]$$

perché  $\{X=Y\} \cup \{X \neq Y\} = \Omega$

$$= P[X=x, X=Y] + \underline{P[X=x, X \neq Y]} = \\ \leq P[X=Y] = 1 - P[X \neq Y] = 0$$

$$\{X=x, X=Y\} = \{\omega: X(\omega) = x, X(\omega) = Y(\omega)\} = \{\omega: Y(\omega) = x, X(\omega) = Y(\omega)\} \\ = \{Y=x, X=Y\}$$

$$= P[X=x, X=Y] = P[Y=x, X=Y] = p_Y(x)$$

2. esistono v.a. con la stessa legge che non sono uguali  
 $\Omega = \{0, 1\}$   $P$ : probabilità uniforme su  $\Omega$   
 $X: \Omega \rightarrow \mathbb{R} \quad X(\omega) = \omega$   
 $Y: \Omega \rightarrow \mathbb{R} \quad Y(\omega) = 1 - \omega$
3. Se  $X, Y$  sono uguali, e se  $f: S \rightarrow S'$ , allora anche  $f(X), f(Y)$  sono uguali
4. Se  $X, Y$  hanno la stessa legge e  $f: S \rightarrow S'$ , allora anche  $f(X), f(Y)$  hanno la stessa legge
- Infatti  $A \subset S' \quad \{f(X) \in A\} = \{X \in f^{-1}(A)\} \quad \searrow$   
 $= \{\omega: f(X(\omega)) \in A\} = \{\omega: X(\omega) \in f^{-1}(A)\}$
- $$P_{f(X)} = P[f(X) \in A] = P[X \in f^{-1}(A)] = P_X[f^{-1}(A)]$$

## ESEMPI

1. Bernoulli (di parametro  $p \in [0, 1]$ )  $X \in \{0, 1\}$

$$P_X(1) = p \quad P_X(0) = 1 - p$$

2. binomiale (di parametri  $n, p$ )  $S_X = \{0, 1, \dots, n\}$

$$\text{se } k \in \{0, 1, \dots, n\} \quad P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

osservazione Se  $X$  è binomiale di parametri  $n, p$ , ci aspettiamo di poter scrivere

$$X \stackrel{\text{legge}}{=} X_1 + \dots + X_n$$

dove  $X_1, X_2, \dots, X_n$  sono Bernoulli di parametro  $p$

osservazione a margine:  $\Omega = \{0, 1, \dots, n\}$   $P(\{\omega\}) = \binom{n}{\omega} p^\omega (1-p)^{n-\omega}$

$X: \Omega \rightarrow \mathbb{R} \quad X(\omega) = \omega$ , allora  $X$  ha legge binomiale

Ma  $\Omega$  è troppo piccolo per costruirvi su  $n$  v.a.

Bernoulli per cui l'uguaglianza  $X = X_1 + \dots + X_n$  valge come uguaglianza di v.a.

3. geometrica di parametro  $p \in [0,1]$   $S_x = \{1, 2, 3, \dots\} = \mathbb{N} \setminus \{0\}$

$$p_x(k) = p(1-p)^{k-1}$$

il primo istante di successo in un esperimento a prove ripetute indipendenti

Mancanza di memoria  $P[X > n+k | X > k] = P[X > n]$

4. Poisson di parametro  $\lambda > 0$  (eventi rari)  $S_x = \mathbb{N}$

$$p_x(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

5. binomiale negativa di parametri  $k \geq 1$ ,  $p \in [0,1]$ ,

numero di prove (senza contare i successi) necessarie per ottenere  $k$  successi (= numeri di insuccessi necessari per ottenere esattamente  $k$  successi)

$$p_x(n) = \binom{n+k-1}{n} p^k (1-p)^n \quad n \geq 1$$

## DISTRIBUZIONI CONGIUNTE

esempio Si lanciano due dadi

$X$ : risultato del primo dado

$Y$ : risultato del secondo dado

$(X, Y)$

$\Omega = \{1, \dots, 6\}$

$P$ : probabilità uniforme

$\omega \in \Omega$   $\omega = (\omega_1, \omega_2)$

$X(\omega) = \omega_1$

$Y(\omega) = \omega_2$

Se però consideriamo quest'altra coppia  $(X, X)$

distribuzione congiunta

$X_1: \Omega \rightarrow S_1$

$\vdots$

$X_d: \Omega \rightarrow S_d$

$X = (X_1, \dots, X_d): \Omega \rightarrow S_1 \times \dots \times S_d$

le leggi marginali non det.  
la legge congiunta

Chiameremo distribuzione congiunta la legge di  $(X_1, \dots, X_d)$

distribuzioni marginali Supponiamo  $X = (X_1, \dots, X_d)$  su  $(\Omega, \mathbb{P})$

con distribuzione congiunta  $\mathbb{P}_X$  (o  $p_X$ )

Chiameremo distribuzioni (o leggi) marginali le leggi delle sottofamiglie di  $(X_1, \dots, X_d)$ , ed in particolare le leggi di ogni  $X_1, \dots, X_d$

Lemma Se  $X = (X_1, \dots, X_d)$  con densità discreta  $p_X$  allora dette  $p_{X_1}, \dots, p_{X_d}$  le densità discrete marginali, si ha

$$\bullet p_{X_1}(x) = \sum_{(x_2, \dots, x_d)} p_X(x, x_2, \dots, x_d)$$

$\vdots$

$$p_{X_d}(x) = \sum_{(x_1, \dots, x_{d-1})} p_X(x_1, \dots, x_{d-1}, x)$$

dimi  $\{X_1 = x\} = \bigcup_{x_2, \dots, x_d} \{X_1 = x, X_2 = x_2, \dots, X_d = x_d\}$

$$\mathbb{P}[X_1 = x] = \sum_{x_2, \dots, x_d} \mathbb{P}[X_1 = x, X_2 = x_2, \dots, X_d = x_d]$$

$$= \sum_{x_2, \dots, x_d} \mathbb{P}[X = (x, x_2, \dots, x_d)]$$

$$= \sum_{x_2, \dots, x_d} p_X(x, x_2, \dots, x_d)$$

ovvero:

$$X: \Omega \rightarrow S_1 \times \dots \times S_d \quad X = (X_1, \dots, X_d)$$

$$S_{X_i} = X_i(\Omega) \subset S_i \quad S_X = X(\Omega) \subset S_1 \times \dots \times S_d$$

Quel che accade è che  $S_X \subset S_{X_1} \times S_{X_2} \times \dots \times S_{X_d}$

$$\sum_{x_2, \dots, x_d} = \sum_{x_2 \in S_{X_2}} \sum_{x_3 \in S_{X_3}} \dots \sum_{x_d \in S_{X_d}}$$

Se avessimo scritto

$$\sum_{x_2 \in S_2} \sum_{x_3 \in S_3} \dots \sum_{x_d \in S_d} p_X(x, x_2, \dots, x_d)$$

Esempio  $X_1, X_2$  lancio di due dadi v.a.

- $(X_1, X_2)$   $P[X_1 = X_2] = \frac{1}{6}$  sono diverse in legge
- $(X_1, X_1)$   $P[X_1 = X_1] = 1$  dunque sono diverse

$$X: \Omega \rightarrow S_1$$

$$Y: \Omega \rightarrow S_2$$

## INDIPENDENZA

variabili aleatorie

indipendenza:  $X, Y$  v.a. su  $(\Omega, \mathcal{P})$  sono indipendenti se per ogni  $A_1 \subset S_1, A_2 \subset S_2$

$$P[X \in A_1, Y \in A_2] = P[X \in A_1] P[Y \in A_2]$$

ovvero, per ogni  $A_1, A_2$

$$\{X \in A_1\} \quad \{Y \in A_2\} \quad \text{indipendenti}$$

indipendenza (per famiglie finite)

$X_1, \dots, X_n$  v.a. su  $(\Omega, \mathcal{P})$   $X_i: \Omega \rightarrow S_i \quad i = 1, \dots, n$  indipendenti se per ogni  $A_1 \subset S_1, \dots, A_n \subset S_n$

$$P[X_1 \in A_1, \dots, X_n \in A_n] = P[X_1 \in A_1] \cdots P[X_n \in A_n]$$

(anche la probabilità di intersezione di sottofamiglie fattorizza in automatico, perché se voglio far sparire degli indici basta prendere  $A_i = S_i$  l'intero spazio  $S_i$ )

Nel caso generale:

indipendenza  $(X_i)_{i \in I}$  v.a. su  $(\Omega, \mathcal{P})$   $X_i: \Omega \rightarrow S_i$

sono indipendenti se per ogni  $J \subset I$  finito e ogni  $A_i \subset S_i$

$$i \in J \quad P\left(\bigcap_{i \in J} \{X_i \in A_i\}\right) = \prod_{i \in J} P[X_i \in A_i] \quad \text{definizione principale}$$

Osservazione  $A_1 = \{x_1\} \quad A_2 = \{x_2\}$ . Indipendenza:

$$(*) \quad P[X_1 = x_1, X_2 = x_2] = P[X_1 = x_1] P[X_2 = x_2] \leftarrow$$

(\*)  $\Leftrightarrow$  indipendenza di  $X_1, X_2$ . ( $\Leftarrow$ ) ovvio

Viceversa, se  $A_1 \subset S_1, A_2 \subset S_2$  (possiamo supporre  $A_1, A_2$  numerabili altrimenti si rimpiazzano  $A_i$  con  $A_i \cap X_i(\Omega)$ )

$$P[X_1 \in A_1, X_2 \in A_2] = P\left[\bigcup_{x_1 \in A_1} \bigcup_{x_2 \in A_2} \{X_1 = x_1, X_2 = x_2\}\right] \quad \text{certamente numerabile}$$

$$\text{poiché } \sum_{x_1 \in A_1} \sum_{x_2 \in A_2} P[X_1 = x_1, X_2 = x_2] = \sum_{x_1 \in A_1} \sum_{x_2 \in A_2} P[X_1 = x_1] P[X_2 = x_2]$$

$$= \sum_{x_1 \in A_1} \sum_{x_2 \in A_2} P[X_1 = x_1] P[X_2 = x_2]$$

fattorizzano  $\Rightarrow \left( \sum_{x_1 \in A_1} P[X_1 = x_1] \right) \left( \sum_{x_2 \in A_2} P[X_2 = x_2] \right)$

prop. dell'unione  $= P[X_1 \in A_1] P[X_2 \in A_2]$

Analogamente per famiglie finite/arbitrarie

### Osservazioni

1)  $(X_i)_{i \in I}$  indipendenti  $X_i: \Omega \rightarrow S_i$   $A_i \subset S_i$  per ogni  $i \in I$

allora  $(\{X_i \in A_i\})_{i \in I}$  sono eventi indipendenti

2)  $(X_i)_{i \in I}$  indipendenti e  $I' \subset I$ , allora  $(X_i)_{i \in I'}$  indipendenti

3) L'indipendenza è una proprietà della legge (congiunta), ovvero se

•  $(X_1, \dots, X_n)$  indipendenti

•  $(X_1, \dots, X_n)$  e  $(Y_1, \dots, Y_n)$  hanno la stessa legge (congiunta)

allora  $(Y_1, \dots, Y_n)$  sono indipendenti

$$P[Y_1 = x_1, \dots, Y_n = x_n] = P[X_1 = x_1, \dots, X_n = x_n]$$

se la legge congiunta è  $= \prod_{i=1}^n P[X_i = x_i] = \prod_{i=1}^n P[Y_i = x_i]$

le stesse allora anche le leggi marginali sono le stesse

4)  $(X_i)_{i \in I}$  indipendenti  $J_1, J_2 \subset I$   $J_1 \cap J_2 = \emptyset$  allora le

due famiglie  $(X_i)_{i \in J_1}$  e  $(X_i)_{i \in J_2}$  sono indipendenti,

nel senso: per ogni  $F_1 \subset J_1$  finito e  $F_2 \subset J_2$  finito, per ogni

$A_i \subset S_i$   $i \in F_1$  e  $B_i \subset S_i$   $i \in F_2$

$$P\left[\bigcap_{i \in F_1} \{X_i \in A_i\} \cap \bigcap_{i \in F_2} \{X_i \in B_i\}\right] = P\left[\bigcap_{i \in F_1} \{X_i \in A_i\}\right] \cdot P\left[\bigcap_{i \in F_2} \{X_i \in B_i\}\right]$$

5)  $A_1, \dots, A_n$  indipendenti  $\Leftrightarrow (\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_n})$  indipendenti

dove per un evento  $A \subset \Omega$   $\mathbb{1}_A(\omega) \begin{cases} 1 & \text{se } \omega \in A \\ 0 & \text{se } \omega \notin A \end{cases} \quad \omega \in \Omega$

infatti

$$P[\mathbb{1}_{A_1} = x_1, \dots, \mathbb{1}_{A_n} = x_n] = P\left[\bigcap_{i=1}^n A_i^{x_i}\right] \quad x_i \in \{0, 1\}$$

6)  $X, Y$  indipendenti,  $X: \Omega \rightarrow S_1, Y: \Omega \rightarrow S_2$  e  $f: S_1 \rightarrow S_1', g: S_2 \rightarrow S_2'$  allora  $f(X)$  e  $g(Y)$  sono indipendenti

Esempio  $n$  gettoni numerati da 1 a  $n$  estratti, uno dopo l'altro ordinatamente. Siano  $X_1, \dots, X_n$   $X_i = \text{il numero del gettone estratto alla } i\text{-esima estrazione}$

• con reinserimento  $\Omega = \{1, \dots, n\}^{\{1, \dots, n\}}$   $P$  uniforme

$$X_i(\omega) = \omega_i \quad \omega = (\omega_1, \dots, \omega_n) \in \Omega$$

$X_1, \dots, X_n$  sono indipendenti (chiaro)

$$P[X_1 = x_1, \dots, X_n = x_n] = P[\omega \in \Omega : \omega = (x_1, \dots, x_n)] = \frac{1}{n^n}$$

$$P[X_i = x_i] = \frac{1}{n}$$

• senza reinserimento  $X_1, \dots, X_n$  non sono indipendenti

$$\Omega = \text{permutazioni di } \{1, \dots, n\} = S_n \quad P = \text{prob. uniforme}$$

$$X_i(\omega) = \omega_i \quad \omega = (\omega_1, \dots, \omega_n) \in \Omega$$

$$P[X_i = x_i] = \frac{1}{n} \quad x_i \in \{1, \dots, n\}$$

$$P[X_1 = x_1, \dots, X_n = x_n] \stackrel{?}{=} \prod_{i=1}^n P[X_i = x_i]$$

$$= 0 \text{ quando } (x_1, \dots, x_n) \text{ non sono tutti distinti} \quad \frac{1}{n!} \quad \frac{1}{n^n} \text{ sempre } \neq 0$$

per verificare la non-indipendenza ci basta verificare 1 solo caso in cui le cose vanno male

(infatti sono diverse anche nel caso in cui  $x_1, \dots, x_n$  siano distinti)

Osservazione Date  $P_1, \dots, P_n$   $n$  densità discrete  $p_i: S_i \rightarrow \mathbb{R}$

Esistono  $n$  v.a.  $X_1, \dots, X_n$  con leggi rispettivamente  $P_1, \dots, P_n$ ?

Attraverso la costruzione canonica costruiamo  $X_1, X_2, \dots, X_n$  come richiesto, inoltre  $X_1, \dots, X_n$  indipendenti.

Osserviamo infatti che (\*) significa che se  $X_1, \dots, X_n$  sono indipendenti, allora

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

Con la costruzione canonica data le densità discrete

$$P(x_1, \dots, x_n) = \prod_{i=1}^n p_i(x_i)$$

possiamo costruire una v.a.  $X$  a valori in  $S_1 \times \dots \times S_n$  che ha  $p$  come densità discreta

$X = (X_1, \dots, X_n)$   $X_i: \Omega \rightarrow S_i$  legge marginale di  $X_i$  è  $p_i$

*conoscere le densità discrete marginali non aiuta a conoscere la densità congiunta (solo nel caso di variabili aleatorie indipendenti)*

proposizione  $(X_1, \dots, X_n)$  v.a. su  $(\Omega, \mathbb{P})$   $X_i: \Omega \rightarrow S_i$  con densità discreta  $p_{X_i}$ . Sia  $P_{(X_1, \dots, X_n)}$  la densità di  $(X_1, \dots, X_n)$ . Sono equivalenti

- $(X_1, \dots, X_n)$  indipendenti

- $P_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = p_{X_1}(x_1) \dots p_{X_n}(x_n)$  per ogni  $x_1 \in S_1, \dots, x_n \in S_n$

## VALORE ATTESO

Ricordiamo che per serie a termini positivi o assolutamente convergenti valgono le proprietà

- commutatività: se  $\sigma: \mathbb{N} \rightarrow \mathbb{N}$  biunivoca, allora

$$\sum_{n=0}^{\infty} a_n = \sum_{n=0}^{\infty} a_{\sigma_n}$$

- associatività: se  $(A_n)_{n \geq 0}$  una partizione di  $\mathbb{N}$ , allora

$$\sum_{n=0}^{\infty} a_n = \sum_{n=0}^{\infty} \left( \sum_{k \in A_n} a_k \right)$$

*può non valere per somme non ass. conv.*

## Valore atteso

Data una v.a. reale  $X: \Omega \rightarrow \mathbb{R}$  su  $(\Omega, \mathbb{P})$ , sia  $p$  la densità discreta associata a  $\mathbb{P}$ . !!

Se  $\sum_{\omega \in \Omega} |X(\omega)| p(\omega) < \infty$

(diremo che  $X$  è integrabile)

*nel caso generale il val. att.*

*si potrà def. come un integrale*

allora si definisce valore atteso (o speranza matematica, speranza

o valor medio):  $E[X] = \sum_{\omega \in \Omega} X(\omega) p(\omega)$  ( $E$  da expected value o espérance)

Se poi  $X \geq 0$  la def. vale anche senza la specifica sull'integrabilità perché serve a stabilire che la somma sia assolut. convergente, se  $X \geq 0$  allora la somma è a termini positivi, quindi ok.

lemma  $(\Omega, \mathbb{P})$  spazio di probabilità discreta

• Se  $X$  è una v.a. reale su  $(\Omega, \mathbb{P})$  allora

$$X \text{ integrabile} \Leftrightarrow \sum_{x \in \mathbb{R}} |x| p_x(x) < \infty$$

dove  $p_x$  è la densità discreta di  $X$

• Se  $X$  è una v.a. reale su  $(\Omega, \mathbb{P})$  integrabile o positiva, allora  $\mathbb{E}[X] = \sum_{x \in \mathbb{R}} x p_x(x)$  si riduce alla somma per gli  $x \in \mathbb{R}$  per cui la densità discreta sia positiva

• Se  $X: \Omega \rightarrow S$  v.a. su  $(\Omega, \mathbb{P})$ , se  $\varphi: S \rightarrow \mathbb{R}$  tale che  $\varphi(X)$  integrabile o positiva, allora

$$\mathbb{E}[\varphi(X)] = \sum_{x \in S} \varphi(x) p_x(x)$$

dim  $S_x = \{x \in \mathbb{R} : \mathbb{P}[X=x] > 0\}$   $\Omega = \bigcup_{x \in S_x} X^{-1}(x)$  partizione

$$\begin{aligned} \sum_{\omega \in \Omega} |X(\omega)| p(\omega) &\stackrel{\text{abs.}}{=} \sum_{x \in S_x} \sum_{\omega \in X^{-1}(x)} |X(\omega)| p(\omega) = \sum_{x \in S_x} |x| \sum_{\omega \in X^{-1}(x)} p(\omega) \\ &= \sum_{x \in S_x} |x| \mathbb{P}[X^{-1}(x)] = \sum_{x \in S_x} |x| \mathbb{P}[X=x] \\ &= \sum_{x \in S_x} |x| p_x(x) \end{aligned}$$

La seconda affermazione, ovvero  $\mathbb{E}[X] = \sum_{x \in S_x} x p_x(x) \left( = \sum_{x \in \mathbb{R}} x p_x(x) \right)$  segue analogamente

Valutiamo il valore atteso di  $\varphi(X)$

$$\begin{aligned} \sum_{\omega \in \Omega} \varphi(X(\omega)) p(\omega) &= \sum_{x \in S_x} \sum_{\omega \in X^{-1}(x)} \varphi(X(\omega)) p(\omega) \\ &= \sum_{x \in S_x} \varphi(x) \sum_{\omega \in X^{-1}(x)} p(\omega) \\ &= \sum_{x \in S_x} \varphi(x) p_x(x) \end{aligned}$$

|| ci basta la densità discreta di  $X$

## PROPRIETÀ

1) se  $a \in \mathbb{R}$   $E[a] = a$

2)  $a \in \mathbb{R}$ ,  $X$  v.a. reale integrabile/positiva  $E[aX] = a E[X]$

3) se  $P[X=0]=1$  allora  $E[X]=0$

4) se  $X \geq 0$  "q.c." (che significa  $P[X \geq 0] = 1$ ) e  $E[X] = 0$ ,  
allora  $X=0$  q.c. (ovvero  $P[X=0]=1$ ) "q.c." = quasi certamente

5) Se  $P[X=Y]=1$ ,  $X$  integrabile, allora anche  $Y$  è integrabile  
e  $E[X] = E[Y]$

6)  $X, Y$  integrabili/positive e  $P[X \leq Y] = 1$ , allora  $E[X] \leq E[Y]$

7) se  $X, Y$  integrabili, allora anche  $X+Y$  integrabile e  
 $E[X+Y] = E[X] + E[Y]$

dim. ② Supponiamo  $X$  integrabile (dovremmo verificare che  $aX$  è integrabile)  
 $E[aX] = \sum_{\omega \in \Omega} (aX)(\omega) p(\omega) = \sum_{\omega \in \Omega} aX(\omega) p(\omega) = a E[X]$

③  $P[X=0]=1$  significa  $p_X(x) = \begin{cases} 1 & \text{se } x=0 \\ 0 & \text{altrimenti} \end{cases}$

$$E[X] = \sum_{x \in \mathbb{R}} x p_X(x) = 0$$

④  $E[X] = \sum_{x \in \mathbb{R}} x p_X(x)$

in quanto  $p_X(x)=0$  se  $x < 0$   
 $\sum_{x \geq 0} x p_X(x) = 0$   
↑  
"0" ← "≥ 0"

quindi  $x p_X(x) = 0$  per ogni  $x \leq 0$

se  $x > 0$  allora  $p_X(x) = 0$

Ovvero  $p_X(x) = 0$  per ogni  $x \neq 0$ . Poiché

$$1 = \sum_{x \in \mathbb{R}} p_X(x) = p_X(0)$$

⑤  $P[X=Y]=1$ , allora  $X, Y$  hanno la stessa legge

$$\sum_{x \in \mathbb{R}} |x| p_Y(x) = \sum_{x \in \mathbb{R}} |x| p_X(x) < \infty \quad \sum x p_Y(x) = \sum x p_X(x)$$

⑥  $X \leq Y$  significa  $P[\{\omega: X(\omega) \leq Y(\omega)\}] = 1$

$$\sum_{\omega \in \Omega} X(\omega) p(\omega) = \sum_{\omega \in \{\omega: X(\omega) \leq Y(\omega)\}} X(\omega) p(\omega) \leq \sum_{\omega \in \{\dots\}} Y(\omega) p(\omega) = \sum_{\omega \in \Omega} Y(\omega) p(\omega)$$

$$\begin{aligned}
 \textcircled{7} \sum_{\omega \in \Omega} |(X+Y)(\omega)| p(\omega) &= \sum_{\omega \in \Omega} |X(\omega) + Y(\omega)| p(\omega) \\
 &\leq \sum_{\omega \in \Omega} |X(\omega)| p(\omega) + |Y(\omega)| p(\omega) \\
 &= \sum_{\omega \in \Omega} |X(\omega)| p(\omega) + \sum_{\omega \in \Omega} |Y(\omega)| p(\omega) < \infty
 \end{aligned}$$

$$\begin{aligned}
 \sum_{\omega \in \Omega} (X+Y)(\omega) p(\omega) &= \sum_{\omega \in \Omega} X(\omega) p(\omega) + Y(\omega) p(\omega) \\
 &= \sum_{\omega \in \Omega} X(\omega) p(\omega) + \sum_{\omega \in \Omega} Y(\omega) p(\omega)
 \end{aligned}$$

Esercizio 2.5  
29-03-21  
(Di Gesù)

### Problema 1.

Rifare il Problema 10 del Foglio 3 considerando come insieme dei risultati possibili  $\Omega = S_n \times \{0, 1\}^{(n-1)}$  e un'opportuna densità discreta  $p$  su  $\Omega$ .

### Problema 10. (Foglio 3)

Un gruppo di  $n$  studenti viene chiamato per un esame orale con una singolare modalità: gli studenti vengono numerati da 1 a  $n$  (ogni studente ha un numero differente). Si procede poi nel modo seguente:

1. si estrae un numero tra gli  $n$  disponibili. Se lo studente corrispondente al numero estratto non ha ancora sostenuto l'esame, allora sarà il candidato successivo;
2. se invece il numero estratto corrisponde ad uno studente già esaminato, allora l'estrazione si dirà sprecata e si procederà ad estrarre un numero tra i soli numeri non ancora estratti.

Si ripete poi la modalità illustrata fino a quando tutti gli studenti hanno sostenuto la prova.

Per  $j \in \{1, 2, \dots, n-1\}$ , sia  $A_j$  l'evento

$A_j = \{\text{non si spreca una estrazione per selezionare il } (j+1)\text{-esimo candidato}\}.$

- Calcolare la probabilità di  $A_j$  per ogni  $j=1, 2, \dots, n-1$ .
- Se  $1 \leq j \leq n-2$  dire se  $A_j$  e  $A_{j+1}$  sono indipendenti.
- Se  $1 \leq j \leq n-5$ , dire se  $A_j$  e  $A_{j+4}$  sono indipendenti.
- Determinare tutti i sottoinsiemi  $J \subset \{1, 2, \dots, n-1\}$  tali che gli eventi  $(A_j)$ , con  $j \in J$  siano indipendenti.

$$\begin{aligned}
 \Omega &= S_n \times \{0, 1\}^{n-1} \\
 &\quad (\sigma, (w_2, \dots, w_n))
 \end{aligned}$$

1 sprecata  
0 non sprecata

$w_1$  è ovviamente 0 poiché il primo non può essere già stato chiamato

$$p(\sigma, (w_2, \dots, w_n)) = \frac{1}{n} \prod_{k=2}^n \left( \frac{1}{n} \right)^{w_k} \left( \dots \right)^{1-w_k}$$

l'estrazione del  $k$ -esimo studente è indipendente dalle altre

$$w_k = 0 \quad \sigma(k) \quad \frac{1}{n}$$

$$w_k = 1 \quad \sigma(k) \quad \frac{1}{n-(k-1)} - \frac{1}{n} = \frac{k-1}{n(n-k+1)}$$

$$\sigma(k) \quad \} \quad 1, \dots, n \quad \sigma(k-1), \sigma(1), \dots, \sigma(\dots), \sigma(n)$$

$$\frac{1}{n-(k-1)}$$

$$\begin{aligned} 1) p(\sigma, (w_2, \dots, w_n)) &= \frac{1}{n} \prod_{k=2}^n \left( \frac{1}{n} \right)^{1-w_k} \left( \frac{k-1}{n(n-k+1)} \right)^{w_k} \\ &= \frac{1}{n} \prod_{k=1}^{n-1} \left( \frac{1}{n} \right)^{1-w_{k+1}} \left( \frac{k}{n(n-k)} \right)^{w_{k+1}} \end{aligned}$$

sono uguali in realtà

$$2) p(\sigma, (w_1, \dots, w_n)) = \frac{1}{n!} \prod_{k=1}^{n-1} \left( \frac{k}{n} \right)^{w_k} \left( \frac{n-k}{n} \right)^{1-w_k} \quad \text{formula alternativa}$$

poiché sono indipendenti si prende il .  
permutazioni

Bisogna poi verificare che è una densità:

$$\sum_{\substack{\sigma \in S_n \\ w_i \in \{0,1\}}} \left( \frac{1}{n} \prod_{k=1}^{n-1} \left( \frac{1}{n} \right)^{1-w_{k+1}} \left( \frac{k}{n(n-k)} \right)^{w_{k+1}} \right)$$

viene da  $w_1$

$$= \frac{1}{n} \sum_{\substack{\sigma \in S_n \\ w_i \in \{0,1\} \\ i \geq 3}} \underbrace{\left( \frac{1}{n} \prod_{k=2}^{n-1} \left( \frac{1}{n} \right)^{1-w_{k+1}} \left( \frac{k}{n(n-k)} \right)^{w_{k+1}} \right)}_{w_2=0} + \underbrace{\frac{1}{n(n-1)} \prod_{k=2}^{n-1} \left( \frac{1}{n} \right)^{1-w_{k+1}} \left( \frac{k}{n(n-k)} \right)^{w_{k+1}}}_{w_2=1}$$

$$(*) \quad \frac{1}{n} + \frac{1}{n(n-1)} = \frac{1}{n-1}$$

$$= \frac{1}{n} \frac{1}{n-1} \sum_{\substack{\sigma \in S_n \\ w_i \in \{0,1\} \\ i \geq 3}} \prod_{k=2}^{n-1} \left( \frac{1}{n} \right)^{1-w_{k+1}} \left( \frac{k}{n(n-k)} \right)^{w_{k+1}} = \frac{1}{n!} \sum_{\sigma \in S_n} 1 = \frac{n!}{n!} = 1$$

### Problema 7.

Assumiamo di essere in grado di taroccare un dado qualunque modificando a nostro piacere la distribuzione di probabilità per l'uscita delle sei facce. Possiamo taroccare due dadi in modo tale che la somma dei punteggi ottenuti con un lancio dei due dadi si distribuisca uniformemente sull'insieme  $\{2, \dots, 12\}$ ?

*È impossibile*

$$p(x) = p_1 x + p_2 x^2 + \dots + p_6 x^6 \quad 1^\circ \text{ dado}$$

$$\text{Voglio } p_1 + \dots + p_6 = 1$$

$$q(x) = q_1 x + q_2 x^2 + \dots + q_6 x^6 \quad 2^\circ \text{ dado}$$

$$\text{Voglio } q_1 + \dots + q_6 = 1$$

$$p(x)q(x) = p_1 q_1 x^2 + (p_1 q_2 + p_2 q_1) x^3 + \dots$$

$$\sum_{i=1}^6 p_i q_{n-i} \quad q_i = 0 \quad i < 1$$
$$q_i = 0 \quad i > 6$$

$$p(x)q(x) = \frac{1}{11} x^2 + \frac{1}{11} x^3 + \dots + \frac{1}{11} x^{12}$$
$$= \frac{1}{11} x^2 (1 + \dots + x^{10}) = \frac{x^{11} - 1}{x - 1}$$

*ha come radici tutte le radici 11<sup>e</sup> dell'unità tranne l'1  $\rightarrow$  non si può dividere in due polinomi di grado 5*

$$p(x) = x(p_1 + p_2 x + \dots + p_6 x^5)$$

$$q(x) = x(q_1 + q_2 x + \dots + q_6 x^5)$$

*Metodo semplice: "a mano"*

*(correzione del prof)*

$$P(\overset{\substack{\uparrow \\ \text{uscita} \\ 1^\circ \text{ dado}}}{X_1} + \overset{\substack{\uparrow \\ \text{uscita} \\ 2^\circ \text{ dado}}}{X_2} = k) = \sum_{j=1}^{k-1} P(X_1 + X_2 = k, X_1 = j) =$$

$$= \sum_{j=1}^{k-1} P(X_2 = k-j, X_1 = j) = \sum_{j=1}^{k-1} \underbrace{P(X_2 = k-j)}_{q_{k-j}} \underbrace{P(X_1 = j)}_{p_j}$$

$$= \frac{1}{11} \dots$$

$$\Rightarrow p_1 q_1 = \frac{1}{11}$$

$$p_1 q_2 + p_2 q_1 = \frac{1}{11}$$

⋮

$$p_6 q_6 = \frac{1}{11}$$

$$p_1 q_1 = \frac{1}{11} \Rightarrow q_1 = \frac{1}{11 p_1}$$

$$p_6 q_6 = \frac{1}{11} \Rightarrow q_6 = \frac{1}{11 p_6}$$

$$p_1 q_6 + p_2 q_5 + p_3 q_4 + p_4 q_3 + p_5 q_2 + p_6 q_1 = \frac{1}{11}$$

$$\Rightarrow p_1 q_6 + p_6 q_1 \leq \frac{1}{11} \Rightarrow \cancel{\frac{1}{11}} \frac{p_1}{p_6} + \cancel{\frac{1}{11}} \frac{p_6}{p_1} \leq \cancel{\frac{1}{11}}$$

$$\Rightarrow \frac{p_1^2 + p_6^2}{p_6 p_1} \leq 1 \quad \checkmark$$

$$\begin{aligned} x + \frac{1}{x} &\leq 1 \\ x^2 + 1 &\leq x \\ x^2 - x + 1 &\leq 0 \\ \underbrace{(x-1)^2}_{\geq 0} + \underbrace{x}_{> 0 \text{ per hp}} &\leq 0 \\ \text{sempre} \end{aligned}$$

### Problema 8.

All'incirca quanti lanci  $n$  di una moneta occorrono affinché lo scarto tra numero di Teste e numero di Croci uscite sia compreso nell'intervallo  $[(-5/100)n, (5/100)n]$  con probabilità almeno  $9/10$ ?

$$S_n = \text{numero di teste} \quad S_c = n - S_n$$

$$S_n - S_c = S_n - (n - S_n) = 2S_n - n$$

$$-\frac{5}{100} \leq 2S_n - n \leq \frac{5}{100}$$

$$\frac{1}{2} \left( -\frac{5}{100} + n \right) \leq S_n \leq \left( \frac{5}{100} + n \right) \frac{1}{2}$$

$$\mathbb{P}[X] \rightarrow \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \geq \frac{9}{10}$$

$$n \frac{1}{2} + \frac{\sqrt{n}}{2} a - \frac{1}{2} \left( -\frac{5}{100} + n \right)$$

$$\frac{n}{2} + \frac{\sqrt{n}}{2} a = -\frac{n}{40} - \frac{n}{2}$$

$$a = -\frac{1}{20} \sqrt{n}$$

$$b = \frac{1}{20} \sqrt{n}$$

$$\int_{-\frac{1}{20}\sqrt{n}}^{\frac{1}{20}\sqrt{n}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \rightarrow \int_0^{\frac{1}{20}\sqrt{n}} \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}}_{F(y)} dy \geq \frac{9}{10} \quad n \geq 1089$$

$$= 2 \left[ \int_{-\infty}^b F(y) dy - \underbrace{\int_{-\infty}^0 F(y) dy}_{\frac{1}{2}} \right]$$

NOTA BENE

$$P(X) = \underbrace{2 \int_{-\infty}^b \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy}_{\Phi(b)} - 1$$

↑  
evento che ci interessa

$$\text{con } b = \frac{\sqrt{n}}{20} \quad \int_{-b}^b \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

$$P(X) = 2 \Phi\left(\frac{\sqrt{n}}{20}\right) - 1 \geq \frac{9}{10} \Rightarrow 2 \Phi\left(\frac{\sqrt{n}}{20}\right) \geq \frac{19}{10} \Rightarrow \Phi\left(\frac{\sqrt{n}}{20}\right) \geq \frac{19}{20}$$

$$\Rightarrow \sqrt{n} \geq 20 \Phi^{-1}\left(\frac{19}{20}\right)$$

$$n \geq 400 \left[ \Phi^{-1}\left(\frac{19}{20}\right) \right]^2 \quad n = 1083$$

### Problema 3.

Un sacchetto contiene un pari numero di gettoni rossi e blu. Due giocatori pescano a turno un gettone (con reinserimento). Se i gettoni sono di colore uguale, i giocatori procedono con un'altra estrazione. Altrimenti vince chi ha pescato il gettone rosso. Determinare la probabilità che

- un giocatore vinca al primo turno;
- un giocatore vinca al k-esimo turno;
- la partita sia patta dopo n turni.

$$\text{Sia } \Omega = \{ (R, R), (R, B), (B, R), (B, B) \} = \{ (R, B) \}^2$$

dove R = si pesca il gettone rosso e B = si pesca il gettone blu e l'ordine indica se è stato pescato dal giocatore 1 o dal giocatore 2.

$$\text{Sia } A = \{ \text{"un giocatore vince al primo turno"} \} = \{ (R, B), (B, R) \}$$

$$\text{Altra auto: } P(A) = \frac{|A|}{|\Omega|} = \frac{2}{4} = \frac{1}{2}$$

La probabilità che un giocatore vinca al  $k$ -esimo turno è  $\frac{1}{2^{k-1}} \cdot \frac{1}{2} = \frac{1}{2^k}$ ,  
mentre la probabilità che la partita sia patta dopo  $n$  turni è  $\frac{1}{2^n}$ .

### NOTA BENE

$\Omega_0 = \{R, B\}^2$ ,  $p_0$  uniforme.

Poiché le prove sono  $\infty$  bisognerebbe prendere  $\Omega^\infty = \Omega^\infty$

Con  $n$  prove indipendenti ho  $\Omega = \Omega_0^n$ , con  $p$  uniforme ovvero  $p$  "prodotto"

Con  $\Omega^\infty$  non posso avere la prob. uniforme  $\rightarrow p$  "prodotto"

poiché  $\Omega^\infty$  non è più numerabile, lo spazio di probabilità

non è più discreto  $\rightarrow$  devo scegliere una  $\sigma$ -algebra

$\Omega = \mathbb{N} \rightarrow$  insieme numerabile e discreto

$$p(w) = q(1-q)^{n-1} \quad \text{con } q = \frac{1}{2}$$

Lezione 9  
30-03-2021  
(Romito)

### VALORE ATTESO

$(\Omega, \mathbb{P})$  spazio di probabilità discreta  $X: \Omega \rightarrow \mathbb{R}$  v.a. reale

Sia  $p$  la densità discreta associata a  $\mathbb{P}$   $p(a) = \mathbb{P}(\{w\})$   $w \in \Omega$

Se  $x$  è integrabile, ovvero,

$$\sum_{w \in \Omega} |X(w)| p(w) < \infty$$

allora si definisce

$$\mathbb{E}[X] = \sum_{w \in \Omega} X(w) p(w)$$

Abbiamo visto

$$\mathbb{E}[X] = \sum_{x \in \mathbb{R}} x p_x(x) = \sum_{x \in \mathbb{R}} x \mathbb{P}[X=x]$$

dove  $p_x$  è la densità discreta della v.a. quando

$$\sum_{x \in \mathbb{R}} |x| p_x(x) < \infty$$

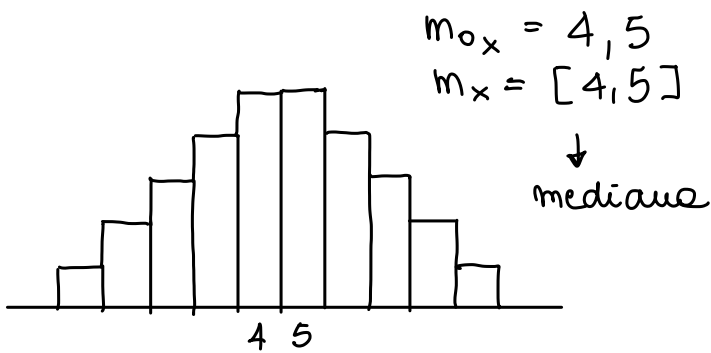
Cosa rappresenta il valore atteso?  
Rappresenta una sintesi di valori  
E' un indicatore di centralità

Indicatori di centralità:  
data una v.a. reale  $X$

- valore atteso
- mediana (valore di probabilità intermedia): ogni valore  $m_x$  tale che :  $P[X \leq m_x] \geq \frac{1}{2}$  e  $P[X \geq m_x] \geq \frac{1}{2}$
- moda (valore più probabile): ogni valore  $m_{ox}$  tale che

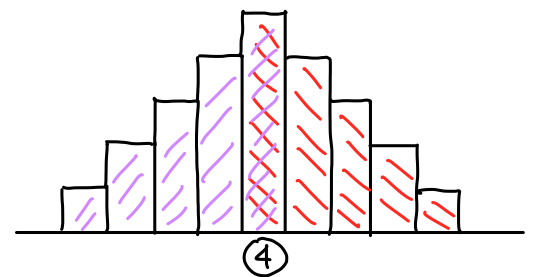
$$p_x(x) \leq p_x(m_{ox}) \text{ per ogni } x$$

(a livello di statistica descrittiva sono interessanti anche i massimi locali)



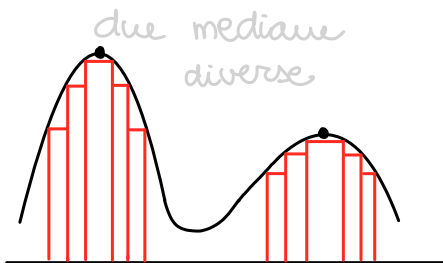
$$n = 9$$

$$p = 1/2$$



densità discreta Binomiale

$$n = 8 \quad p = \frac{1}{2}$$



Esempio: popolazione  
con all'interno  
un'altra popolazione

es: l'età delle persone  
in un'università



Esempio: Distribuzione  
del reddito di una popolazione

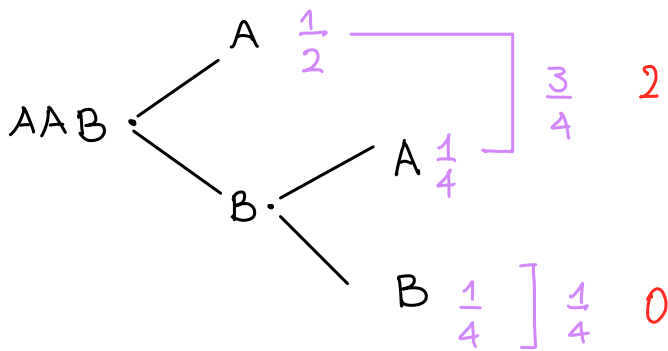
## Valor medio binomiale

$X$  binomiale  $n, p$   $p_x(k) = \binom{n}{k} p^k (1-p)^{n-k}$   $k = 0, 1, \dots, n$

$$E[X] = \sum_{k=0}^n k p_x(k) = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = (?)$$

Oservazione Consideriamo il problema: un gioco che si svolge a più mani e ognuna delle giocatrici ha ugual probabilità di vincere. Vince la prima che si aggiudica tre mani e vince la posta che è stata pagata per l'ingresso al gioco. Il problema riguarda come dividere la posta se il gioco si interrompesse prima della conclusione.

- se il gioco si interrompe A A B B  $\rightarrow$  in parti uguali
- se il gioco si interrompe A A B



Posta in gioco è 2  $\frac{3}{4} \cdot 2 + \frac{1}{4} \cdot 0 = \frac{3}{2}$

esempio Un sacchetto con  $2n$  gettoni, di cui  $n$  rossi e  $n$  blu. Si estraggono  $2n$  gettoni e si vuole calcolare il numero medio di gettoni rossi (ovvero il valore atteso del numero di gettoni estratti)

- estrazioni senza rimpiazzamento:  $n$
- estrazioni con rimpiazzamento: numero di gettoni è binomiale

di parametri,  $2n$  e  $\frac{1}{2}$

$G$ : numero di gettoni: binomiale  $(2n, \frac{1}{2})$

$X_1, \dots, X_{2n}$  v.a. Bernoulli, indipendenti e ognuna di parametro  $\frac{1}{2}$

Se andiamo a confrontare  $G$   $X_1 + \dots + X_{2n}$  sono diverse come variabili aleatorie MA sono uguali in legge.

$G$  legge  $X_1 + \dots + X_{2n}$  poiché sono uguali in legge, il loro valore atteso è lo stesso perché ciò dipende solo dalla legge

quindi

$$\begin{aligned} E[G] &= E[X_1 + \dots + X_{2n}] \\ &= E[X_1] + E[X_2] + \dots + E[X_{2n}] = 2n \cdot \frac{1}{2} = n \end{aligned}$$

$$X_1: \begin{cases} 0 & \frac{1}{2} \\ 1 & \frac{1}{2} \end{cases} \quad E[X_1] = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$$

Osservazione: Possiamo scrivere una v.a. binomiale di parametri  $n, p$  come:

$$X_1 + \dots + X_n$$

con  $X_1, \dots, X_n$  indipendenti e con distr. di Bernoulli di parametri  $p$

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

← valore limite nella legge dei grandi numeri

in quanto

$$X_i = \begin{cases} 0 & 1-p \\ 1 & p \end{cases} \quad E[X_i] = 1 \cdot p + 0(1-p) = p \quad E\left[\frac{X_1 + \dots + X_n}{n}\right] = p$$

momenti Se  $X$  è una v.a. reale

• se  $p \in \mathbb{R}$  il momento assoluto di ordine  $p$  di  $X$  è  $E[|X|^p]$

• se  $p \in \mathbb{Z}$  e se  $E[|X|^p] < \infty$  allora  $X$  ha momento di ordine  $p$  uguale a  $E[X^p]$

lemma Se  $X$  v.a. reale, se  $1 \leq p \leq q$  e se  $E[|X|^q] < \infty$  allora anche  $E[|X|^p] < \infty$

dimi introduciamo una notazione: per  $A$  evento,  $A \subset \Omega$

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{se } \omega \in A \\ 0 & \text{se } \omega \notin A \end{cases} \quad \text{indicatrice dell'evento } A$$

$$\mathbb{E}[|X|^p] = \mathbb{E}[|X|^p (\mathbb{1}_{\{|X| \geq 1\}} + \mathbb{1}_{\{|X| < 1\}})]$$

in quanto  $\mathbb{1}_A + \mathbb{1}_{A^c} = 1$

$$= \mathbb{E}[|X|^p \mathbb{1}_{\{|X| \geq 1\}}] + \mathbb{E}[|X|^p \mathbb{1}_{\{|X| < 1\}}]$$

$$|X|^p \mathbb{1}_{\{|X| < 1\}} \leq 1 \quad \text{con probabilità 1} \rightarrow \mathbb{E}[|X|^p \mathbb{1}_{\{|X| < 1\}}] \leq 1$$

$$|X|^p \mathbb{1}_{\{|X| \geq 1\}} \leq |X|^q \mathbb{1}_{\{|X| \geq 1\}} \leq |X|^q \quad \text{con prob. 1}$$

$$\rightarrow \mathbb{E}[|X|^p \mathbb{1}_{\{|X| \geq 1\}}] \leq \mathbb{E}[|X|^q \mathbb{1}_{\{|X| \geq 1\}}] \leq \mathbb{E}[|X|^q]$$

In conclusione

$$\mathbb{E}[|X|^p] \leq 1 + \mathbb{E}[|X|^q]$$

□

osservazione: Per il valore atteso vale la disuguaglianza di Jensen:

se  $X$  è una v.a. reale, se  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$  è una funzione convessa

se  $X$  e  $\varphi(X)$  sono integrabili (o a valori positivi)

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$$

il lemma precedente si può dimostrare  $S_n |X|$  e  $\varphi(X) = x^{q/p}$

VARIANZA Sia  $X$  una v.a. dotata di momento secondo (ovvero  $\mathbb{E}(X^2) < \infty$ )

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

oss anche  $X - \mathbb{E}[X]$  ha momento secondo perché

$$(X - \mathbb{E}[X])^2 \leq 2(X^2 + \mathbb{E}[X]^2)$$

$$\text{sd}(X) = \sqrt{\text{Var}(X)} \quad \text{deviazione standard}$$

osservazioni:

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

in quanto  $(X - \mathbb{E}[X])^2 = X^2 - 2\mathbb{E}[X] \cdot X + \mathbb{E}[X]^2$   
 $= \mathbb{E}[X^2] - \mathbb{E}[X]^2$

• Problema di minimizzazione

$$\min_m \mathbb{E}[(X - m)^2]$$

ha soluzione  $m = \mathbb{E}$

infatti  $\mathbb{E}[(X - m)^2] = \mathbb{E}[X^2] - 2m\mathbb{E}[X] + m^2$

• In effetti se invece consideriamo  $\min_m \mathbb{E}[|X - m|]$  la soluzione è la mediana

•  $X$  è costante se e solo se  $\text{Var}(X) = 0$  infatti  $\text{Var}(X) = 0$  significa

$$\mathbb{E}[(X - \mathbb{E}[X])^2] = 0 \text{ quindi } (X - \mathbb{E}[X])^2 = 0 \text{ con probabilità } 1, \text{ ovvero}$$

$$X = \mathbb{E}[X] \text{ con probab } 1$$

• Supponiamo che  $X$  sia binomiale di parametri  $n, p$

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

$$\mathbb{E}[X] = np$$

$$\mathbb{E}[X^2] = \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k}$$

$$\left| \begin{array}{l} \text{Var}(X) = np(1-p) \\ \text{sd}(X) = \sqrt{np(1-p)} \end{array} \right.$$

↑  
 $\mathbb{E}[\varphi(X)] = \sum_{x \in \mathbb{R}} \varphi(x) p_X(x)$

Ci ricordiamo che

$$X \stackrel{\text{legge}}{=} X_1 + \dots + X_n$$

$$(X_i)_{i=1, \dots, n}$$

Bernoulli ( $p$ )  
indipendenti

$$\begin{aligned} X^2 &= \sum_{i,j=1}^n X_i X_j = \sum_{i=1}^n X_i^2 + 2 \sum_{i < j} X_i X_j \\ &= \sum_{i=1}^n X_i + 2 \sum_{i < j} X_i X_j \end{aligned}$$

$$\mathbb{E}[X^2] = \sum_{i=1}^n \mathbb{E}[X_i] + 2 \sum_{i < j} \mathbb{E}[X_i X_j]$$

$$= np + 2 \sum_{i < j} \mathbb{E}[X_i X_j] = np + 2p^2 \cdot \frac{n(n-1)}{2}$$

$$\mathbb{E}[X_i, X_j] = X_i X_j = \begin{cases} 1 & \text{se } X_i = X_j = 1 \\ 0 & \text{altrimenti} \end{cases}$$

$$= \mathbb{P}[X_i = 1, X_j = 1] = \mathbb{P}[X_i = 1] \mathbb{P}[X_j = 1] = p^2$$

Teorema Siano  $X, Y$  v.a. su  $(\Omega, \mathcal{P})$  indipendenti e dotate di momento primo

Allora  $X \cdot Y$  ha momento primo e

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

dim  $(X, Y) : \Omega \rightarrow \mathbb{R}^2$ . Se  $p_x, p_y$  sono le densità discrete di  $X, Y$

allora  $p_{(x,y)}(x, y) = p_x(x) p_y(y)$  per ogni  $x, y \in \mathbb{R}$

$$\varphi : \mathbb{R}^2 \rightarrow \mathbb{R} \quad \varphi(x, y) = |x \cdot y|$$

$$\begin{aligned} \mathbb{E}[|X \cdot Y|] &= \mathbb{E}[\varphi(x, y)] = \sum_{x, y \in \mathbb{R}} \varphi(x, y) p_{(x,y)}(x, y) \\ &= \sum_{x \in \mathbb{R}} \sum_{y \in \mathbb{R}} |x| |y| p_x(x) p_y(y) = \left( \sum_{x \in \mathbb{R}} |x| p_x(x) \right) \left( \sum_{y \in \mathbb{R}} |y| p_y(y) \right) < \infty \\ &\quad = \mathbb{E}(|X|) = \mathbb{E}(|Y|) \end{aligned}$$

$$t(x, y) = x \cdot y$$

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[\varphi(x, y)] = \sum_{x \in \mathbb{R}} \sum_{y \in \mathbb{R}} x y p_x(x) p_y(y) = \mathbb{E}[X] \mathbb{E}[Y]$$

Osservazioni ① se  $X, Y$  non sono indipendenti, non è detto che il prodotto sia integrabile, nell'ipotesi che  $X, Y$  siano integrabili.

In generale  $XY$  è integrabile se  $X, Y$  hanno momento secondo

$$|XY| \leq \frac{1}{2}(X^2 + Y^2)$$

② Non vale il viceversa, ovvero esistono  $X, Y$  non indipendenti ma per le quali valgono le conclusioni del teorema

esempio

$X$  uniforme su  $\{-1, 0, 1\}$  e  $Y = |X|$

• non indipendenti:  $P[X=0, Y=1] = 0 \neq \frac{1}{3} \cdot \frac{2}{3} = P[X=0] P[Y=1]$

• scorrelate:  $\mathbb{E}[X] = 0$   $\mathbb{E}[Y] = \frac{2}{3}$   $\mathbb{E}[XY] = \mathbb{E}[X|X|] = \mathbb{E}[X] = 0$

Covarianza  $X, Y$  v.a. reali su  $(\Omega, \mathcal{P})$  dotate di momento secondo

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

Proprietà:

$$\textcircled{1} \text{ cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y]$$

② se  $X, Y$  indipendenti allora  $\text{Cov}(X, Y) = 0$  (scorrelate) ma non vale il viceversa

③ Cov simmetrica:  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$

④ Cov bilineare  $\text{Cov}(\lambda X_1 + \mu X_2, Y) = \lambda \text{Cov}(X_1, Y) + \mu \text{Cov}(X_2, Y)$

⑤  $\text{Cov}(X, X) = \text{Var}(X)$

⑥  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$

infatti  $(X+Y)^2 \leq 2(X^2 + Y^2)$

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

$$= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - \mathbb{E}[X]^2 - 2\mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[Y]^2$$

$$= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

⑥' In generale se  $X_1, \dots, X_n$  hanno momento secondo

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

⑥'' Se  $X_1, \dots, X_n$  sono indipendenti (è sufficiente scorrelate)

$$\text{Var}\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n \text{Var}(X_j)$$

Usiamo questa formula per calcolare la varianza di una binomiale  $(n, p)$

$X$  binomiale  $\overset{\text{legge}}{\sim} X_1 + \dots + X_n$   $X_1 \dots X_n$  Bernoulli( $p$ )

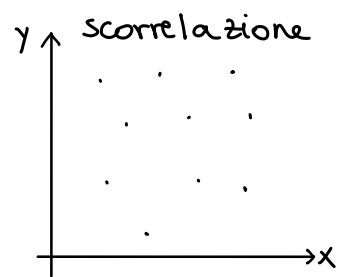
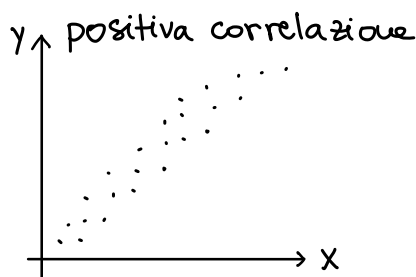
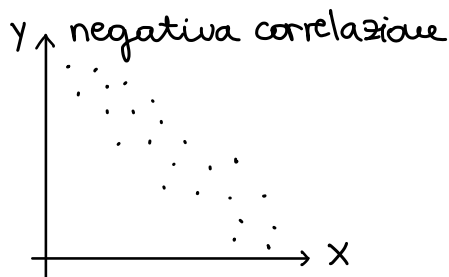
$$\mathbb{E}[X] = np$$

$$\begin{aligned} \text{Var}(X) &= \text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n) = n\text{Var}(X_1) \\ &= np(1-p) \end{aligned}$$

$$\text{Var}(X_1) = \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2 = \mathbb{E}[X_1] - \mathbb{E}[X_1]^2 = p(1-p)$$

$$\text{De Moivre - Laplace} \quad \mathbb{P}\left[\underbrace{np}_{\mathbb{E}S_n} + \underbrace{a\sqrt{np(1-p)}}_{\text{sd}(S_n)} \leq S_n \leq \underbrace{np}_{\mathbb{E}S_n} + \underbrace{b\sqrt{np(1-p)}}_{\text{sd}(S_n)}\right] \rightarrow \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

$$\mathbb{P}\left[a \leq \frac{S_n - \mathbb{E}S_n}{\text{sd}(S_n)} \leq b\right] \rightarrow \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$



Correlazione  $X, Y$  v.a. dotate di momento secondo e non costanti

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\text{sd}(X) \text{sd}(Y)}$$

noto anche come coefficiente di correlazione (o di Pearson)

### Proprietà

- ①  $X, Y$  indipendenti,  $\rho(X, Y) = 0$  (non vale il viceversa)
- ②  $\rho$  simmetrica
- ③  $|\rho(X, Y)| \leq 1$
- ④  $|\rho(X, Y)| = 1$  se e solo se esistono  $a, b$  tali che  $Y = aX + b$  e  $a, \rho(X, Y)$  sono concordi

dim

$$X' = \frac{X - \mathbb{E}[X]}{\text{sd}(X)}$$

$$Y' = \frac{Y - \mathbb{E}[Y]}{\text{sd}(Y)}$$

$$\mathbb{E}[X'] = \mathbb{E}[Y'] = 0$$

$$\text{Var}(X') = \text{Var}(Y') = 1$$

$$\rho(X', Y') = \rho(X, Y) = \text{cov}(X', Y') = \mathbb{E}[X' Y']$$

$$0 \leq \mathbb{E}[(X' + Y')^2] = \mathbb{E}[(X')^2] + 2\mathbb{E}[X' Y'] + \mathbb{E}[(Y')^2]$$

$$= 1 + 2\rho(X, Y) + 1$$

quindi  $0 \leq 2(1 + \rho(X, Y))$  cioè  $\rho(X, Y) \geq -1$

$0 \leq \mathbb{E}[(X' - Y')^2] = 2(1 - \rho(X, Y))$  quindi  $\rho(X, Y) \leq 1$

Per la ④ Se  $Y = aX + b$   $\text{Var}(Y) = a^2 \text{Var}(X)$

$$\text{Cov}(X, Y) = a \mathbb{E}[(X - \mathbb{E}[X])^2] = a \text{Var}(X)$$

$$\rho(X, Y) = \frac{a \text{Var}(X)}{\text{sd}(X) |a| \text{sd}(X)} = \frac{a}{|a|} = \text{sgn}(a)$$

Supponiamo ora  $|\rho(X, Y)| = 1$ , ad esempio  $\rho(X, Y) = 1$

$$0 \leq \mathbb{E}[(X' - Y')^2] = 2(1 - \rho(X, Y)) = 1$$

quindi  $\mathbb{E}[(X' - Y')^2] = 0$

cioè  $(X' - Y')^2 = 0$  con probabilità 1

$$X' = Y'$$

$$\frac{X - \mathbb{E}[X]}{\text{sd}(X)} = \frac{Y - \mathbb{E}[Y]}{\text{sd}(Y)} \quad \text{ovvero} \quad Y = \frac{a}{\text{sd}(Y)} + \frac{b}{\text{sd}(X)} \left[ \mathbb{E}[Y] - \frac{\text{sd}(Y)}{\text{sd}(X)} \mathbb{E}[X] \right]$$

Se poi  $\rho(X, Y) = -1$  avremo  $Y' = -X'$

□

Osservazione Correlazione è correlazione lineare

$\rho(X, X^2)$  correlazione  $\neq$  causalità "spurious correlations"

Distribuzione condizionale  $X, Y$  v.a. su  $(\Omega, \mathbb{P})$

la legge condizionale di  $Y$  dato  $X = x$  (con  $\mathbb{P}[X = x] > 0$ )

è data da:  $S_{Y|X} = \{y: P_{(X,Y)}(x, y) > 0\}$

$$P_{Y|X}(y|x) = \mathbb{P}[Y = y | X = x]$$

Valore atteso condizionato

Se  $\mathbb{P}[X = x] > 0$  il valore atteso condizionato a  $X = x$  è

$$\mathbb{E}[Y | X = x] = \sum_y y P_{Y|X}(y|x)$$

ovvero il valore atteso valutato per mezzo della legge condizionale

Il valore atteso di  $Y$  condizionato a  $X$

$$\mathbb{E}[Y|X] = \sum_{x: \mathbb{P}[X=x] > 0} \mathbb{E}[Y | X = x] \mathbb{1}_{\{X=x\}}$$

Osservazioni ①  $\mathbb{E}[Y|X]$  è una funzione di  $X$

②  $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$

infatti

$$\mathbb{E}[\mathbb{E}[Y|X]] = \sum_{x: \mathbb{P}[X=x] > 0} \mathbb{E}[Y|X=x] \mathbb{E}[\mathbb{1}_{\{X=x\}}]$$

$$= \sum_{x: \mathbb{P}[X=x] > 0} \mathbb{E}[Y|X=x] \mathbb{P}[X=x]$$

in quanto  $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}[A]$

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[Y|X]] &= \sum_{x: P[X=x] > 0} \sum_y y P_{Y|X}(y|x) P[X=x] \\
&= \sum_{x: P[X=x] > 0} \sum_y y P[X=x] P[Y=y|X=x] \\
&= \sum_{x: P[X=x] > 0} \sum_y y P[Y=y, X=x] \\
&= \sum_y y P[Y=y] = \mathbb{E}[Y]
\end{aligned}$$

$$\{Y=y\} = \bigcup_{x \in X(\Omega)} \{Y=y, X=x\} \quad \text{unione disgiunta}$$

$$P[Y=y] = \sum_x P[Y=y, X=x]$$

## VARIABILI ALEATORIE A VALORI INTERI

$X: \Omega \rightarrow \mathbb{N}$  oppure  $\mathbb{Z}$

lemma (formula di convoluzione)  $X, Y$  indipendenti e intere allora  $p_{X+Y}(n) = \sum_{j \in \mathbb{Z}} p_X(j) p_Y(n-j) \quad n \in \mathbb{Z}$

dim  $\{X+Y=n\} = \bigcup_{j \in \mathbb{Z}} \{X=j, Y=n-j\}$

$$\begin{aligned}
P[X+Y=n] &= \sum_{j \in \mathbb{Z}} P[X=j, Y=n-j] \\
&= \sum_{j \in \mathbb{Z}} P[X=j] P[Y=n-j]
\end{aligned}$$

esempio  $X, Y$  v.a. indipendenti e di Poisson di parametri  $\lambda, \mu$

Allora anche  $X+Y$  è di Poisson di parametro  $\lambda+\mu$

$X$  di Poisson  $p_X(n) = \frac{\lambda^n}{n!} = e^{-\lambda} \quad n \geq 0$

esercizio: usare la formula di convoluzione e la formula del binomio

funzione generatrice delle probabilità

$X$  è una v.a. a valori in  $\mathbb{N}$

$$g_x(t) = \mathbb{E}[t^X] = \sum_{n=0}^{\infty} p_x(n) t^n$$

definita nell'insieme di convergenza della serie di potenze

raggi di convergenza  $\geq 1$

proprietà

•  $X, Y$  a valori in  $\mathbb{N}$

$g_x = g_y$  in un intorno di 0  $\Leftrightarrow X, Y$  hanno la stessa legge

• Se  $X, Y$  sono indipendenti, allora  $g_{X+Y} = g_X g_Y$

•  $\mathbb{E}[X] = \lim_{t \rightarrow 1^-} g'_x(t)$

•  $\mathbb{E}[X(X-1)] = \lim_{t \rightarrow 1^-} g''_x(t)$

dim •  $g_x^{(n)}(0) = n! p_x(n)$

•  $g_{X+Y}(t) = \mathbb{E}[t^{X+Y}] = \mathbb{E}[t^X \cdot t^Y] = \mathbb{E}[t^X] \mathbb{E}[t^Y] = g_X(t) g_Y(t)$

$X, Y$  indep.  $\longrightarrow t^X, t^Y$  indep.

□

Esercizio 6  
12 - 04 - 21  
(Di Gesù)

### Problema 7.

Sia  $e(p) = \mathbb{E}(f(X_1, \dots, X_n))$ , con  $X_1, X_2, \dots, X_n$  variabili aleatorie indipendenti e con distribuzione di Bernoulli di parametro  $p$ , e  $f: \{0, 1\}^n \rightarrow \mathbb{R}$  una funzione crescente nel senso seguente:

vale  $f(x) \leq f(y)$  se  $x = (x_1, \dots, x_n)$  e  $y = (y_1, \dots, y_n)$  soddisfano  $x_k \leq y_k$  per ogni  $k = 1, \dots, n$ .

Dimostrare che  $e(p_1) \leq e(p_2)$  per  $p_1 \leq p_2$ .

$p_1 \leq p_2 \Rightarrow e(p_1) \leq e(p_2)$

$e(p) = \mathbb{E} f(X_1, \dots, X_n)$  1<sup>a</sup> idea: induzione

2<sup>a</sup> idea: prendiamo lo stesso spazio di probabilità sia per le  $X_k$  e le  $Y_k$  e li scegliamo dipendenti in modo di forzare  $Y_k \geq X_k$

$\mathbb{E} f(X_1, \dots, X_n) \leq \tilde{\mathbb{E}} f(Y_1, \dots, Y_n)$   
( $\Omega, \mathbb{P}$ ) ( $\tilde{\Omega}, \tilde{\mathbb{P}}$ )

$$Y_K \sim \text{Ber}(p_2) \quad p_2 > 0$$

$$A_K \sim \text{Ber}\left(\frac{p_1}{p_2}\right), \quad X_K = A_K Y_K = \begin{cases} 1 & \text{con prob } p_1 \\ 0 & \text{con prob } 1-p_1 \end{cases}$$

$$X_K = 1 \Rightarrow Y_K = 1$$

$$\text{quindi } Y_K(\omega) \geq X_K(\omega) \quad \forall \omega$$

### Problema 8.

- (i) Sia  $X$  una variabile aleatoria a valori in  $\mathbb{N}$ . Dimostrare l'uguaglianza  $E(X) = \sum_{k=0}^{\infty} P[X > k]$ .
- (ii) Ci sono  $b$  palline blu e  $r$  palline rosse in un contenitore. Le palline vengono estratte a caso una dietro l'altra fino a quando esce una pallina blu. Mostrare che il valore atteso del numero di estrazioni effettuate è  $(b+r+1)/(b+1)$

$$i) \text{ Sia } X \text{ una v.a. a valori in } \mathbb{N} \Rightarrow E(X) = \sum_{k=0}^{\infty} P[X > k]$$

$$P[X > k] = \sum_{i>k} p_x(i) = \sum_{i \in \mathbb{N}} \mathbb{1}_{(i>k)} p_x(i), \quad \forall k \in \mathbb{N}$$

Da cui si ottiene:

$$\sum_{k=0}^{\infty} P(X > k) = \sum_{k=0}^{\infty} \sum_{i=0}^{\infty} \mathbb{1}_{(i>k)} p_x(i) = \sum_{i=0}^{\infty} \left( \sum_{k=0}^{\infty} \mathbb{1}_{(i>k)} \right) p_x(i) = \sum_{i=0}^{\infty} i p_x(i) = E[X]$$

Visto che  $\sum_{k=0}^{\infty} \mathbb{1}_{(i>k)} = i$  infatti

$$\mathbb{1}_{(i>k)} = \begin{cases} 1 & \text{se } 0 \leq k \leq i-1 \\ 0 & \text{se } k \geq i \end{cases}$$

ii) Sia  $X$  la variabile aleatoria che rappresenta il numero di estrazioni fino a quella della pallina blu.

La probabilità che la prima pallina estratta sia rossa e la seconda blu è  $p_x^{(1)} = \frac{r}{b+r} \cdot \frac{b}{b+r-1}$ .

La probabilità che le prime 2 palline estratte siano rosse e la terza blu è  $p_x^{(2)} = \frac{r}{b+r} \cdot \frac{r-1}{b+r-1} \cdot \frac{b}{b+r-2}$ .

Ripetendo il ragionamento per  $k$  volte abbiamo:

$$p_x^{(k)} = \frac{r!}{(r-k)!} \cdot \frac{(b+r-k-1)!}{(b+r)!} \cdot b$$

A questo punto voglio calcolare il valore atteso di  $X$ .

$$\mathbb{E}[X] = \sum_{k \in \mathbb{R}} k P_X(k) = \sum_{k \in \mathbb{R}} k P(X=k)$$

che per il punto (i) è uguale a

$$\sum_{k=0}^{\infty} P(X > k) = \sum_{k=0}^{\infty} \frac{r}{b+r} \cdot \frac{r-1}{b+r-1} \cdots \frac{r-k+1}{b+r-k+1} = \sum_{k=0}^r \frac{r!}{(r-k)!} \cdot \frac{(b+r-k)!}{(b+r)!} = \frac{r! b!}{(b+r)!} \sum_{k=0}^r \frac{(b+r-k)!}{(r-k)! b!}$$

che per la definizione dei binomiali posso scrivere come

$$\frac{r! b!}{(b+r)!} \sum_{k=0}^r \binom{b+r-k}{b} = \frac{r! b!}{(b+r)!} \sum_{n=0}^r \binom{b+n}{b}$$

$$\text{Voglio vedere che } \frac{r! b!}{(b+r)!} \sum_{n=0}^r \binom{b+n}{b} = \frac{b+r+1}{b+1}$$

Questo è vero se e solo se

$$\sum_{n=0}^r \binom{b+n}{b} = \frac{(b+r+1)!}{r! (b+1)!} = \binom{b+r+1}{b+1}$$

$$\text{Poiché } \binom{b+r+1}{b+1} = \binom{b+r}{r} + \binom{b+r}{r-1}, \text{ iterando si ha:}$$

$$\binom{b+r+1}{r+1} = \binom{b+r}{r} + \binom{b+r-1}{r-1} + \binom{b+r-2}{r-2} + \dots$$

$$\dots = \sum_{k=0}^r \binom{b+r-k}{r-k} = \sum_{n=0}^r \binom{b+n}{b} \quad \text{per la proprietà del binomiale}$$

idea 2:  $\Omega = S_{b+r}$   $P$  uniforme  $X: \Omega \rightarrow \mathbb{N}$   
 $\sigma \mapsto \min\{i \mid \sigma(i) \text{ blu}\}$

$$k > r \quad P[X > k] = 0$$

$$k = 0, \dots, r \quad P[X > k] = \binom{r}{k} \frac{k! (b+r-k)!}{(b+r)!}$$

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} P[X > k] = \sum_{k=0}^r P[X > k] = \dots \quad \text{come prima}$$

$$(*) \quad \underbrace{\sum_{n=0}^r \binom{b+n}{b}}_{=d(b,r)} = \underbrace{\binom{b+r+1}{b+1}}_{=B(b,r)} \quad \forall b, r \in \mathbb{N}$$

$$\sum_{r=0}^{\infty} d(b,r) x^r = \sum_{r=0}^{\infty} B(b,r) x^r \quad \forall b \in \mathbb{N} \quad \forall |x| \text{ suff. piccolo}$$

$$(1+y)^{-m} = \sum_{k=0}^{\infty} (-1)^k \binom{m+k-1}{k} y^k \quad y \neq -1, m \in \mathbb{N}$$

### Problema 9.

Siano  $X, Y$  e  $Z$  variabili aleatorie indipendenti tutte di legge di Poisson di parametro  $\lambda$ .

- Qual è la legge condizionale di  $X$  dato  $X+Y=n$ ? Si tratta di una legge nota? Quanto vale la media di questa legge condizionale?
- Quanto vale la covarianza delle variabili aleatorie  $X+Y$  e  $X+Z$ ?
- Quanto vale  $P[X+Y=2, X+Z=3]$ ?

• Siano  $X, Y, Z$  variabili aleatorie e sia  $\Omega = \mathbb{N}$ ,  
e si ha  $p_x(k) = p_y(k) = p_z(k) = \frac{\lambda^k}{k!} e^{-\lambda}$

La legge condizionale di  $X$  dato  $X+Y=n$  è

$$S_{X|n} = \{x : P_{(X, X+Y)}(x, n) > 0\}$$

$$P_{X|n}(x, n) = P(X=x | X+Y=n) = \frac{P(X=x, Y=n-x)}{P(X+Y=n)}$$

Poiché le variabili sono indipendenti si ha:

$$P(X=x, Y=n-x) = P(X=x)P(Y=n-x) = e^{-\lambda} \frac{\lambda^x}{x!} e^{-\lambda} \left( \frac{\lambda^{n-x}}{(n-x)!} \right)$$

$$P(X+Y=n) = e^{-2\lambda} \frac{(2\lambda)^n}{n!}$$

Quindi si ha:

$$P(X=x | X+Y=n) = \frac{e^{-2\lambda} \frac{\lambda^x}{x!} \frac{\lambda^{n-x}}{(n-x)!}}{e^{-2\lambda} \frac{(2\lambda)^n}{n!}} = \binom{n}{x} \left(\frac{1}{2}\right)^n$$

Da qui notiamo che si tratta di un binomiale

$B(n, \frac{1}{2})$ , e quindi la sua media sarà:

$$E(X=x | X+Y=n) = np = \frac{n}{2}$$

- La covarianza delle variabili aleatorie  $X+Y$  e  $X+Z$  è  
 $Cov(X+Y, X+Z) = Cov(X, X) + Cov(Y, X) + Cov(X, Z) + Cov(Y, Z)$

Essendo  $X, Y, Z$  indipendenti e poiché la covarianza tra variabili aleatorie indipendenti è pari a 0 abbiamo

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

$$E[X] = \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} = \sum_{n=1}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} = \lambda e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

$$\begin{aligned}
 E[X^2] - (E[X])^2 &= \sum_{n=0}^{\infty} n^2 \frac{\lambda^n}{n!} e^{-\lambda} - \lambda^2 = \sum_{n=0}^{\infty} n(n-1) \frac{\lambda^n}{n!} e^{-\lambda} + \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} - \lambda^2 \\
 &= \sum_{n=0}^{\infty} n(n-1) \frac{\lambda^n}{n!} e^{-\lambda} + \lambda - \lambda^2 = \lambda^2 e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^{n-2}}{(n-2)!} + \lambda - \lambda^2 \\
 &= \lambda^2 + \lambda - \lambda^2 = \lambda
 \end{aligned}$$

Dunque abbiamo  $\text{Cov}(X+Y, X+Z) = \lambda$ .

$$\begin{aligned}
 \bullet P[X+Y=2, X+Z=3] &= P[\{(0,2,3)\} \cup \{(1,1,2)\} \cup \{(2,0,1)\}] \\
 &= P[\{(0,2,3)\}] + P[\{(1,1,2)\}] + P[\{(2,0,1)\}] = \\
 &= e^{-\lambda} \frac{\lambda^0}{0!} \cdot e^{-\lambda} \frac{\lambda^2}{2!} \cdot e^{-\lambda} \frac{\lambda^3}{3!} + e^{-\lambda} \frac{\lambda^1}{1!} \cdot e^{-\lambda} \frac{\lambda^1}{1!} \cdot e^{-\lambda} \frac{\lambda^2}{2!} + e^{-\lambda} \frac{\lambda^2}{2!} \cdot \\
 &\quad \cdot e^{-\lambda} \frac{\lambda^0}{0!} \cdot e^{-\lambda} \frac{\lambda^1}{1!} = e^{-3\lambda} \left( \frac{\lambda^5}{12} + \frac{\lambda^4}{2} + \frac{\lambda^3}{2} \right)
 \end{aligned}$$

### Problema 10.

Vengono lanciate  $n$  monete. Le monete che presentano testa vengono lanciate di nuovo. Determinare il valore atteso e la varianza del numero di teste risultanti.

la probabilità ad ognuno degli  $n$  lanci di ottenere Testa è  $p = \frac{1}{4}$  mentre quella di ottenere Croce è  $(1-p) = \frac{3}{4}$ .

$$\text{Sia } X = X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i \quad \text{con } X_i = \begin{cases} 1 & p = \frac{1}{4} \\ 0 & p = \frac{3}{4} \end{cases}$$

la variabile aleatoria del numero di teste ottenute.

Noto che  $X$  è una binomiale, perciò il suo valore atteso sarà  $E[X] = n \cdot p = n \cdot \frac{1}{4} = \frac{n}{4}$ .

Mentre la sua Varianza sarà:

$$\text{Var}(X) = np(1-p) = n \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{3n}{16}$$

( $p = \frac{1}{4}$  perché  $\begin{matrix} & T & C \\ C & \swarrow & \searrow \\ & T & C \end{matrix}$  e quindi  $P(T) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$  e  $1-p = P(C) = \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4}$ )

Idea 2:

$X = \#$  teste nei primi  $n$  lanci

$Y = \#$  teste nei secondi  $n$  lanci

$$\mathbb{E}[Y] = \sum_{k=0}^n \underbrace{\mathbb{E}[Y|X=k]}_{\frac{k}{2}} \underbrace{\mathbb{P}(X=k)}_{\binom{n}{k} \frac{1}{2^n}}$$

$$= \sum_{k=0}^n \frac{k}{2} \binom{n}{k} \frac{1}{2^n} = \frac{1}{2} \mathbb{E}[X] = \frac{n}{4}$$

Analogo per la  $\text{Var}(X)$

### Problema 6.

In un gioco, il concorrente gioca una somma iniziale  $P > 0$ . La probabilità di vittoria ad ogni turno è  $p$  e la somma giocata viene raddoppiata se vince, dimezzata se perde. Qual è la quantità attesa di denaro posseduta dal giocatore dopo l' $n$ -esimo turno? Calcolarne il limite per  $n \rightarrow \infty$ .

Fissato il numero di turni  $n$ , chiamiamo la quantità attesa di denaro al turno  $n$ ,  $D_n$  e questa si può scrivere come  $D_n = P \cdot X_1 \cdot X_2 \cdot X_3 \cdots X_n$  dove  $X_i$  con  $i \in \{1, \dots, n\}$  sono variabili aleatorie indipendenti che hanno la stessa distribuzione:

$$X_i = \begin{cases} 2 & \text{con probabilità } p \\ \frac{1}{2} & \text{con probabilità } 1-p \end{cases}$$

Quindi si ha  $\mathbb{E}[D_n] = \mathbb{E}[P \cdot X_1 \cdot X_2 \cdots X_n]$  e poiché sono indipendenti fra loro si può scrivere  $X_1 = X_2 = \cdots = X_n$  e quindi  $\mathbb{E}[D_n] = P \cdot \mathbb{E}[X_1]^n$   
 $P \cdot \mathbb{E}[X_1]^n = P \left( 2p + \frac{1}{2}(1-p) \right)^n = P \left( \frac{3p+1}{2} \right)^n$ .

Perciò abbiamo:

$$\lim_{n \rightarrow \infty} \mathbb{E}[D_n] = \begin{cases} 0 & \text{se } 0 \leq p < \frac{1}{3} \\ P & \text{se } p = \frac{1}{3} \\ \infty & \text{se } \frac{1}{3} < p \leq 1 \end{cases}$$

## Miniriepilogo

- Valore atteso, varianza, deviazione standard
- covarianza, correlazione
- variabili aleatorie a valori interi

funzione generatrice (delle probabilità)

$$X \text{ a valori interi (in } \mathbb{N}) \quad g_X(t) = \mathbb{E}[t^X] \\ = \sum_{n=0}^{\infty} p_X(n) t^n$$

ove  $p_X$  è la densità discreta di  $X$ . Ben definita almeno quando  $|t| \leq 1$

- $X, Y$  hanno la stessa legge  $\Leftrightarrow g_X = g_Y$  in un intorno di 0

esempi 1.  $X$  Bernoulli ( $p$ )  $g_X(t) = (1-p) + pt$

2.  $X$  binomiale ( $n, p$ )

$$g_X(t) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} t^k = \sum_{k=0}^n \binom{n}{k} (tp)^k (1-p)^{n-k} = (tp + 1-p)^n$$

$X \stackrel{\text{legge}}{=} X_1 + \dots + X_n$  con  $X_1, \dots, X_n$  Bernoulli( $p$ ) indipendenti

$$g_X(t) \stackrel{\text{indip.}}{=} g_{X_1 + \dots + X_n}(t) = g_{X_1}(t) \cdot \dots \cdot g_{X_n}(t) = (1-p + tp)^n$$

3. geometrica ( $p$ )  $p_X(n) = p(1-p)^{n-1} \quad n \geq 1 \quad X \sim \text{geometrica}(p)$

$$g_X(t) = \sum_{n=1}^{\infty} t^n p(1-p)^{n-1} = tp \sum_{n=1}^{\infty} (t(1-p))^{n-1} = tp \sum_{n=0}^{\infty} (t(1-p))^n \\ = \frac{tp}{1-t(1-p)}$$

$$\mathbb{E}[X] = g'_X(1) = \left( \frac{p(1-t(1-p)) + tp(1-p)}{(1-t(1-p))^2} \right) \Big|_{t=1} = 1 + \frac{p(1-p)}{p^2} \\ = 1 + \frac{1-p}{p} = \frac{1}{p}$$

Dalla definizione:  $\mathbb{E}[X] = \sum_{n=1}^{\infty} np(1-p)^{n-1} = p \frac{d}{dp} \left( - \sum_{n=0}^{\infty} (1-p)^n \right) = \dots$

$$\mathbb{E}[X(X-1)] = g''_X(1) = \left( \frac{p(1-p)}{(1-t(1-p))^2} + \frac{p(1-p)}{(1-t(1-p))^2} + \frac{2tp(1-p)^2}{(1-t(1-p))^3} \right) \Big|_{t=1}$$

$\mathbb{E}[X^2] - \mathbb{E}[X]$

$$= \frac{1-p}{p} + \frac{1-p}{p} + \frac{2(1-p)^2}{p^2} = 2 \left( \frac{p^2(1-p) + (1-p)^2}{p^2} \right)$$

$$\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = \dots$$

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

4. Poisson ( $\lambda$ )  $p_x(n) = e^{-\lambda} \frac{\lambda^n}{n!} \quad n \geq 0 \quad X \sim \text{Poisson}(\lambda)$

$$g_x(t) = \mathbb{E}[t^X] = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} t^n = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} = e^{\lambda(t-1)}$$

$$\mathbb{E}[X] = g'_x(1) = (\lambda e^{\lambda(t-1)})|_{t=1} = \lambda$$

$$\mathbb{E}[X(X-1)] = g''_x(1) = (\lambda^2 e^{\lambda(t-1)})|_{t=1} = \lambda^2$$

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned}$$

funzione generatrice dei momenti

$X$  v.a. reale. la funzione generatrice dei momenti

$$\phi_x: \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\} \quad \phi_x(t) = \mathbb{E}[e^{tx}]$$

osservazione  $\phi_x(0) = 1$ . Ma può accadere che  $\phi_x(t) = +\infty \quad \forall t \neq 0$

proprietà Se  $\phi_x < \infty$  in un intorno di 0

- $\phi_x$  analitica (in ogni aperto dove assume valori finiti)
- $\phi_x^{(n)}(0) = \mathbb{E}[X^n]$  per ogni  $n \geq 0$
- $X, Y$  indipendenti, allora  $\phi_{X+Y}(t) = \phi_x(t) \phi_y(t)$

\* date  $X, Y$  tali che  $\phi_x, \phi_y$  finite in un intorno di 0

$\phi_x = \phi_y$  in un intorno di 0  $\Leftrightarrow X, Y$  hanno la stessa legge

dim Mostriamo la seconda proprietà nel caso discreto

$$(\cdot) \quad \phi_x(t) = \sum_x e^{tx} p_x(x) \leftarrow$$

$$\phi_x^{(n)}(t) = \sum_x x^n e^{tx} p_x(n) \leftarrow$$

Converge in un intorno (eventualmente più piccolo di  $(\cdot)$ ) di 0

$$\phi_x^{(n)}(0) = \sum_x x^n p_x(n) = \mathbb{E}[X^n]$$

Osservazione la soluzione al problema è definire la funzione caratteristica

$$\varphi_x(t) = \mathbb{E}[e^{itx}]$$

## TEOREMI LIMITE

disuguaglianza di Markov

Se  $X$  è una v.a. su  $(\Omega, \mathcal{P})$  con  $X \geq 0$  con prob 1, allora

per ogni  $a > 0$   $\mathbb{P}[X \geq a] \leq \frac{1}{a} \mathbb{E}[X]$

dim Ricordiamo  $\mathbb{1}_A$  per un evento  $A$

$$\mathbb{1}(w) = \begin{cases} 1 & \text{se } w \in A \\ 0 & \text{se } w \notin A \end{cases} \quad \mathbb{1}_A : \Omega \rightarrow \mathbb{R}$$

funzione indicatrice dell'evento  $A$ . Se  $p$  è la densità discreta di  $\mathbb{P}$

$$\mathbb{E}[\mathbb{1}_A] = \sum_{w \in \Omega} \mathbb{1}_A(w) p(w) \quad p(w) = \mathbb{P}(\{w\})$$

$$= \sum_{w \in A} p(w) = \mathbb{P}(A)$$

Osserviamo che, data  $a > 0$   $a \mathbb{1}_{\{X \geq a\}} \leq X$

Infatti se  $w \in \Omega$

$$a \mathbb{1}_{\{X \geq a\}}(w) = \begin{cases} 0 & \text{se } X(w) < a \\ a & \text{se } X(w) \geq a \end{cases} \leq X(w)$$

Quindi  $\mathbb{E}[a \mathbb{1}_{\{X \geq a\}}] \leq \mathbb{E}[X]$

$$a \mathbb{P}[X \geq a]$$

## disuguaglianza di Chebychev

Se  $X$  è una v.a. su  $(\Omega, \mathbb{P})$  con momento secondo finito, allora per ogni  $a > 0$

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

dim. Applichiamo la disug. di Markov alla v.a.  $|X - \mathbb{E}[X]|^2$

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq a] = \mathbb{P}[|X - \mathbb{E}[X]|^2 \geq a^2]$$

in quanto

$$\{\omega: |X(\omega) - \mathbb{E}[X]| \geq a\} = \{\omega: (X(\omega) - \mathbb{E}[X])^2 \geq a^2\}$$

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

□

## Convergenza in probabilità

Siano date una successione  $(X_n)_{n \geq 0}$  di v.a., e una v.a.  $X$ , definite su  $(\Omega, \mathbb{P})$

Si dice che  $(X_n)_{n \geq 0}$  converge in probabilità a  $X$  se:

$$\text{per ogni } \varepsilon > 0 \quad \lim_{n \rightarrow \infty} \mathbb{P}[|X_n - X| \geq \varepsilon] = 0$$

## legge (debole) dei grandi numeri

Se per ogni  $n$  sono date  $X_1, \dots, X_n$  indipendenti e con uguale distribuzione (indipendente da  $n$ ), che ammettano momento secondo, detti  $\mu = \mathbb{E}[X_1]$ ,  $\sigma^2 = \text{Var}(X_1)$ , e posto  $S_n = X_1 + \dots + X_n$

allora  $\frac{S_n}{n} \xrightarrow{\text{in probabilità}} \mu$  ↪ finita

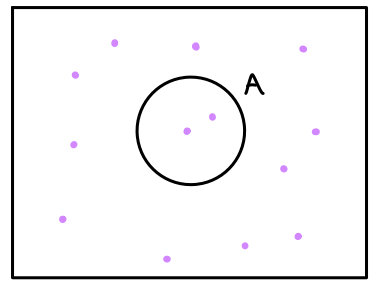
$$\uparrow$$
$$\frac{X_1 + \dots + X_n}{n}$$

media empirica → da non confondere con la media (valore atteso)

↪ ci permette di calcolare il valore atteso

Osservazione Questa è esattamente la legge dei grandi numeri già vista, con la differenza che questa vale per qualunque distribuzione

Supponiamo di avere un riquadro con un bersaglio, vogliamo calcolare l'area di questo bersaglio. Come facciamo? Effettuiamo un certo numero  $n$  di tiri.



metodo Montecarlo

Ci sono due possibilità: colpire l'area  $A$  del bersaglio (successo) oppure colpire l'area  $A^c$  (insuccesso).

Per  $n \rightarrow \infty$  notiamo che il numero di successi tende all'area del bersaglio.

Abbiamo appena descritto il metodo Montecarlo.

Osservazione la probabilità di un evento noi la calcoliamo come frequenza delle volte in cui questo evento accade.

Osservazione fissata una legge di densità  $p_0$  per ogni  $n$  costruiamo  $X_1, \dots, X_n$  indipendenti e tutte con leggi di densità  $p_0$

$$\text{per ogni } t > 0 \quad \mathbb{P} \left[ \left| \frac{X_1 + \dots + X_n}{n} - m \right| \geq \varepsilon \right] \rightarrow 0$$

dim fissato  $\varepsilon > 0$ , applichiamo la disuguaglianza di Chebychev a  $S_n - \mathbb{E}[S_n]$ . Infatti

$$\frac{S_n}{n} - m = \frac{S_n}{n} - \mathbb{E}(X_1) = \frac{1}{n} (S_n - n \mathbb{E}(X_1)) = \frac{S_n - \mathbb{E} S_n}{n}$$

$$\begin{aligned} \mathbb{P} \left[ \left| \frac{S_n}{n} - m \right| \geq \varepsilon \right] &= \mathbb{P} \left[ |S_n - \mathbb{E}(S_n)| \geq \varepsilon n \right] \quad \text{chebychev} \\ &\leq \frac{1}{\varepsilon^2 n^2} \text{Var}(S_n) \end{aligned}$$

$$\begin{aligned} \text{Var}(S_n) &= \text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n) \\ &= n \sigma^2 \end{aligned}$$

quindi

$$\mathbb{P} \left[ \left| \frac{S_n}{n} - m \right| \geq \varepsilon \right] \leq \frac{1}{\varepsilon^2 n^2} n \sigma^2 = \frac{\sigma^2}{\varepsilon^2 n} \rightarrow 0 \quad \text{per } n \rightarrow \infty$$

□

Corollario la legge dei grandi numeri continua a valere nella formulazione del teorema precedente se le v.a. sono solo correlate

Passiamo al teorema del limite centrale

Riformuliamo De Moivre-Laplace con il linguaggio delle v.a.

$X_1, \dots, X_n$  v.a. di Bernoulli ( $p$ )

$S_n = X_1 + \dots + X_n$  numero di successi

$$\bullet \mathbb{E}[S_n] = np \quad \text{Var}(S_n) = np(1-p) \quad \text{sd}(S_n) = \sqrt{np(1-p)}$$

$$\mathbb{P}[np + a\sqrt{np(1-p)} \leq S_n \leq np + b\sqrt{np(1-p)}] \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}y^2} dy$$

Possiamo scriverlo come

$$\mathbb{P}\left[a \leq \frac{S_n - \mathbb{E}[S_n]}{\text{sd}(S_n)} \leq b\right] \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}y^2} dy$$

### TEOREMA DEL LIMITE CENTRALE

indipendenti  
da  $n$

Se  $X_1, \dots, X_n$  sono v.a. indipendenti con ugual distribuzione di media comune  $m$  e varianza comune  $\sigma^2$ , posto  $S_n = X_1 + \dots + X_n$  per ogni  $a < b$

$$\mathbb{P}\left[a \leq \frac{S_n - \mathbb{E}[S_n]}{\text{sd}(S_n)} \leq b\right] \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}y^2} dy$$

---

### Chiarimento sul valore atteso condizionato:

legge condizionale  $P_{X|Y}(x) = \mathbb{P}[Y=y | X=x]$

$P_{Y|X}(\cdot | x)$  è una densità "p di  $y$  dato  $x$ "

$$\bullet P_{Y|X}(\cdot | x) \geq 0$$

$$\bullet \sum_y P_{Y|X}(y | x) = \sum_y \mathbb{P}[Y=y | X=x]$$

$$= \mathbb{P}\left[\bigcup_y \{Y=y\} | X=x\right] = 1$$

Il valore atteso di  $Y$  condizionato a  $X=x$  è il valore atteso rispetto alla legge condizionale

$$E[Y|X=x] = \sum_y y P_{Y|X}(y|x)$$

Detto questo, osserviamo che abbiamo una funzione

$$x \mapsto E[Y|X=x] \quad \text{per } x \text{ tale che } P[X=x] > 0$$

Valore atteso condizionato

$$E[Y|X] = \sum_x E[Y|X=x] \mathbb{1}_{\{X=x\}}$$

$$E[Y|X] \text{ è una v.a. } E[Y|X]: \Omega \rightarrow \mathbb{R}$$

$$\omega \in \Omega \longrightarrow X(\omega) \longrightarrow E[Y|X](\omega) = E[Y|X=x]$$

$$E[Y|X](\omega) = \sum_x E[Y|X=x] \mathbb{1}_{\{X=x\}}(\omega) = \begin{cases} 1 & \text{se } \omega \in \{X=x\} \rightarrow X(\omega)=x \\ 0 & \text{se } \omega \notin \{X=x\} \rightarrow X(\omega) \neq x \end{cases}$$

tutte = 0 tranne che  
per  $x = X(\omega)$

lezione 12  
14-04-2021  
(Rovito)

TEOREMA DEL LIMITE CENTRALE  $\rightarrow$  lo si ritrova negli esperimenti fisici

Se  $X_1, \dots, X_n$  sono v.a. indipendenti con ugual distribuzione di media comune  $m$  e varianza comune  $\sigma^2$ , posto  $S_n = X_1 + \dots + X_n$

per ogni  $a < b$

$\rightarrow$  dipenderà dalla legge di partenza

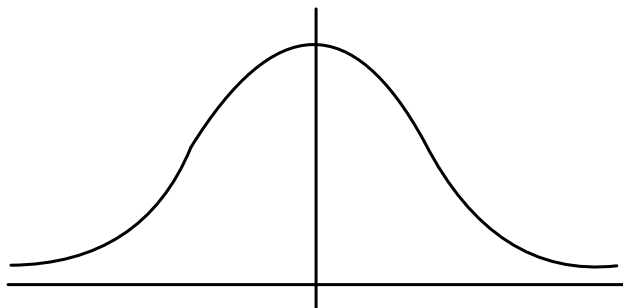
$$P\left[a \leq \frac{S_n - E[S_n]}{\text{sd}(S_n)} \leq b\right] \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}y^2} dy$$

$\rightarrow$  tende a crescere  
 $\rightarrow \text{sd}(S_n) = \sigma\sqrt{n}$

alla fine troviamo  
la distribuzione  
Gaussiana

Osservazioni

1. Estensione del teorema di De Moivre - Laplace
2. Il Teorema del limite centrale si riferisce a una legge
3. Il TLC ha valore universale



distribuzione gaussiana

## Osservazione

LGN:  $\frac{S_n}{n} \rightarrow m$  (in probabilità)

$$\text{TLC: } \frac{S_n - \mathbb{E}[S_n]}{\text{sd}(S_n)} = \frac{S_n - m \cdot n}{\sigma \sqrt{n}} = \sqrt{n} \frac{\frac{S_n}{n} - m}{\sigma}$$

$\sqrt{n} \frac{\frac{S_n}{n} - m}{\sigma} \rightarrow$  gaussiana velocità di converg. al più  $\frac{1}{\sqrt{n}}$  (lenta)

Montecarlo per quanto abbia un tasso di convergenza "povero" è un metodo che funziona perché indipendente dalla regolarità ecc.

Un altro tasso di convergenza

$$P\left[\left|\frac{S_n}{n} - m\right| \geq \varepsilon\right] \leq O\left(\frac{1}{n}\right)$$

il tasso di convergenza non è migliorabile

$$\varepsilon = \frac{1}{n^\alpha}$$

questo però è migliorabile

$$\left(\frac{\sigma^2}{\varepsilon^2 n}\right) \frac{\sigma^2}{n^{1-2\alpha}}$$

teorema (disuguaglianza di concentrazione)

Se  $X_1, X_2, \dots, X_n$  sono indipendenti e con legge di Bernoulli(p)

$p \in (0, 1)$  posto  $S_n = X_1 + \dots + X_n$ , allora per ogni  $\varepsilon > 0$  esiste

$H(p, \varepsilon) > 0$  tale che

$$P\left[\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right] \leq 2e^{-nH(p, \varepsilon)}$$

dim  $L(s) = \mathbb{E}[e^{sX_1}] = (1-p) + pe^s$

Per indipendenza  $\mathbb{E}[e^{sS_n}] = L(s)^n$

$$\left\{\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right\} = \left\{\frac{S_n}{n} \geq p + \varepsilon\right\} \cup \left\{\frac{S_n}{n} \leq p - \varepsilon\right\}$$

$$P\left[\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right] \leq P\left[\frac{S_n}{n} \geq p + \varepsilon\right] + P\left[\frac{S_n}{n} \leq p - \varepsilon\right]$$

Inoltre è sufficiente dimostrare il teorema per  $\varepsilon$  suff. piccolo

Consideriamo  $P[\frac{S_n}{n} \geq p + \varepsilon]$  e  $\varepsilon$  tale che  $p + \varepsilon < 1$

&  $a \in (0, 1)$ . Dalla disug. di Markov, per ogni  $s > 0$

$$P\left[\frac{S_n}{n} \geq a\right] = P\left[e^{s(\frac{S_n}{n} - a)} \geq 1\right] \leq E\left[e^{s(\frac{S_n}{n} - a)}\right] \\ \leq e^{-as} L\left(\frac{s}{n}\right)^n = e^{-n(a \frac{s}{n} - \log L(\frac{s}{n}))}$$

Per  $t > 0$  scegliamo  $s = nt$  in modo che

$$P\left[\frac{S_n}{n} \geq a\right] \leq e^{-n(at - \log L(t))} \\ \leq e^{-nL^*(a)}$$

con  $L^*(a) = \sup_{t>0} (at - \log L(t))$

Dobbiamo dimostrare che  $L^*(a) > 0$

Consideriamo  $\varphi(t) = at - \log L(t) = at - \log((1-p) + pe^t)$

$$\lim_{t \rightarrow -\infty} \varphi(t) = -\infty \quad \lim_{t \rightarrow +\infty} \varphi(t) = -\infty$$

•  $\varphi(0) = 0$

•  $\varphi' = a - \frac{L'}{L} \quad L' = L'' = pe^t > 0$

$$\varphi'' = -\frac{L''}{L} + \frac{(L')^2}{L^2} = \underbrace{\frac{L'}{L^2}}_{>0} \underbrace{(L' - L)}_{pe^t - (1-p+pe^t) = -(1-p)} < 0$$

$\varphi'' < 0$  ovvero  $\varphi$  concava

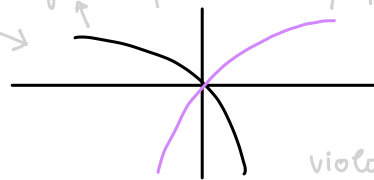
•  $\varphi'(0) = a - \frac{L'(0)}{L(0)} = a - p$

si verifica se  $\varphi'(0) < 0$

si verifica se  $\varphi'(0) > 0$

La funzione deve fare 0 in 0 e deve avere un solo max.

Noi vorremmo la funzione viola, che ha max positivo.



Se  $a > p$ ,  $\varphi$  ha massimo con valore positivo. In conclusione

$$L^*(p + \varepsilon) > 0 \text{ e } P\left[\frac{S_n}{n} \geq p + \varepsilon\right] \leq e^{L^*(p + \varepsilon) \cdot n}$$

Passiamo al termine  $P\left[\frac{S_n}{n} \leq p + \varepsilon\right]$

$$y_i = 1 - x_i \quad i = 1, \dots, n$$

$Y_1, \dots, Y_n$  indipendenti e di Bernoulli di parametro  $1-p$

$$S_n' = Y_1 + \dots + Y_n = n - S_n$$

$$\frac{S_n'}{n} = 1 - \frac{S_n}{n}$$

$$\mathbb{P}\left[\frac{S_n}{n} \leq p + \varepsilon\right] = \mathbb{P}\left[\frac{S_n'}{n} \geq (1-p) + \varepsilon\right] \leq e^{-nL^*(1-p+\varepsilon)}$$

dai conti precedenti, poiché  $1-p+\varepsilon < 1$

Concludiamo

Se  $\varepsilon > 0$  è tale che  $0 < \varepsilon < \min(p, 1-p)$  e se

$$H(p, \varepsilon) = \min(L^*(p+\varepsilon), L^*(1-p+\varepsilon)) > 0$$

allora  $\mathbb{P}[|\frac{S_n}{n} - p| \geq \varepsilon] \leq e^{-nH(p, \varepsilon)}$

□

## MODELLI CONTINUI

$\sigma$ -algebra

Dato un insieme  $\Omega$ , un insieme  $\mathcal{F} \subset \mathcal{P}(\Omega)$  è una  $\sigma$ -algebra se

- $\emptyset, \Omega \in \mathcal{F}$
- se  $A \in \mathcal{F}$ , allora  $A^c \in \mathcal{F}$
- se  $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$ , allora  $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}$

probabilità Dato  $(\Omega, \mathcal{F})$ , una funzione  $\mathbb{P}: \mathcal{F} \rightarrow \mathbb{R}$  è una probabilità se

- $\mathbb{P}(\Omega) = 1$
- $0 \leq \mathbb{P}(A) \leq 1$  per ogni  $A \in \mathcal{F}$
- se  $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$  a due a due disgiunti  $\mathbb{P}(\bigcup_n A_n) = \sum_n \mathbb{P}(A_n)$

Osservazione Problema: definire una probabilità uniforme sull'intervallo  $[0, 1]$ .

Necessariamente i singoletti  $\{x\}$ ,  $x \in [0, 1]$  devono avere probabilità zero.

L'uniformità la recuperiamo richiedendo che range di valori di uguale ampiezza abbiano la stessa probabilità (ovvero

scegliamo come probabilità la misura di Lebesgue su  $[0,1]$

evento: qualunque elemento di  $\mathcal{F}$

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

Osservazione Cosa "funziona" di quanto fatto finora nel modello generale

- proprietà elementari (inclusione-esclusione, ecc...)
- densità finita discreta **NO**
- probabilità condizionata: definizione, formula di Bayes, formula di disintegrazione (nel caso di partizioni univ.  $\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$ )
- continuità dall'alto/basso

$(\Omega, \mathcal{F}) \quad \mathbb{P}: \mathcal{F} \rightarrow \mathbb{R} \quad \text{t.c.}$

- $\mathbb{P}(\Omega) = 1$

- $0 \leq \mathbb{P}(A) \leq 1$  per ogni  $A \in \mathcal{F}$

- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$  per ogni  $A, B \in \mathcal{F}$ , con  $A \cap B = \emptyset$

allora sono equivalenti

- $\sigma$ -additività: per ogni  $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$  a due a due disgiunti  $\mathbb{P}(\bigcup A_n) = \sum \mathbb{P}(A_n)$

- continuità dall'alto:  $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$ ,  $A_{n+1} \subset A_n \dots \mathbb{P}(\bigcap A_n) \dots$

- continuità dal basso:  $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$ ,  $A_n \subset A_{n+1} \dots \mathbb{P}(\bigcup A_n) \dots$

- indipendenza:  $(A_i)_{i \in I} \subset \mathcal{F}$  eventi indipendenti se

$$\forall J \subseteq I \text{ finito si ha } \mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j)$$

Eserciz. 7

19/04/2021

(Di Genà)

### Problema 1.

Sia dato un numero  $n \geq 1$ . Si lancia  $n$  volte un dado regolare e siano  $X_1, X_2, \dots, X_n$  i risultati. Si dice che avviene un calo all'estrazione  $k \geq 2$  se  $X_k < X_{k-1}$ . Sia  $D_n$  il numero di cali nei lanci effettuati.

- Calcolare media e varianza di  $D_n$ .
- Calcolare la probabilità che non ci siano cali.

(i) Trucco: Usare la linearità

$$\text{Sia } C_K = \begin{cases} 1 & \text{se } X_K < X_{K-1} \\ 0 & \text{altrimenti} \end{cases} \quad K=2, \dots, n$$

$$\mathbb{E}[D_n] = \sum_{K=2}^n \mathbb{E}[C_K] = \sum_{K=2}^n P(C_K = 1)$$

$$P(C_K = 1) = \frac{\frac{6-5}{2}}{36} = \frac{15}{36} = \frac{5}{12}$$

$$\text{Quindi } \mathbb{E}[D_n] = \frac{5(n-1)}{12}$$

$$\text{Var}(D_n) = \mathbb{E}[D_n^2] - \mathbb{E}[D_n]^2 =$$

$$\begin{aligned} &= \sum_{K,j=2}^n \mathbb{E}[C_K C_j] - \frac{25(n-1)^2}{144} \\ &= \sum_{K=2}^n \mathbb{E}[C_K^2] + 2 \sum_{K=2}^{n-1} \sum_{j=K+1}^n \mathbb{E}[C_K C_j] - \frac{25(n-1)^2}{144} \\ &= \frac{5(n-1)}{12} - \frac{25(n-1)^2}{144} + 2 \sum_{K=2}^{n-1} \sum_{j=K+1}^n \mathbb{E}[C_K C_j] \end{aligned}$$

Definiamo 2 casi:

Sia  $K = 2, \dots, n-1$

$$j = K+2, \dots, n$$

allora  $\mathbb{E}[C_K C_j] = \mathbb{E}[C_K] \mathbb{E}[C_j] = \frac{25}{144}$  perché  $C_K C_j$  sono indipendenti

se invece  $j = K+1$

$$\mathbb{E}[C_K C_{K+1}] = P(C_K = 1, C_{K+1} = 1) = \frac{\frac{6 \cdot 5 \cdot 4}{3!}}{216} = \frac{20}{216} = \frac{5}{54}$$

$$\text{Var}(D_n) = \dots = \frac{35}{432} (n+1)$$

$$(ii) \quad P(D_n = 0) \quad \text{es: } \begin{array}{c} 11 \dots 1 \\ 11 \dots 2 \dots 5 \dots 6 \dots \\ 33 \dots 44 \dots 55 \dots \end{array}$$

Sia  $A_K = \text{"escono esattamente } K \text{ numeri distinti"}$

$$P(D_n = 0) = \sum_{K=1}^6 P(D_n = 0, A_K)$$

$$P(D_n = 0, A_k) = \frac{\binom{6}{k} \binom{n-1}{k-1}}{6^n}$$

$$\Rightarrow P(D_n = 0) = \frac{\sum_{k=0}^5 \binom{n-1}{k} \binom{6}{k+2}}{6^n} = \frac{\sum_{k=0}^5 \binom{n-1}{k} \binom{6}{5-k}}{6^n} = \frac{\binom{n+5}{5}}{6^n}$$

idea 2:  $n = n_1 + \dots + n_6 \quad n_k \geq 0$

### Problema 3.

Siano  $X, Y$  variabili aleatorie indipendenti e a valori in  $N$ , e  $p \in (0, 1)$  tali che per ogni  $n, k \geq 0$ , con  $k \leq n$ ,

$$P[X=k | X+Y=n] = \binom{n}{k} p^k (1-p)^{n-k}$$

- Calcolare  $E[s^X t^Y | X+Y=n]$  per ogni  $s, t > 0$  e  $n \in N$ .
- Esprimere  $E[s^X t^Y]$  in termini della funzione generatrice delle probabilità di  $X+Y$ .
- Dimostrare che  $X, Y$  hanno entrambe distribuzione di Poisson.

$$P[X=k | X+Y=n] = \binom{n}{k} p^k (1-p)^{n-k}$$

$$(i) \quad E[s^X t^Y | X+Y=n] = \sum_{k=0}^n s^k t^{n-k} P[X=k | X+Y=n] =$$

$$= \sum_{k=0}^n \binom{n}{k} (sp)^k [t(1-p)]^{n-k} = [sp + t(1-p)]^n$$

$$(ii) \quad E[s^X t^Y] = \sum_{n=0}^{\infty} E[s^X t^Y | X+Y=n] P[X+Y=n] \quad \text{fatto generale}$$

$$G_{X+Y}(r) = E[r^{X+Y}] = \sum_{n=0}^{\infty} r^n P(X+Y=n)$$

$$\Rightarrow E[s^X t^Y] = G_{X+Y}(sp + t(1-p)) \quad \text{da sostituire al posto di } r$$

$$(iii) \quad G_X(s) G_Y(t)$$

$$\stackrel{||}{=} E[s^X 1^Y] E[t^Y 1^X] = G_{X+Y}(sp + t(1-p)) G_{X+Y}(p + t(1-p))$$

$$G_{X+Y}(sp + t(1-p)) = G_{X+Y}(sp + t(1-p)) G_{X+Y}(p + t(1-p)) \quad \text{uguaglianza funzionale}$$

Mostriamo che  $X+Y$  è una poissoniana, ovvero che

$$\exists \lambda > 0: G_{X+Y}(r) = e^{\lambda(r-1)} \longrightarrow F(r) = G(r+1) = e^{\lambda r}$$

$$\begin{cases} F(a+b) = F(a) F(b) \\ F(0) = 1 \end{cases}$$

$$\Rightarrow \exists \lambda \in \mathbb{R} : F(r) = e^{\lambda r}$$

$$\Rightarrow G(r) = e^{\lambda(r-1)} \quad \text{e quindi per forza } \lambda \geq 0 \rightarrow \text{si trova facilmente un assurdo}$$

$$G_X(s) = \mathbb{E} s^X = G_{X+Y}(s p + 1 - p) = G_{X+Y}((s-1)p + 1) = e^{\lambda p(s-1)}$$

$\Rightarrow X$  è di Poisson di parametro  $\lambda p$

$$G_Y(t) = \dots$$

### Problema 10.

Una scatola contiene  $n$  palline rosse e  $n$  palline blu. Le palline vengono estratte (senza reinserimento) due per volta, fino a svuotare la scatola. Sia  $X$  il numero di volte in cui le due palline estratte sono di colore differente.

(i) Determinare tutti i possibili valori di  $k$  per cui  $P(X = k) > 0$ .

(ii) Calcolare la probabilità che  $X = 0$  e che  $X = n$ .

(iii) Calcolare il valore atteso di  $X$ .

(i) determinare  $k$  :  $P(X = k) > 0$

$n$  pari  $k = 0, 2, 4, \dots, n$

$n$  dispari  $k = 1, 3, \dots, n$

(ii)  $P(X = 0) = 0$  se  $n$  dispari

Assumiamo  $n$  pari

$$\text{Allora } P(X = 0) = \frac{\binom{n}{n/2}}{\binom{2n}{n}} = \frac{\frac{n!}{(\frac{n}{2})! (\frac{n}{2})!}}{\frac{(2n)!}{n! n!}} = \frac{(n!)^3}{(\frac{n}{2})!^2 (2n)!}$$

$$P(X = n) = \frac{2^n}{\binom{2n}{n}} = \frac{2^n n! n!}{(2n)!}$$

(iii)  $\mathbb{E}[X] = ?$

Sia  $z_k = \begin{cases} 1 & \text{se alla } k\text{-esima estrazione ne escano due diverse} \\ 0 & \text{altrimenti} \end{cases}$

$$X = \sum_{k=1}^n z_k \Rightarrow \mathbb{E}[X] = \sum_{k=1}^n \mathbb{E}[z_k] = \sum_{k=1}^n P(z_k = 1)$$

$$P(z_k = 1) = \frac{2 \binom{2n-2}{n-1}}{\binom{2n}{n}} = \frac{2 (2n-2)! n! n!}{(2n)! (n-1)! (n-1)!} = \frac{2 n^2 (2n-2)!}{(2n)!} = \frac{n}{2n-1}$$

$$\Rightarrow E[z_k] = \frac{n}{2n-1} \Rightarrow E[X] = \frac{n^2}{2n-1}$$

## Problema 2.

Un sacchetto contiene  $r$  gettoni rossi. Si inserisce nel sacchetto un gettone blu e poi si estrae (con reinserimento) un gettone. Si ripete questo procedimento (di inserire cioè un gettone blu e poi estrarre un gettone) fino a quando si estrae un gettone rosso. Detto  $N$  il numero di estrazioni fatte,

- mostrare che  $P[N = \infty] = 0$ ;

## NOTA BENE

$$\lim_{n \rightarrow \infty} P(N = k) = 0 \not\Rightarrow P(N = \infty) = 0$$

↓  
Poiché  $P(\Omega) = 1$  questa cosa è ovvia

$$N: \Omega \rightarrow \mathbb{N} \cup \{\infty\}$$

idea 1:

$$\sum_{k=1}^{\infty} P(N = k) = 1 \Rightarrow P(N = \infty) = 1 - P\left(\bigcup_{k=1}^{\infty} \{N = k\}\right) = 0$$

idea 2:

$$P(N = \infty) \leq P(N > k) \quad \forall k \quad \text{per } k \rightarrow \infty \rightarrow 0$$

Lezione 13  
20-04-2021  
(Romito)

$$(\Omega, \mathcal{F}, P) \quad (S, \mathcal{G}) \quad S = (\mathbb{R}, \mathbb{R}^d, \text{insieme di simboli})$$

↓  
dove  $\mathcal{G}$  è una  $\sigma$ -algebra su  $S$ , tipicamente se

$S = \mathbb{R}^d \quad \mathcal{G} = \mathcal{B}(\mathbb{R}^d) = \sigma\text{-algebra dei Boreliani.}$

la più piccola  $\sigma$ -algebra che contiene gli aperti

$X: \Omega \rightarrow S$  è una v.a. se

$$X^{-1}(A) \in \mathcal{F} \quad \text{per ogni } A \in \mathcal{G}$$

(ovvero  $X$  è misurabile)

Se  $S = \mathbb{R}$ ,  $\mathcal{G} = \mathcal{B}(\mathbb{R})$ , allora sono equivalenti

- $X$  v.a.

- $X^{-1}((-\infty, x]) \in \mathcal{F}$  per ogni  $x \in \mathbb{R}$

## legge di una v.a.

$(\Omega, \mathcal{F}, P)$   $(S, \mathcal{S})$   $X: \Omega \rightarrow S$  v.a., la legge  $P_x$  di  $X$  è la misura di probabilità  $P_x: \mathcal{S} \rightarrow \mathbb{R}$  definita da

$$P_x[A] = P[X \in A] (= P[\{\omega \in \Omega : X(\omega) \in A\}]) \quad A \in \mathcal{S}$$

## Osservazione

- la legge  $P_x$  non può più essere caratterizzata dalla densità discreta
- costruzione canonica: se  $\tilde{P}$  probabilità su  $(S, \mathcal{S})$ , allora  $\Omega = S$ ,  $\mathcal{F} = \mathcal{S}$ ,  $\tilde{P} = \tilde{P}$   $X: \Omega \rightarrow S$  definita da  $X(\omega) = \omega$  è una v.a. (perché  $X^{-1}(A) = A$ ) con legge  $P_x = \tilde{P}$
- composizione di una v.a. con una funzione misurabile rimane una v.a.  $X$  v.a. su  $(\Omega, \mathcal{F}, P)$  a valori  $(S, \mathcal{S})$  e  $\varphi: S \rightarrow S'$   $(S', \mathcal{S}')$  misurabile (ovvero  $\varphi^{-1}(B) \in \mathcal{S}$  per ogni  $B \in \mathcal{S}'$ ). Se  $B \in \mathcal{S}'$   
$$\varphi(X)^{-1}(B) = X^{-1}(\varphi^{-1}(B)) \in \mathcal{F}$$

quindi  $\varphi(X)$  è una v.a.

- uguaglianza in legge:  $X=Y$  in legge se  $P_x = P_y$

La legge si può caratterizzare attraverso la funzione generatrice dei momenti  $(S=\mathbb{R})$   $\phi_x(t) = E[e^{tx}]$

- uguaglianza di v.a.:  $X=Y$  con probab. 1  $P[X=Y]=1$
- leggi congiunte marginali:  $X: \Omega \rightarrow S_1 \times S_2$   $(S_1, \mathcal{S}_1)$   $(S_2, \mathcal{S}_2)$   
 $X = (X_1, X_2)$   $P_x$  legge congiunta  
 $P_{x_1}[A] = P_x[A \times S_2]$   $A \in \mathcal{S}_1$

- indipendenza.  $(\Omega, \mathcal{F}, P)$   $(X_i)_{i \in I}$   $X_i: \Omega \rightarrow S_i$   $(S_i, \mathcal{S}_i)$   
 $(X_i)_{i \in I}$  indipendenti se per ogni  $J \subset I$  finite, per ogni  $A_i \in \mathcal{S}_i$   $i \in J$   
$$P\left[\bigcap_{i \in J} \{X_i \in A_i\}\right] = \prod_{i \in J} P[X_i \in A_i]$$

Funzione di ripartizione (o di distribuzione oppure funzione cumulata di probabilità)

$X$  v.a. reale su  $(\Omega, \mathcal{F}, \mathbb{P})$  la funzione di ripartizione  $F_X$  di  $X$  è

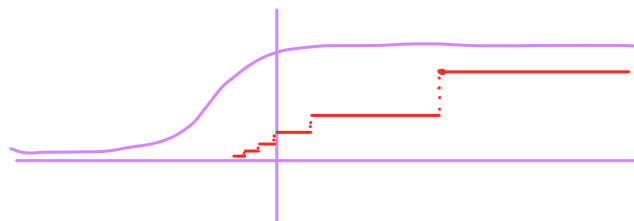
$$F_X(x) = \mathbb{P}[X \leq x] \quad x \in \mathbb{R}$$

proprietà  $X$  sia v.a. reale su  $(\Omega, \mathcal{F}, \mathbb{P})$

•  $F_X$  è monotona crescente

•  $F_X$  è continua a destra

$$\lim_{x \rightarrow +\infty} F_X(x) = 1 \quad \lim_{x \rightarrow -\infty} F_X(x) = 0$$



dim •  $x_1 \leq x_2 \quad (-\infty, x_1] \subset (-\infty, x_2] \rightarrow \{X \in (-\infty, x_1]\} \subset \{X \in (-\infty, x_2]\}$

$$F_X(x_1) = \mathbb{P}[X \in (-\infty, x_1]] \leq \mathbb{P}[X \in (-\infty, x_2]] = F_X(x_2)$$

• c.a.d. = per ogni  $x_0 \in \mathbb{R}_1 \quad \lim_{x \rightarrow x_0^+} F_X(x) = F_X(x_0)$

È sufficiente verificare che, se  $x_n \downarrow x_0$ , allora  $F_X(x_n) \downarrow F_X(x_0)$

Sia  $x_n \downarrow x_0 \quad A_n = \{X \in (-\infty, x_n]\}$ , allora  $A_{n+1} \subset A_n$

$$\mathbb{P}(\bigcap_n A_n) = \lim_n \mathbb{P}(A_n) = \lim_n \mathbb{P}[X \leq x_n] = \lim_n F_X(x_n)$$

$$\bigcap_n A_n = \{X \in (-\infty, x_0]\} \quad (\rightarrow F_X(x_0) = \mathbb{P}(\bigcap A_n)) \quad \text{infatti}$$

\*  $\omega \in \bigcap_n A_n \rightarrow X(\omega) \leq x_n$  per ogni  $n$  al limite  $X(\omega) \leq x_0$

\*  $X(\omega) \leq x_0 \rightarrow X(\omega) \leq x_n$  per ogni  $n \rightarrow \omega \in \bigcap A_n$

•  $\lim_{x \rightarrow +\infty} F_X(x)$ , è sufficiente mostrare che per ogni  $x_n \uparrow +\infty, F_X(x_n) \uparrow 1$

•  $\lim_{x \rightarrow -\infty} F_X(x) \quad x_n \downarrow -\infty \quad A_n = \{X \in (-\infty, x_n]\} \quad A_{n+1} \subset A_n = \emptyset \quad \square$

Teorema Se  $F: \mathbb{R} \rightarrow \mathbb{R}$  è una funzione tra le tre proprietà precedenti allora esiste uno spazio  $(\Omega, \mathcal{F}, \mathbb{P})$  e su di esso una v.a.  $X$  t.c.  $F_X = F$ . In particolare la funzione di ripartizione caratterizza la distribuzione.

dim È sufficiente costruire la legge, il resto segue dalla costruzione canonica. Bisogna definire una misura di probabilità su  $\mathbb{R}$

$$\mu((-\infty, x]) = F(x) \quad x \in \mathbb{R}$$

o anche

$$\mu((a, b]) = F(b) - F(a)$$

Ad esempio

$$F(x) = \begin{cases} 0 & x \leq 0 \\ x & x \in (0, 1) \\ 1 & x \geq 1 \end{cases}$$

in tal caso  $\mu_F((a, b]) = F(b) - F(a) = b - a \quad 0 \leq a < b \leq 1$

quindi  $\mu_F$  è la misura di Lebesgue su  $[0, 1]$

Osservazione  $F_X$  funzione di ripartizione,  $a < b$

$$\begin{aligned} F_X(b) - F_X(a) &= P[X \in (-\infty, b]] - P[X \in (-\infty, a]] \\ &= P[\{X \in (-\infty, b]\} \setminus \{X \in (-\infty, a]\}] \\ &= P[X \in (a, b]] \end{aligned}$$

Osservazione discontinuità della funzione di ripartizione

$X$  v.a.,  $F_X$  funzione di ripartizione, allora se  $x \in \mathbb{R}$

$$\lim_{y \rightarrow x^-} F_X(y) = P[X < x]$$

infatti è sufficiente verificare che, per ogni  $x_n \uparrow x$ ,  $F_X(x_n) \uparrow P[X < x]$

Sia  $A_n = \{X \in (-\infty, x_n]\}$   $A_n \subset A_{n+1}$  quindi

$$\lim_n F_X(x_n) = P[\bigcup_n A_n]$$

Ora però  $\bigcup_n A_n = \{X \in (-\infty, x)\}$

Osserviamo che se  $x$  è punto di discontinuità di  $F_X$

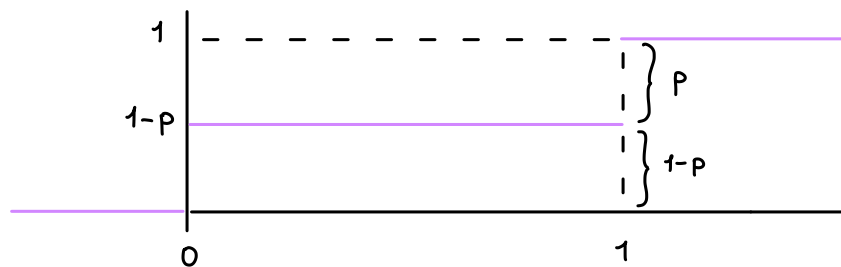
$$P[X < x] = \lim_{y \rightarrow x^-} F_X(y) < \lim_{y \rightarrow x^+} F_X(y) = P[X \leq x]$$

$$P[X \leq x] - P[X < x] = P[X = x]$$

La discontinuità di  $F_X$  sono i valori  $x \in \mathbb{R}$  t.c.  $P[X = x] > 0$

esempio  $X$  v.a. di Bernoulli di parametro  $p$ . Se  $x \in \mathbb{R}$

$$F_X(x) = P[X < x] = \begin{cases} 0 & x < 0 \\ 1-p & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$



funzione di ripartizione: caso multidimensionale

Se  $X$  è una v.a. su  $\mathbb{R}^d$   $X = (X_1, \dots, X_d)$

la funzione di ripartizione  $F_X: \mathbb{R}^d \rightarrow \mathbb{R}$  definita come

$$F_X(x) = \mathbb{P}[X_1 \leq x_1, X_2 \leq x_2, \dots, X_d \leq x_d]$$

per ogni  $x = (x_1, \dots, x_d)$

Valgono le proprietà

- $F_X$  monotonica crescente rispetto a ogni variabile
- $F_X$  è continua a destra rispetto a ogni variabile
- per ogni  $i = 1, \dots, d$   $\lim_{x_i \rightarrow -\infty} F_X(x_1, \dots, x_d) = 0$
- per ogni  $i = 1, \dots, d$   $\lim_{x \rightarrow +\infty} F_X(x_1, \dots, x_d) = F_{X^{(i)}}(x^{(i)})$   
dove  $X^{(i)} = (X_1, X_2, \dots, X_{i-1}, X_{i+1}, \dots, X_d)$  e  
 $x^{(i)} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d)$
- $F_X$  identifica la distribuzione, nel senso che se  $X, Y$  v.a.

$$\mathbb{P}_X = \mathbb{P}_Y \Leftrightarrow F_X = F_Y$$

proposizione  $X, Y$  v.a. reali allora

$X, Y$  indipendenti  $\Leftrightarrow F_{(X,Y)}(x,y) = F_X(x)F_Y(y)$  per ogni  $x, y \in \mathbb{R}$

dim Se  $X, Y$  sono indipendenti, per ogni  $x, y \in \mathbb{R}$

$$\begin{aligned} F_{(X,Y)}(x,y) &= \mathbb{P}[X \leq x, Y \leq y] = \mathbb{P}[X \leq x] \mathbb{P}[Y \leq y] \\ &= F_X(x) F_Y(y) \end{aligned}$$

Viceversa costruiamo  $(X', Y')$  indipendenti e tali che  $X' = X, Y' = Y$  in legge.

In particolare  $F_{X'} = F_X$  e  $F_{Y'} = F_Y$ . Per quanto dimostrato prima

$$F_{(x', y')} = F_{x'} F_{y'} = F_x F_y = F_{(x, y)}$$

quindi  $(X, Y) = (X', Y')$  in legge, quindi  $X, Y$  indipendenti

□

## VALORE ATTESO

$(\Omega, \mathcal{F}, \mathbb{P})$   $X$  v.a. reale

• se  $X \geq 0$ , allora si definisce

$$\mathbb{E}[X] = \sup \{ \mathbb{E}[Y] : Y \text{ v.a. reale discreta t.c. } 0 \leq Y \leq X \}$$

• se poi  $\mathbb{E}[|X|] < \infty$ , posto  $X_+ = \max(X, 0)$  e  $X_- = \max(-X, 0)$

$$\mathbb{E}[X] = \mathbb{E}[X_+] - \mathbb{E}[X_-]$$

(si ha  $|X| = X_+ + X_-$  e  $X = X_+ - X_-$ )

Valgono le proprietà

• omogeneità  $\mathbb{E}[aX] = a \mathbb{E}[X]$  (dati  $X$  integrabile e  $a \in \mathbb{R}$ , anche  $a \cdot X$  integrabile e vale l'uguaglianza)

• linearità  $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$  ( $X, Y$  integrabile, lo è anche  $X + Y$  e vale l'uguaglianza)

• positività: •  $X \leq Y$  con probabilità 1, allora  $\mathbb{E}[X] \leq \mathbb{E}[Y]$

•  $X \geq 0$  con probabilità 1 e  $\mathbb{E}[X] = 0$ , allora

$X = 0$  con probabilità 1

• varianza, covarianza, correlazione

• se  $X, Y$  indipendenti e integrabili/positive, allora  $X \cdot Y$  è integrabile/positivo e

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \mathbb{E}[Y]$$

Osservazione Non valgono le proprietà associate alla densità discreta

Osservazione Ripercorrendo la dimostrazione della LGN si vede che queste funzionano anche in questo contesto

$$\text{Markov: } a \mathbb{1}_{\{X \geq a\}} \leq X \quad \mathbb{E}[\mathbb{1}_{\{X \geq a\}}] = \mathbb{P}[X \geq a]$$

Il teorema del limite vale nel contesto generale

## VARIABILI ALEATORIE (ASSOLUTAMENTE) CONTINUE

densità (continua) Una densità continua è una funzione  $f: \mathbb{R} \rightarrow \mathbb{R}$  tale che:

- $f \geq 0$
- $\int_{\mathbb{R}} f(x) dx = 1$

Una v. a.  $X$  è assolutamente continua se esiste una densità continua  $f_x$  t.c.

$$P[X \in A] = \int_A f_x(x) dx$$

In tal caso, se  $A = (-\infty, x]$

$$F_x(x) = P[X \in (-\infty, x]] = \int_{-\infty}^x f_x(x) dx$$

e se poi  $a < b$   $A = [a, b]$  (o  $(a, b]$ ,  $[a, b)$ ,  $(a, b)$ )

$$F_x(b) - F_x(a) = \int_a^b f_x(x) dx$$

osservazioni: 1. se  $X$  è ass. continua,  $P[X=x]=0$  per ogni  $x \in \mathbb{R}$

2. sotto opportune condizioni  $F_x' = f_x$

lezione 14  
21-04-2021  
(Rovito)

## VARIABILI ALEATORIE (ASSOLUTAMENTE) CONTINUE

$X: \Omega \rightarrow \mathbb{R}$  è ass. continua se esiste  $f_x: \mathbb{R} \rightarrow \mathbb{R}$  (densità) tale che

$$P[X \in A] = \int_A f_x(x) dx$$

In particolare

$$F_x(x) = \int_{-\infty}^x f_x(x) dx$$

Se  $a < b$   $F_x(b) - F_x(a) = \int_a^b f_x(x) dx$

$F_x$  continua

$f_x$  è una densità se

- $f_x(x) \geq 0$  per q.o.  $x \in \mathbb{R}$
- $\int_{\mathbb{R}} f_x(x) dx = 1$

### Caso multidimensionale

Una densità è una funzione  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  tale che

- $f(x) \geq 0$  per q.o.  $x \in \mathbb{R}^d$

$$\int_{\mathbb{R}^d} f(x) dx = 1$$

Una v.a.  $X: \Omega \rightarrow \mathbb{R}^d$  è ass. continua se esiste una densità  $f_x$  t.c.

$$\mathbb{P}[X \in A] = \int_A f_x(x) dx \quad A \in \mathcal{B}(\mathbb{R}^d)$$

In particolare se  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$

$$\begin{aligned} F_x(x) &= \mathbb{P}[X_1 \leq x_1, \dots, X_d \leq x_d] \\ &= \mathbb{P}[X \in (-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_d)] \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_d} \end{aligned}$$

Si ottiene

$$\partial_{x_1} \dots \partial_{x_d} F_x(x_1, \dots, x_d) = f_x(x_1, \dots, x_d)$$

proposizione  $X, Y$  v.a. con densità rispettivamente  $f_x, f_y$  allora  $X, Y$  indipendenti

se e solo se  $f_{(x,y)}(x,y) = f_x(x)f_y(y)$  per q.o.  $(x,y)$

In effetti  $X, Y$  sono indipendenti se e solo se  $f_{(x,y)}(x,y) = g(x)h(y)$

dim Se  $X, Y$  indipendenti

$$\begin{aligned} F_{(x,y)}(x,y) &= F_x(x)F_y(y) = \left( \int_{-\infty}^x f_x(x') dx' \right) \left( \int_{-\infty}^y f_y(y') dy' \right) = \int_{-\infty}^x \int_{-\infty}^y f_x(x') f_y(y') dx' dy' \\ &= \int_{(-\infty, x] \times (-\infty, y]} f_x(x') f_y(y') dx' dy' \end{aligned}$$

quindi  $(X, Y)$  ha densità e  $f_{(x,y)}(x,y) = f_x(x)f_y(y)$

Viceversa se  $f_{(x,y)} = f_x f_y$

$$\begin{aligned} F_{(x,y)}(x,y) &= \int_{-\infty}^x \int_{-\infty}^y f_x(x') f_y(y') dx' dy' = \left( \int_{-\infty}^x f_x(x') dx' \right) \left( \int_{-\infty}^y f_y(y') dy' \right) \\ &= F_x(x)F_y(y) \end{aligned}$$

□

Osservazione la proposizione si estende al caso di un numero finito di v.a.

densità marginali  $X: \Omega \rightarrow \mathbb{R}^d$  v.a.  $X = (X_1, \dots, X_d)$  con densità  $f_x$

allora per ogni  $J \subset \{1, \dots, d\}$  la v.a.  $(X_i)_{i \in J}: \Omega \rightarrow \mathbb{R}^{\#J}$

ha densità. In particolare per ogni  $i = 1, \dots, d$  la v.a.  $X_i$  ha densità:

$$f_{x_i}(x) = \int_{\mathbb{R}^{d-1}} f_x(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_d) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_d$$

dim. Dimostriamo il solo caso  $J = \{1\}$

$$\begin{aligned} F_{X_1}(x) &= P[X_1 \leq x] = P[X \in (-\infty, x] \times \mathbb{R}^{d-1}] \\ &= \int_{(-\infty, x] \times \mathbb{R}^{d-1}} f_X(y) dy = \int_{-\infty}^x \left( \int_{\mathbb{R}^{d-1}} f_X(y_1, \dots, y_d) dy_2 \dots dy_d \right) dy_1 \quad \square \end{aligned}$$

densità condizionale  $(X, Y)$  ha densità  $f_{(X,Y)}$ . La densità condizionale di  $Y$  dato  $X$

$$f_{Y|X}(y|x) = \begin{cases} \frac{f_{(X,Y)}(x,y)}{f_X(x)} & \text{se } f_X(x) \neq 0 \\ 0 & \text{altrimenti} \end{cases}$$

Osseviamo che  $\underbrace{f_X(x)}_{\text{se } = 0} = \underbrace{\int_{\mathbb{R}} f_{(X,Y)}(x,y) dy}_{\text{allora } = 0}$

### VALORE ATTESO

Se  $X$  v.a. reale con densità  $f_X$ , allora

$$E[X] = \int_{\mathbb{R}} x f_X(x) dx \quad \text{se} \quad \int_{\mathbb{R}} |x| f_X(x) dx < \infty$$

nel caso di v.a. discrete

$$\sum x p_X(x) \quad \text{se} \quad \sum |x| p_X(x) < \infty$$

In generale, se  $X$  è v.a. su  $\mathbb{R}^d$  con densità  $f_X$  e se  $\varphi: \mathbb{R}^d \rightarrow \mathbb{R}$  misurabile allora  $\varphi(X)$  è una v.a. e

$$E[\varphi(X)] = \int_{\mathbb{R}^d} \varphi(x) f_X(x) dx \quad \text{se} \quad \int_{\mathbb{R}^d} |\varphi(x)| f_X(x) dx < \infty$$

### Formula del cambio di variabile

$X$  v.a. su  $\mathbb{R}^d$  con densità  $f_X$ , supponiamo  $f_X = 0$  fuori da  $A$

Sia  $\varphi: A \rightarrow B$  diffeomorfismo, allora la v.a.  $Y = \varphi(X)$  ha densità  $f_Y$

$$f_Y(y) = f_X(\varphi^{-1}(y)) |\det D\varphi^{-1}(y)| \quad y \in B$$

dimi Dalla formula del cambio di variabili, se  $V \subset B$  Boreliano

$$P[Y \in V] = P[X \in \varphi^{-1}(V)] = \int_{\varphi^{-1}(V)} f_X(x) dx \quad y = \varphi(x)$$

$$= \int_V f_X |\varphi^{-1}(y)| |\det D \varphi^{-1}(y)| dy$$

Osservazioni 1. se  $d=1$   $\xrightarrow{\text{se siamo in } \mathbb{R}}$   $f_Y(y) = f_X(\varphi^{-1}(y)) \left| \frac{d\varphi^{-1}}{dy}(y) \right|$   $\xrightarrow{\frac{1}{\varphi'(\varphi^{-1}(y))}}$

2. Se  $X$  è una v.a. con densità  $f_X$  e se  $f_X = 0$  su un aperto  $B$ , allora  $P[X \in B] = 0$

### formula di convoluzione

Se  $X, Y$  sono v.a. reali indipendenti con densità  $f_X, f_Y$ , allora  $X+Y$  ha densità  $f_{X+Y}(z) = f_X \star f_Y(z) = \int_{\mathbb{R}} f_X(x) f_Y(z-x) dx$

dimi  $f_{(X,Y)}(x,y) = f_X(x) f_Y(y)$

$$\begin{aligned} F_{X+Y}(z) &= P[X+Y \leq z] = P[(X,Y) \in \overbrace{\{(x,y) \in \mathbb{R}^2 : x+y \leq z\}}^{A_z}] \\ &= \int_{A_z} f_X(x) f_Y(y) dx dy = \int_{\mathbb{R}} dx \int_{-\infty}^{z-x} f_X(x) f_Y(y) dy \\ &= \int_{\mathbb{R}} f_X(x) \left( \int_{-\infty}^{z-x} f_Y(y) dy \right) dx = \int_{\mathbb{R}} F_Y(z-x) f_X(x) dx \end{aligned}$$

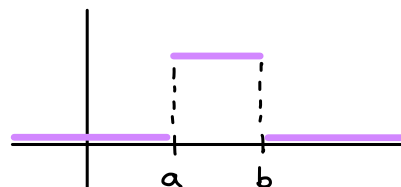
D'altra parte se  $f_X \star f_Y$  è la densità, dobbiamo ottenere  $F_Y(z-x)$  da

$$\begin{aligned} \int_{-\infty}^z f_X \star f_Y(u) du &= \int_{-\infty}^z \int_{\mathbb{R}} f_X(x) f_Y(u-x) du dx = \int_{\mathbb{R}} f_X(x) \left( \int_{-\infty}^z f_Y(u-x) du \right) dx \\ v &= u-x \\ &= \int_{\mathbb{R}} f_X(x) \left( \int_{-\infty}^{z-x} f_Y(v) dv \right) dx = \int_{\mathbb{R}} f_X(x) F_Y(z-x) dx \quad \square \end{aligned}$$

### Esempi

1. Distribuzione uniforme  $X$  ha distribuzione uniforme su  $[a,b]$  se

$$f_X(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x)$$



Per valutare il valore atteso

$$\int_{\mathbb{R}} |x| f_X(x) dx = \int_a^b \frac{1}{b-a} |x| dx < \infty$$

quindi

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_x(x) dx = \frac{1}{b-a} \int_a^b x dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}$$

$$\mathbb{E}[X^2] = \mathbb{E}[\varphi(X)] = \int_{\mathbb{R}} x^2 f_x(x) dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{a^2 + ab + b^2}{3}$$

$\varphi: \mathbb{R} \rightarrow \mathbb{R}$   $\varphi(x) = x^2$   $\mathbb{E}[|\varphi(x)|] < \infty \leftarrow$  vero perché  $X$  ha momenti di ogni ordine

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{3}(a^2 + ab + b^2) - \frac{1}{4}(a^2 + b^2) = \dots = \frac{1}{12}(b-a)^2$$

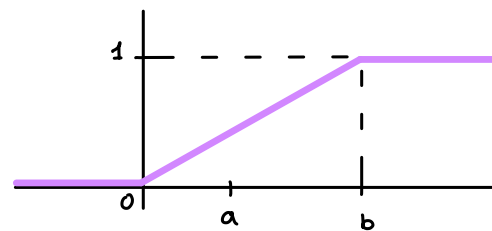
Funzione generatrice dei momenti

$$\phi_x(t) = \mathbb{E}[e^{tx}] = \int_{\mathbb{R}} e^{tx} f_x(x) dx = \frac{1}{b-a} \int_a^b e^{tx} dx = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

$$\varphi(x) = e^{tx}$$

La funzione di ripartizione. Se  $x \in \mathbb{R}$

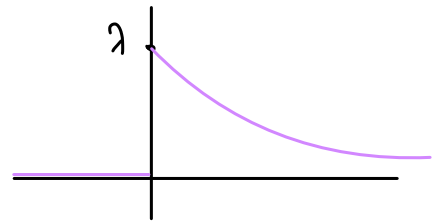
$$F_x(x) = \begin{cases} 0 & \text{se } x \leq a \\ \text{se } x \in (a, b): \int_{-\infty}^x f_x(y) dy = \int_a^x \frac{1}{b-a} dy = \frac{x-a}{b-a} \\ 1 & \text{se } x \geq b \end{cases}$$



2. esponenziale  $X$  ha distribuzione esponenziale di parametro  $\lambda > 0$

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

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



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

Una v.a. esponenziale ha la proprietà di mancanza di memoria:

$$\text{per ogni } x, y > 0 \quad \mathbb{P}[X > x+y \mid X > x] = \frac{\mathbb{P}[X > x+y]}{\mathbb{P}[X > x]} = e^{-\lambda y} = \mathbb{P}[X > y]$$

$$\mathbb{P}[X > x+y] = \mathbb{P}[X > x] \mathbb{P}[X > y] \longrightarrow \varphi(x) \varphi(y) = \varphi(x+y)$$

$$\varphi(x) = \mathbb{P}[X > x] = 1 - F_x(x)$$

Funzione generatrice dei momenti. Se  $t < \lambda$

$$\phi_x(t) = \mathbb{E}[e^{tx}] = \int_{\mathbb{R}} e^{tx} f_x(x) dx = \int_0^{+\infty} \lambda e^{-\lambda x} e^{tx} dx = \lambda \int_0^{+\infty} e^{-(\lambda-t)x} dx = \frac{\lambda}{\lambda-t}$$

3. Gaussiana di parametri  $m, \sigma^2$

$$f_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-m)^2}$$

$$\mathbb{E}[X] = m \quad \text{Var}(X) = \sigma^2$$

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy$$

funzione di  
ripartizione  
di una Gaussiana  
standard

se  $X \sim N(m, \sigma^2)$  con  $\frac{X-m}{\sigma}$  è Gaussiana standard

$$F_x(x) = \mathbb{P}[X \leq x] = \mathbb{P}\left[\frac{X-m}{\sigma} \leq \frac{x-m}{\sigma}\right] = \Phi\left(\frac{x-m}{\sigma}\right)$$

4. Cauchy  $f_x(x) = \frac{1}{\pi(1+x^2)}$

non ha valore atteso finito  $\int_{\mathbb{R}} |x| f_x(x) dx = \int_{\mathbb{R}} \frac{|x|}{\pi(1+x^2)} dx = \infty$

5. Gamma di parametri  $r \geq 1$  intero e  $\lambda > 0$

$$f_x(x) = \frac{1}{\Gamma(r)} \lambda^r x^{r-1} e^{-\lambda x} \mathbb{1}_{[0, \infty)}$$

• se  $r=1$  esponenziale

$$\mathbb{E}[X] = \frac{r}{\lambda} \quad \text{Var}(X) = \frac{r}{\lambda^2}$$

• se  $r$  è intero una Gamma( $r, \lambda$ ) ha la stessa distribuzione della somma di  $r$  v.a. esponenziali indipendenti di parametro  $\lambda$

•  $\chi^2$  (chi-quadro):  $n \geq 1$  intero (gradi di libertà): caso particolare Gamma( $\frac{n}{2}, \frac{1}{2}$ )

$$\mathbb{E}[X] = n \quad \text{Var}(X) = 2n$$

v.a. con legge  $\chi^2(n)$  ha la stessa distribuzione della somma dei quadrati di  $n$  Gaussiane standard indipendenti

• Student  $n$  gradi di libertà

$$f_x(x) = c_n \left(1 + \frac{x^2}{n}\right)^{-\frac{1}{2}(n+1)}$$

$$E[X] = 0$$

$$\text{Var}(X) = \begin{cases} \frac{n}{n-2} & \text{se } n > 2 \\ +\infty & \text{se } n = 1, 2 \end{cases}$$

$X$  con distribuzione di Student, ha la legge di  $\sqrt{n} \frac{z}{s}$

dove  $z$  è Gaussiana standard,  $s$  è  $\chi^2(n)$ , e  $z, s$  indipendenti

lezione 15  
27-04-2021  
(Romito)

• formula di convoluzione

Se  $X, Y$  sono v.a. reali indipendenti con densità  $f_x, f_y$ , allora  $X+Y$  ha densità  $f_{x+y}(x) = f_x \star f_y(z) = \int_{\mathbb{R}} f_x(x) f_y(z-x) dx$

Gaussiana  $N(m, \sigma^2)$   $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-m)^2}$

Somma di Gaussiane

Se  $X \in N(m_1, \sigma_1^2)$ ,  $Y \in N(m_2, \sigma_2^2)$  e se  $X, Y$  indipendenti allora  $X+Y \in N(m_1+m_2, \sigma_1^2+\sigma_2^2)$

dim 1. con la formula di convoluzione

$$X - m_1 \in N(0, \sigma_1^2)$$

$$Y - m_2 \in N(0, \sigma_2^2)$$

Non è restrittivo supporre  $m_1 = m_2 = 0$

$$f_{x+y}(z) = \int_{\mathbb{R}} f_x(x) f_y(z-x) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{\mathbb{R}} e^{-\frac{1}{2\sigma_1^2}x^2} e^{-\frac{1}{2\sigma_2^2}(z-x)^2} dx$$

$$\frac{1}{2\sigma_1^2}x^2 + \frac{1}{2\sigma_2^2}(z-x)^2 = \frac{z^2}{2(\sigma_1^2+\sigma_2^2)} + \frac{1}{2}\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)\left(x - \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}z\right)^2$$

$$f_{x+y}(z) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}} \underbrace{\int_{\mathbb{R}} e^{-\frac{1}{2}\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)\left(x - \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}z\right)^2} dx}_{\text{con il cambio di variabile } x' = x - \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}z}$$

$$= \int_{\mathbb{R}} e^{-\frac{1}{2}\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)(x')^2} dx' = \sqrt{2\pi \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}} \rightarrow \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$

quindi

$$f_{X+Y}(z) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} e^{-\frac{z^2}{2(\sigma_1^2 + \sigma_2^2)}}$$

2. con la funzione generatrice dei momenti  $\phi_x(t) = E[e^{tx}]$

Supponiamo  $U$  sia  $N(m, \sigma^2)$ , allora  $\frac{U-m}{\sigma}$  è  $N(0, 1)$

$$P\left[\frac{U-m}{\sigma} \leq x\right] = P[U \leq m + \sigma x] = \int_{-\infty}^{m+\sigma x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-m)^2} dy$$

$y' = \frac{y-m}{\sigma}$

$$\phi_U(t) = E[e^{tU}] = E\left[e^{t(m + \sigma \frac{U-m}{\sigma})}\right] = E[e^{t(m + \sigma z)}]$$

dove  $z = \frac{U-m}{\sigma}$   $= e^{tm} \phi_z(\sigma t)$

$$\phi_z(t) = E[e^{tz}] = \int_{\mathbb{R}} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2 - 2tz)} dz$$

$$z^2 - 2tz = (z-t)^2 - t^2$$

$$= \frac{1}{\sqrt{2\pi}} e^{+\frac{1}{2}t^2} \underbrace{\int_{\mathbb{R}} e^{-\frac{1}{2}(z-t)^2} dz}_{= 2\pi} = e^{+\frac{1}{2}t^2}$$

Quindi  $\phi_U(t) = e^{tm + \frac{1}{2}\sigma^2 t^2}$

Se ora  $X \sim N(m_1, \sigma_1^2)$  e  $Y \sim N(m_2, \sigma_2^2)$

$$\phi_{X+Y}(t) = \phi_X(t) \phi_Y(t) = e^{t(m_1+m_2) + \frac{1}{2}(\sigma_1^2 + \sigma_2^2)t^2}$$

□

Osservazione Riepurgiamo al teorema del limite centrale

Supponiamo  $X_1, \dots, X_n$  sono  $N(m, \sigma^2)$

$$S_n = X_1 + \dots + X_n \sim N(m \cdot n, n\sigma^2)$$

$$\frac{S_n - E[S_n]}{\text{sd}(S_n)} \sim N(0, 1)$$

Il "limite" si realizza per  $n$  finito.

# STATISTICA

- Statistica descrittiva
- Statistica inferenziale

la statistica è una disciplina di analisi dei dati

## Statistica descrittiva

Sintesi dei dati

$x_1, \dots, x_n$

$y_1, \dots, y_n$

↓  
scopo della stat.  
descrittiva.

media campionaria  
= media aritmetica

$$\bar{x} = \frac{1}{n} (x_1 + \dots + x_n)$$

Può essere fatto in  
maniera grafica,  
numerica ecc...

varianza campionaria

$$\bar{\sigma}^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2$$

$x_1, \dots, x_n$

$y_1, \dots, y_n$

$$\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y})$$

↳ covarianza  
campionaria

Statistica inferenziale → quella che studieremo

→ Analisi dei dati effettuata attraverso un modello

Consideriamo il lancio ripetuto di una moneta per  $n$  volte

$x_1, \dots, x_n$

$x_i \in \{0, 1\}$

Conoscenza del valore  $p$  della probabilità di ottenere 1

• problema 1: qual è il valore di  $p$

"soluzione" del problema

Interpretiamo i lanci all'interno di un modello probabilistico

$X_1, \dots, X_n$  v.a. Bernoulli indipendenti di parametro  $p$

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{\text{in probab.}} p$$

Allora un valore candidato per  $p$  sarà

$$\bar{p} = \frac{1}{n} (x_1 + \dots + x_n)$$

Non possiamo aspettarci di aver trovato il valore esatto

• la legge dei grandi numeri è un teorema limite

quindi  $\bar{p} \neq p$

- più fondamentale: a  $n$  finito sopravvivono le componenti casuali del problema e il valore stimato ha fluttuazioni casuali che possono sensibilmente cambiare il valore della stima

problema 2 trovare un range di valori "ragionevoli" di  $p$  attraverso i dati.

Ricordiamo il teorema di De Moivre - Laplace ( $b=a=q$ )

$$np - q\sqrt{np(1-p)} \leq s_n \leq np + q\sqrt{np(1-p)}$$

con probabilità

$$\int_{-q}^q \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 2\Phi(q) - 1$$

dove

$$\Phi(q) = \int_{-\infty}^q \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

è la funzione di ripartizione di una Gaussiana standard

$$p - q \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \bar{p} \leq p + q \frac{\sqrt{p(1-p)}}{\sqrt{n}}$$

$$\bar{p} - \frac{q}{\sqrt{n}} \sqrt{p(1-p)} \leq p \leq \bar{p} + \frac{q}{\sqrt{n}} \sqrt{p(1-p)}$$

In conclusione

$$p \in \left( \bar{p} - \frac{q}{\sqrt{n}} \sqrt{p(1-p)}, \bar{p} + \frac{q}{\sqrt{n}} \sqrt{p(1-p)} \right)$$

con probabilità approssimativamente  $2\Phi(q) - 1$

Osserviamo che  $p(1-p) \leq \frac{1}{4}$  quindi

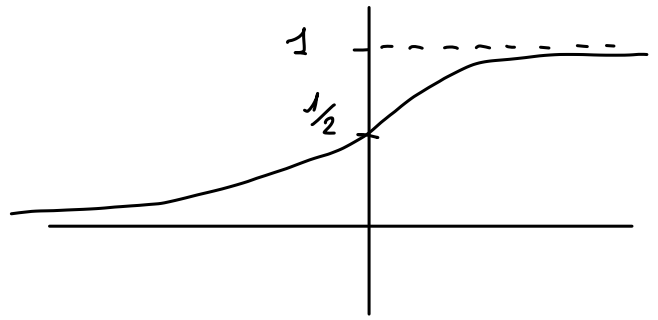
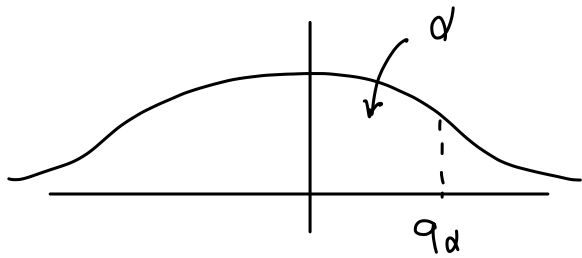
$$p \in \left( \bar{p} - \frac{q}{2\sqrt{n}}, \bar{p} + \frac{q}{2\sqrt{n}} \right) \text{ con probab. } \approx 2\Phi(q) - 1$$

quantile di ordine  $d \in [0, 1]$  è il valore  $q_d$  tale che

$$\Phi(q_d) = d$$

(funzione inversa di  $\Phi$ )

$$2\Phi(q) - 1 = d \quad \Phi(q) = \frac{d+1}{2}$$



$2\Phi(q) - 1$	50%	75%	90%	95%	99%
$q$	0.77	1.15	1.65	1.96	2.58

problema 3 possiamo "mettere alla prova" i valori "candidati" di  $p$  che abbiamo determinato?

Dato un valore  $p_0$ , scoprire se  $p_0$  è compatibile con i dati

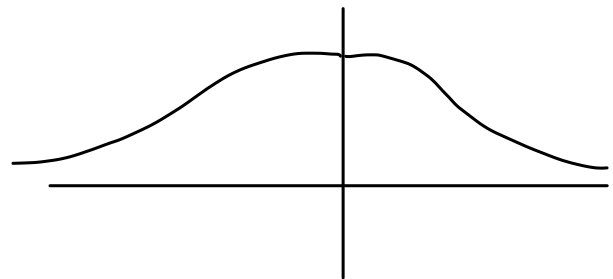
$$\frac{\sqrt{n}(\bar{p} - p_0)}{\sqrt{p(1-p)}} \rightarrow \frac{1}{2}$$

Partiamo dall'ipotesi che effettivamente  $p_0$  sia un buon valore di  $p$ . Possiamo allora stimare

$$\sqrt{p(1-p)} \approx \sqrt{p_0(1-p_0)}$$

e quindi studiamo

$$\frac{\sqrt{n}(\bar{p} - p_0)}{\sqrt{p_0(1-p_0)}} = q$$



Se  $q$  è grande

- $p_0$  non è il valore adeguato
- è accaduto un caso fortuito e raro

Se  $q$  è "piccolo"

- $p_0$  è plausibile

## Conclusioni

- stima dei parametri statistici
- regioni di fiducia
- test statistici

Esercizio 8  
28-04-2021  
(Di Gesù)

### Problema 3.

Siano  $X$  e  $Y$  variabili aleatorie indipendenti, dove  $X$  abbia distribuzione geometrica di parametro  $1/2$  e  $Y$  abbia distribuzione di Poisson di parametro  $1/2$ . Sia  $(U_n)_{n \geq 1}$  una successione di variabili aleatorie indipendenti e tutte con la stessa legge di  $U = X - 2Y$ . Stimare  $n$  affinché le variabili aleatorie  $S_n = U_1 + \dots + U_n$  superi il livello  $2n$  con probabilità al più  $0.1$ .

$$\forall a > 0 \quad \mathbb{P}[|S_n - \mathbb{E}[S_n]| \geq a] \leq \frac{1}{a^2} \text{Var}[S_n]$$

$$\mathbb{E}[S_n] = \mathbb{E}\left[\sum_{i=1}^n U_i\right] = \sum_{i=1}^n \mathbb{E}[U_i] = n$$

$$\mathbb{E}[U_i] = \mathbb{E}[X - 2Y] = \mathbb{E}[X] - 2\mathbb{E}[Y] = 2 - 2 \cdot \frac{1}{2} = 1$$

$$\mathbb{P}[X = k] = \left(\frac{1}{2}\right)^k$$

$$\mathbb{P}[Y = i] = e^{-\frac{1}{2}} \frac{1}{2^i i!}$$

$$\begin{aligned} \text{Var}(S_n) &= \text{Var}\left(\sum_{i=1}^n U_i\right) = \sum_{i=1}^n \text{Var}(U_i) = \sum_{i=1}^n (\mathbb{E}[U_i^2] - \mathbb{E}[U_i]^2) = \\ &= \sum_{i=1}^n (\mathbb{E}[U_i^2] - 1) \end{aligned}$$

$$\begin{aligned} \text{Var}(U) &= \text{Var}(X - 2Y) = \text{Var}(X) + \text{Var}(-2Y) = \text{Var}(X) + 4\text{Var}(Y) \\ &= 4 \end{aligned}$$

$$\text{Var}(X) = \frac{(1 - \frac{1}{2})}{\frac{1}{4}} = 2$$

$$\text{Var}(Y) = \frac{1}{2}$$

$$\text{Var}(S_n) = \sum_{i=1}^n 4 = 4n$$

$$\mathbb{P}[|S_n - n| \geq a] \leq \frac{1}{a^2} \cdot 4n$$

$$a=n \quad \mathbb{P}[|S_n - n| \geq n] \leq \frac{1}{n^2} \cdot 4n$$

$$|S_n - n| \geq n \Leftrightarrow S_n - n \geq n \vee S_n - n \leq -n$$

$$\mathbb{P}[S_n - n \geq n] \leq \mathbb{P}[|S_n - n| \geq n] \leq \frac{1}{n^2} \cdot 4n$$

$$\mathbb{P}[S_n \geq 2n] \leq \frac{4}{n}$$

$$\frac{4}{n} = 0.1 \quad n \geq 40$$

soluzione gdg:

$$\mathbb{P}(S_n > 2n) = \mathbb{P}(e^{\lambda S_n} > e^{2\lambda n}) \underset{\text{mark}}{\leq} e^{-2\lambda n} \mathbb{E} e^{\lambda S_n}$$

$$= e^{-2\lambda n} \mathbb{E}[e^{\lambda U_1}] \dots \mathbb{E}[e^{\lambda U_n}]$$

$$= e^{-2\lambda n} [\mathbb{E}[e^{\lambda X}]]^n [\mathbb{E}[e^{-2\lambda Y}]]^n$$

$$= e^{-n [2\lambda - \log \mathbb{E}[e^{\lambda X}] - \log \mathbb{E}[e^{-2\lambda Y}]]}$$

$\phi(\lambda)$

basta trovare  $\lambda$  t.c.  $\phi(\lambda) > 0$

( $n \geq 23$ )

### Problema 6.

Sia  $X$  una variabile aleatoria con media nulla e momento quarto finito. Dimostrare che esiste una costante  $C > 0$  tale che per ogni  $\varepsilon > 0$  ed ogni intero positivo  $n$

$\mathbb{P}[|X_n| \geq \varepsilon] \leq C/(\varepsilon^4 n^2)$ , con  $X_n$  la media empirica di  $n$  copie indipendenti di  $X$ . Hp 3

Th

"Più momenti si hanno e più la media empirica converge alla media in probabilità"  $\rightarrow$  idea

$$\mathbb{P}[|\bar{X}_n| \geq \varepsilon] \leq \frac{C}{\varepsilon^4 n^2}$$

$X_1, \dots, X_n$  indipendenti

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n} \quad \text{media empirica}$$

per la disuguaglianza di Markov:

$$\mathbb{P}[|\bar{X}_n|^4 \geq \varepsilon^4] \leq \frac{1}{\varepsilon^4} \mathbb{E}[\bar{X}_n^4] \rightarrow \frac{1}{n^4} \mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)^4\right]$$

$\hookrightarrow |\bar{X}_n| \geq \varepsilon$  poiché è tutto positivo

essendoci il fattore  $\varepsilon^4$ , ottenendo una stima ragionevole riusciamo a dimostrare la tesi

per omogeneità del valore atteso

$$\frac{1}{n^4} \mathbb{E} \left[ \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n X_i X_j X_k X_l \right] \rightarrow \text{perché è la } 4^{\text{a}} \text{ potenza di un polinomio}$$

$$= \frac{1}{n^4} \left( \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \mathbb{E} [X_i X_j X_k X_l] \right)$$

linearità  
del valore atteso

voglio ottenere  $n^2$   
cancellazioni

sono  $n^4$  termini  
ma essendo indipendenti  
ci sono delle cancellazioni

Gli indici possono essere tutti uguali, uguali a due a due, oppure diversi. Prendiamo ad esempio il caso  $j=k=l$  e  $i$  diverso: allora avrò, per indipendenza:  $\underbrace{\mathbb{E}[X_i]}_{=0} \mathbb{E}[X_j^3] = 0$ .

Lo stesso ragionamento vale per gli altri casi in cui gli indici sono diversi, tranne nel caso in cui lo sono due a due.

Allora gli unici casi ammessi sono:

- gli indici sono tutti uguali  $\rightarrow n$  modi
- gli indici sono uguali due a due  $\rightarrow n(n-1) \cdot \frac{4!}{2!2!} = 6n(n-1)$

per lo scopo  
dell'es. basta  
che è un num  
finito

scelgo  
il 1° num  
"A"  
scelgo  
il 2° num  
"B"  
"autogrammi"  
di A A B B

$$\mathbb{E}[\bar{X}_n^4] = \frac{1}{n^4} \left( \sum_{i=1}^n \underbrace{\mathbb{E}[X_i^4]}_{\substack{\text{momento quarto} \\ \text{che chiamo } \delta}} + \underbrace{6n(n-1)}_{\text{combinaz.}} \underbrace{\mathbb{E}[X_i^2]}_{\substack{\text{momento secondo} \\ \text{che chiamo } \beta}} \underbrace{\mathbb{E}[X_j^2]}_{\substack{\text{momento secondo} \\ \text{che chiamo } \beta}} \right) =$$

$$= \frac{1}{n^4} \left( n\delta + 6n^2\beta^2 - 6n\beta^2 \right) = \frac{1}{n^4} \left( 6n^2\beta^2 + (\delta - 6\beta^2)n \right) =$$

$$= \frac{1}{n^2} \left( 6\beta^2 + \frac{\delta - 6\beta^2}{n} \right)$$

volendo si potrebbe  
"aggiustare" in questo  
modo:  $6\beta^2 + n\left(\frac{\delta - 6\beta^2}{n}\right) =$   
 $= \delta + 1 > 0$

Torniamo alla  
stima iniziale  
 $\rightarrow$

$$\mathbb{P} [ |\bar{X}_n| \geq \varepsilon ] \leq \frac{1}{\varepsilon^4 n^2} \left( 6\beta^2 + \frac{\delta - 6\beta^2}{n} \right)$$

poiché questa è  
una costante la  
tesi è dimostrata

□

**Problema 5.**

Sia  $Z = \max\{X^2, Y^2\}$  con  $X, Y$  variabili aleatorie standardizzate aventi covarianza  $\rho$ .

(i) Si mostri che  $E[Z] = 1$  nel caso particolare in cui  $|\rho| = 1$ .

(ii) Si mostri che in generale valgono le stime  $1 \leq E[Z] \leq 1 + \sqrt{1 - \rho^2}$ .

$$(i) \quad Y = \frac{sd(Y)}{sd(X)} X + E[X] \quad \pm \quad \frac{sd(Y)}{sd(X)} E[X]$$

poiché  $|\rho| = 1$

per via del modulo

$$Y = \pm X \rightarrow \text{poiché } |\text{cov}(X, Y)| = 1$$

c'è una relazione lineare:  $Y = aX + b$

$$Z = \max\{X^2, Y^2\}$$

$$Z = \max\{X^2, X^2\} = X^2$$

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

ragionamento gdg:  
passando al valore atteso  $b=0$  (per forza) e passando al momento secondo si ha  $a^2 = 1 \Rightarrow a = \pm 1$

$$ii) \quad Z \geq X^2 \text{ ovvio} \rightarrow E[Z] \geq 1$$

$$\max\{s, t\} = \frac{s+t + |t-s|}{2}$$

vero in generale

$$Z = \frac{1}{2} (X^2 + Y^2 + |X^2 - Y^2|)$$

$$\Rightarrow E[Z] = \frac{1}{2} (E[X^2] + E[Y^2] + E[|X^2 - Y^2|])$$

linearità

$$\text{Var } X = 1$$

$$\text{Var } Y = 1$$

perché  $X$  e  $Y$  sono standardizzate

$$E[Z] = 1 + \frac{1}{2} E[|X - Y| |X + Y|]$$

$$\frac{1}{2}(1+1)$$

diff. di quadrati

Oss:  $\text{cov}$  è un prodotto scalare

$$E[AB] = \text{cov}(A, B) \text{ perché sono standardizzate}$$

Disuguaglianza di Cauchy-Schwartz

$$E[AB] \leq \sqrt{E[A^2] E[B^2]} \quad (\text{vero in generale})$$

$$A = |X - Y|$$

$$B = |X + Y|$$

$$E[|X - Y| |X + Y|] \leq \sqrt{E[(X - Y)^2] E[(X + Y)^2]} =$$

$$E[X - Y] = \text{Var}(X) + \text{Var}(Y) - 2\rho = 2 - 2\rho$$

$$E[X + Y] = 2 + 2\rho$$

$$E[|X - Y| |X + Y|] \leq \sqrt{4 - 4\rho^2} = 2\sqrt{1 - \rho^2} \Rightarrow E[Z] \leq 1 + \sqrt{1 - \rho^2}$$

### Problema 8.

Un sacchetto contiene  $m$  gettoni, contrassegnati dai numeri da 1 a  $m$ , con  $m \geq 3$ . I gettoni sono estratti con la seguente modalità: ogni gettone estratto viene reinserito nel sacchetto dopo la successiva estrazione. Per  $n \geq 1$ , sia  $S_n$  il numero di volte in cui il gettone  $m$  è estratto nelle prime  $n$  estrazioni.

- Calcolare la media di  $S_n$ .
- Calcolare la varianza di  $S_n$ .

• Sia 
$$X_k = \begin{cases} 1 & \text{esce } \textcircled{m} \text{ alla } k\text{-esima estrazione} \\ 0 & \text{altrimenti} \end{cases}$$

$$\mathbb{E}[X_k] = \mathbb{P}[X_k = 1] =: p_k$$

$$p_1 = \frac{1}{m}$$

$$p_2 = \underbrace{\frac{m-1}{m}}_{1^{\text{a estr.}}} \underbrace{\frac{1}{m-1}}_{2^{\text{a estr.}}} = \frac{1}{m}$$

$$p_3 = \underbrace{\frac{1}{m}}_{1^{\text{a e.}}} \underbrace{\frac{1}{m-1}}_{3^{\text{a e.}}} + \underbrace{\frac{m-1}{m}}_{1^{\text{a e.}}} \underbrace{\frac{m-2}{m-1}}_{2^{\text{a e.}}} \underbrace{\frac{1}{m-1}}_{3^{\text{a e.}}} = \frac{1}{m(m-1)^2} = \frac{1}{m}$$

Per induzione  $p_k = \frac{1}{m} \quad \forall k$

$$\mathbb{E}[S_n] = \mathbb{E}\left[\sum_n X_k\right] = \frac{n}{m}$$

• Gli  $X_k$  NON sono indipendenti

$$\text{Var}(S_n) = \sum_{k,j=1}^n \underbrace{\mathbb{E}[X_k X_j]}_{\text{spezziamolo nel caso in cui } k=j} - \left(\frac{n}{m}\right)^2 =$$

$$= 2 \sum_{k \geq j} \mathbb{E}[X_k X_j] + \frac{n}{m} - \left(\frac{n}{m}\right)^2$$

$$= 2 \sum_{k=1}^{n-1} \sum_{j=0}^{n-k-1} \mathbb{E}[X_k X_{k+j+1}] + \frac{n}{m} - \left(\frac{n}{m}\right)^2$$

⊛ Affermazione  $\forall k=1, \dots, n-1 \quad \forall j=0, \dots, n-k-1$

$$\mathbb{E}[X_k X_{k+j+1}] = \mathbb{P}(X_{k+j+1} = 1, X_k = 1) = \frac{1}{m^2} \left[ 1 - \left(\frac{1}{1-m}\right)^j \right]$$

Dimostrazione per induzione

Caso in cui  $j \neq 0$

$$\text{Var}(S_n) = \frac{n(m-1)(m-2)+2}{m^3} + \frac{2(m-1)}{m^4} \left[ (1-m)^{1-n} - 1 \right]$$

### Dimostrazione (\*)

Fissiamo  $k = 1, \dots, n-1$

•  $j = 0 \checkmark$

• sia  $j = 1, \dots, n-k-1$

Assumiamo che sia vera l'identità per  $j-1$

$$\text{ovvero } \mathbb{P}(X_{k+j} = 1, X_k = 1) = \frac{1}{m^2} \left[ 1 - \left(-\frac{1}{m-1}\right)^{j-1} \right]$$

$$\text{Allora } \mathbb{P}(X_{k+j+1} = 1, X_k = 1) = \mathbb{P}(X_{k+j+1} = 1, X_{k+j} = 0, X_k = 1)$$

$$= \mathbb{P}(X_{k+j+1} = 1 \mid X_{k+j} = 0, X_k = 1) \mathbb{P}(X_k = 1) =$$

$$= \mathbb{P}(X_{k+j+1} = 1 \mid X_{k+j} = 0) \mathbb{P}(X_{k+j} = 0 \mid X_k = 1) \mathbb{P}(X_k = 1)$$

$\forall \varepsilon > 0$

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \frac{1}{m}\right| > \varepsilon\right) \xrightarrow{n \rightarrow \infty} 0 \quad (\text{Chebyshev})$$

lezione 16

03-05-21

(Rovito)

Dei 3 problemi visti la scorsa volta solo il primo trova una reale soluzione soddisfacente per il suo apporto matematico

## TEORIA DEGLI STIMATORI

idea: vogliamo lavorare su un modello probabilistico ma questo dipende dai parametri, che in qualche maniera ci dicono come sono distribuite le v.a. che entrano in gioco nel modello che descrive il fenomeno, dunque lo sp. di prob. dipende in qualche modo dai parametri.

### Modello statistico:

(H) insieme di valori dei parametri

↓ theta maiuscolo

Chiameremo modello statistico la terna  $(\Omega, \mathcal{F}, \mathbb{P}_\theta)_{\theta \in H}$

Supporremo che il modello sia non banale ovvero

se  $\theta_1 \neq \theta_2$ , allora  $P_{\theta_1} \neq P_{\theta_2}$

Campione Sia  $(m_\theta)_{\theta \in \Theta}$  una famiglia di probabilità (su  $\mathbb{R}$  o  $\mathbb{R}^d$  o insieme di simboli, etc. a seconda del fenomeno)

Un campione di  $n$  individui estratto da una popolazione di legge  $m_\theta$  è una famiglia  $X_1, \dots, X_n$  di v.a. indipendenti e con legge  $m_\theta$   
 $X_1, \dots, X_n$  indipendenti su ogni  $(\Omega, \mathcal{F}, P_\theta)$

Bernoulli di  
parametro  $\theta$

per ogni  $\theta$ :  $X_1, \dots, X_n$  hanno ognuna legge  $m_\theta$

Osservazione Possiamo generare un modello statistico attraverso un campione

Supponiamo che le v.a. del campione siano reali. Assegnate  $(m_\theta)_{\theta \in \Theta}$  probabilità su  $\mathbb{R}$ ,

$$\Omega = \mathbb{R}^n$$

$$\mathcal{F} = \mathcal{B}(\mathbb{R}^n)$$

- caso discreto  $P_\theta(\{x_1, \dots, x_n\}) = \prod_{i=1}^n m_\theta(\{x_i\})$

- caso continuo  $P_\theta$  con densità  $\prod_{i=1}^n f_\theta(x_i)$ , dove  $f_\theta$  è la densità associata a  $m_\theta$

$$X_i: \Omega \rightarrow \mathbb{R} \text{ come } X_i(\omega) = \omega_i$$

Statistica Una statistica, all'interno di un modello statistico, è una variabile aleatoria che non dipende dal parametro

Stimatore Una statistica che è in funzione del campione

esempio 1.  $X_1, \dots, X_n$  campione, allora  $X_1$  è uno stimatore

2.  $X_1, \dots, X_n$  campione,  $\frac{1}{n}(X_1 + \dots + X_n)$  è uno stimatore

la sola definizione di stimatore non è sufficiente dobbiamo capire quando uno stimatore è "buono" e quando no

Stimatore corretto (o non distorto)  $(\Omega, \mathcal{F}, P_\theta)_{\theta \in \Theta}$  modello statistico,

$X_1, \dots, X_n$  campione,  $U$  è uno stimatore corretto se

valore atteso  
valutato rispetto  
a  $P_\theta$   $\rightarrow \mathbb{E}_\theta[U] = \theta$  per ogni  $\theta \in \Theta$

poiché ci sono un'infinità di spazi di probabilità, ne fisso uno

Talvolta non troveremo stime del parametro ma di funzioni del param.

stimatore corretto:  $E[\hat{\theta}_n] \rightarrow \theta$ , successione consistente  $(\hat{\theta}_n)_{n \geq 1} : \hat{\theta}_n \xrightarrow{\text{in probab.}} \theta$

esempi 1.  $X_1, \dots, X_n$  campione estratto da una popolazione di Bernoulli

di parametro  $p$  allora la media empirica  $\bar{X} = \frac{1}{n} (X_1 + \dots + X_n)$  è

uno stimatore corretto  $E_p[\bar{X}] = p$  (ma anche  $E[X_1] = p$ )

2. Come prima ma la popolazione è geometrica di parametro

$p$ , allora  $E_p[\bar{X}] = \frac{1}{p}$  (ma anche  $E[X_1] = \frac{1}{p}$ )

quindi  $\bar{X}$  è uno stimatore corretto di  $\theta(p)$  con  $\theta(p) = \frac{1}{p}$

successione di stimatori asintoticamente non distorta:

$(\hat{\theta}_n)_{n \geq 1}$  tale che  $E[\hat{\theta}_n] \rightarrow \theta$

→ più corretto: successione di stimatori  
Stimatore consistente (asintoticamente non distorto)

Se un modello statistico  $(\Omega, \mathcal{F}, P_\theta)_{\theta \in \Theta}$  per ogni  $n$  sia

dato un campione di numerosità  $n$  estratto da una popolazione

di legge  $m_\theta$ , dove  $(m_\theta)_{\theta \in \Theta}$  è una famiglia di probabilità.

La successione di stimatori  $(U_n)_{n \geq 1}$  è consistente se per ogni

$\theta \in \Theta$  e per ogni  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P_\theta[|U_n - \theta| \geq \varepsilon] = 0$$

idea: più num. è il  
campione, più esperimenti  
facciamo, più si ha un buon  
risultato (sperabilmente)

(Anche in questo caso vale la definizione più generale per l'approssimazione di  
funzioni parametro)

Rischio quadratico: Se  $U$  è uno stimatore di  $\theta$

$$R_\theta(U) = E_\theta[(U - \theta)^2]$$

il rischio introduce un ordinamento parziale tra gli stimatori:

se  $U, V$  stimatori,  $U$  è preferibile a  $V$  se  $R_\theta(U) \leq R_\theta(V) \quad \forall \theta \in \Theta$

Vogliamo trovare un metodo per determinare stimatori di parametri

Nel seguito considereremo un campione  $X_1, \dots, X_n$  estratto da una popolazione di legge  $m_\theta$  per  $\theta \in \Theta$

- caso discreto:  $m_\theta$  è una legge discreta identificata dalla densità discreto  $p_\theta$

- caso continuo:  $m_\theta$  è una legge ass. continua identificata dalla densità  $f_\theta$

Verosimiglianza:

$$L_\theta(x_1, \dots, x_n) = \begin{cases} p_\theta(x_1) \cdot \dots \cdot p_\theta(x_n) & \text{caso discreto} \\ f_\theta(x_1) \cdot \dots \cdot f_\theta(x_n) & \text{caso continuo} \end{cases}$$

Nel caso discreto:

$$L_\theta(x_1, \dots, x_n) = P_\theta(X_1 = x_1) P(X_2 = x_2) \cdot \dots \cdot P(X_n = x_n)$$

Prob. di ottenere ciò che si è osservato  $\rightarrow P_\theta[X_1 = x_1, X_2 = x_2, \dots, X_n = x_n]$

Che cos'è quindi la verosimiglianza?

Dati i valori  $x_1, x_2, \dots, x_n$  che sono i possibili esiti/realizzazioni del campione la verosimiglianza è la probabilità che il campione abbia esattamente quei valori  $\rightarrow$  Ci chiediamo quanto siano verosimili quei valori.

Esperimento  $\rightarrow$  Risultato  $\rightarrow$  Osservazione dei parametri  $\rightarrow$  Quanto è verosimile?

Nel caso continuo la verosimiglianza ha lo stesso significato, ma dobbiamo rapportarci ad un modello continuo  $\rightarrow$  integrale anziché somma

Stima di massima verosimiglianza:

Dato un campione di numerosità  $n$  di legge  $m_\theta$ , sia  $L_\theta$  la funzione di verosimiglianza. Uno stimatore  $U$  è uno stimatore di massima verosimiglianza se

valore del parametro  $\rightarrow$   $L_U(X_1, \dots, X_n) = \sup_{\theta \in \Theta} L_\theta(X_1, X_2, \dots, X_n)$

Significa per q.o.  $\omega$

$$L_{U(\omega)}(X_1(\omega), X_2(\omega), \dots, X_n(\omega)) = \sup_{\theta \in \Theta} L_\theta(X_1(\omega), \dots, X_n(\omega))$$

esempi: 1. popolazioni Bernoulli di parametro  $p \in \mathbb{H} = [0,1]$   $p$  parametro

$$P_p(X) = p^x (1-p)^{1-x} \quad x \in \{0,1\}$$

$$L_p(x_1, \dots, x_n) = p^{x_1 + \dots + x_n} (1-p)^{n - (x_1 + \dots + x_n)}$$

Se  $X_1, \dots, X_n$  campione, dobbiamo massimizzare

$$p \mapsto L_p(X_1, \dots, X_n) = p^{n\bar{x}} (1-p)^{n-\bar{x}}$$

media empirica

spesso conviene massimizzare il logaritmo della funzione

$$\log L_p(X_1, \dots, X_n) = n\bar{x} \log p + n(1-\bar{x}) \log(1-p)$$

↳ come si massimizza? con il calcolo differenziale

NB Noi stiamo massimizzando il parametro

$$\frac{d}{dp} \log L_p = \frac{n\bar{x}}{p} - \frac{n(1-\bar{x})}{1-p}$$

$$\frac{d}{dp} \log L_p = 0 \quad \text{se} \quad p = \bar{x}$$

$$\left. \frac{d^2}{dp^2} (\log L_p) \right|_{p=\bar{x}} < 0 \quad \text{Abbiamo trovato il massimo}$$

2. Campione da una popolazione geometrica di parametro  $p \rightarrow \bar{x}$

↳ tecnicamente  $\bar{x}$  non è uno stimatore corretto perché  $E = \frac{1}{p}$  però è uno stim. corretto di un'opportuna funzione dello stimatore

3. Campione di una popolazione Poisson di parametro  $\lambda > 0 \rightarrow \bar{x}$

↳ la media empirica è uno stimatore corretto

4. Campione da una popolazione uniforme di parametro  $\theta > 0$

$$\text{densità} \rightarrow f_\theta(x) = \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x) \quad x \in \mathbb{R}$$

$$L_\theta(x_1, \dots, x_n) = \theta^{-n} \mathbb{1}_{[0,\theta]}(x_1) \dots \mathbb{1}_{[0,\theta]}(x_n)$$

Osserviamo che  $\mathbb{1}_{[0,\theta]}(x) = \mathbb{1}_{[x, \infty)}(\theta)$  e che

$$\mathbb{1}_{[0,\theta]}(x_1) \dots \mathbb{1}_{[0,\theta]}(x_n) = \mathbb{1}_{[x_1, \infty)}(\theta) \dots \mathbb{1}_{[x_n, \infty)}(\theta) = \mathbb{1}_{[\max(x_1, \dots, x_n), \infty)}(\theta)$$

quindi

$$L_\theta(x_1, \dots, x_n) = \theta^{-n} \mathbb{1}_{[\max(x_1, \dots, x_n), \infty)}(\theta)$$

Dato il campione  $X_1, \dots, X_n$

$$L_\theta(X_1, \dots, X_n) = \theta^{-n} \mathbb{1}_{[\max(X_1, \dots, X_n), \infty)}(\theta) = \theta^{-n}$$

Quindi

$$\hat{\theta} = \max(X_1, \dots, X_n)$$



In questo caso lo stimatore di massima verosimiglianza non è la media empirica, che rimane però uno stimatore quasi corretto

Osservazione Nel contesto precedente

$$\mathbb{E}_\theta[X_1] = \int_{\mathbb{R}} x f_\theta(x) dx = \frac{1}{\theta} \int_0^\theta x dx = \frac{\theta}{2}$$

Un altro stimatore "naturale" per il valore del parametro  $\theta$  è

$$U = \frac{2}{n} \sum_{k=1}^n X_k$$

Vogliamo confrontare i due stimatori

• Sia  $V = \max(X_1, \dots, X_n)$

• Rischio quadratico per  $U$

$$\mathbb{E}_\theta[X_1] = \frac{\theta}{2} \quad \text{Var}_\theta(X_1) = \frac{\theta^2}{12}$$

quindi  $\mathbb{E}_\theta[U] = \theta$  (corretto) e

$$R_\theta(U) = \mathbb{E}_\theta[(U - \theta)^2] = \text{Var}_\theta(U) = \frac{4}{n^2} \cdot n \frac{\theta^2}{12} = \frac{\theta^2}{3n}$$

Vista come succ. di stimatori  $U_n = \frac{2}{n} \sum_{k=1}^n X_k$ ,  $(U_n)_{n \geq 1}$  è consistente grazie alla disuguaglianza di Chebyshev:

$$\mathbb{P}_\theta[|U_n - \theta| \geq \varepsilon] \leq \frac{1}{\varepsilon^2} \text{Var}_\theta(U_n) = \frac{\theta^2}{3n\varepsilon^2} \rightarrow 0$$

• Rischio quadratico per  $V$

$$\begin{aligned} F_V^\theta(x) &= \mathbb{P}_\theta[V \leq x] = \mathbb{P}_\theta[\max(X_1, \dots, X_n) \leq x] = \mathbb{P}_\theta[X_1 \leq x, \dots, X_n \leq x] \\ &= \prod_{i=1}^n \mathbb{P}_\theta[X_i \leq x] = F_{X_1}^\theta(x)^n \end{aligned}$$

Se  $x \in [0, \theta]$   $F_V^\theta(x) = \left(\frac{x}{\theta}\right)^n$  in quanto  $F_{X_1}^\theta(x) = \frac{x}{\theta}$  per  $x \in [0, \theta]$

la densità è quindi

$$f_\theta(x) = n\theta^{-n} x^{n-1} \mathbb{1}_{[0, \theta]}(x)$$

Calcoliamo i momenti di  $V$ : se  $k \geq 1$

$$\mathbb{E}_\theta[V^k] = \int_{\mathbb{R}} x^k f_\theta(x) dx = \int_0^\theta n\theta^{-n} x^{n+k-1} dx = \frac{n}{n+k} \theta^k$$

quindi

$$\mathbb{E}_{\theta}[V] = \frac{n}{n+1} \theta \quad \mathbb{E}_{\theta}[V^2] = \frac{n}{n+2} \theta^2$$

e

$$R_{\theta}(V) = \mathbb{E}_{\theta}[(V - \theta)^2] = \mathbb{E}_{\theta}[V^2 - 2\theta V + \theta^2] = \frac{n}{n+2} \theta^2 - 2 \frac{n}{n+2} \theta^2 + \theta^2$$

$$= \frac{2\theta^2}{(n+1)(n+2)}$$

In particolare anche questo stimatore, visto come successione  $V_n = \max(X_1, \dots, X_n)$  costituisce una successione consistente di stimatori

Per la disuguaglianza di Markov

$$P[|V_n - \theta| \geq \varepsilon] = P[(V_n - \theta)^2 \geq \varepsilon^2] \leq \frac{1}{\varepsilon^2} R_{\theta}(V_n) = \frac{2\theta^2}{\varepsilon^2(n+1)(n+2)}$$

• Confronto

$$R_{\theta}(U) = \frac{\theta^2}{3n}$$

$$R_{\theta}(V) = \frac{2\theta^2}{(n+1)(n+2)}$$

quindi  $V$  è preferibile a  $U$  (perché  $(n+1)(n+2) > 3n$ )

Una media empirica che è corretta è automaticamente anche consistente (LGN)

Modelli esponenziali Supponiamo  $\Theta \subset \mathbb{R}$

• caso discreto: sia  $m_{\theta}$  probabilità sugli interi con densità  $p_{\theta}$  tale che esistano  $c_{\theta} > 0$  e due funzioni

$$T: \mathbb{N} \rightarrow \mathbb{R} \text{ e } g: \mathbb{N} \rightarrow \mathbb{R}$$

$$\text{tali che } p_{\theta}(k) = c_{\theta} g(k) e^{\theta T(k)} \quad k \in \mathbb{N}$$

• caso continuo: sia  $m_{\theta}$  probabilità su  $\mathbb{R}$  con densità  $f_{\theta}$  tale che esistano  $c_{\theta} > 0$  e due funzioni misurabili

$$T: \mathbb{R} \rightarrow \mathbb{R} \text{ e } g: \mathbb{R} \rightarrow \mathbb{R}$$

$$\text{tali che } f_{\theta}(x) = c_{\theta} g(x) e^{\theta T(x)} \quad x \in \mathbb{R}$$

esempi 1. popolazione geometrica

$$p_{\theta}(k) = p(1-p)^{k-1} = \underbrace{\frac{p}{1-p}}_{c_p} e^{\underbrace{k \log(1-p)}_{T(k) \xrightarrow{g=1}}}$$

2. popolazione esponenziale

$$f_{\theta}(x) = \lambda e^{-\lambda x} \mathbb{1}_{[0, \infty)}(x) = \underbrace{\lambda}_{c_{\lambda}} \underbrace{\mathbb{1}_{[0, \infty)}(x)}_{g(x)} e^{\lambda \underbrace{(-x)}_{T(x)}}$$

3. popolazione uniforme

$$f_{\theta}(x) = \frac{1}{\theta} \mathbb{1}_{[0, \theta]}(x)$$

è questo un modello esponenziale?

osservazione: lo stimatore di massima verosimiglianza è per definizione a valori in  $\Theta$

lezione 17  
04-05-2021  
(Romito)

teorema

Supponiamo  $\Theta \subset \mathbb{R}$  aperto. Supponiamo che la famiglia  $(m_{\theta})_{\theta \in \Theta}$  sia un modello esponenziale, ovvero se  $f_{\theta}$  è la densità associata a  $m_{\theta}$  allora

$$f_{\theta}(x) = c_{\theta} g(x) e^{\theta T(x)}$$

dove  $c_{\theta} > 0$  e  $g: \mathbb{R} \rightarrow \mathbb{R}$ ,  $T: \mathbb{R} \rightarrow \mathbb{R}$  funzioni misurabili tali che

$$\bullet x \mapsto g(x) T(x)^2 e^{\theta T(x)} \text{ integrabile per ogni } \theta \in \Theta$$

Consideriamo per ogni  $n$  un campione di numerosità  $n$  estratto da una popolazione di legge  $m_{\theta}$  e supponiamo

- lo stimatore  $\hat{\theta}_n$  di massima verosimiglianza esiste in  $\Theta$  per ogni  $n$

Allora  $(\hat{\theta}_n)_{n \geq 1}$  è consistente

dim. Poniamo

$$\psi(\theta) = \log \int g(x) e^{\theta^T(x)} dx = -\log c_\theta \quad c_\theta = e^{-\psi(\theta)}$$

Calcoliamo la derivata di  $\psi$ . Ricordiamo il teorema di derivazione sotto il segno di integrale

Data  $f(t, x)$  tale che

- $x \mapsto f(t, x)$  integrabile per ogni  $t$
- $t \mapsto f(t, x)$  derivabile per q.o.  $x$
- $|\partial_t f(t, x)| \leq h(x)$ , con  $h$  integrabile

$$\text{allora } \frac{d}{dt} \int f(t, x) dx = \int \partial_t f(t, x) dx$$

In fatti

$$\begin{aligned} \frac{d}{d\theta} \psi(\theta) &= \frac{1}{\int g(x) e^{\theta^T(x)} dx} \frac{d}{d\theta} \int g(x) e^{\theta^T(x)} dx = c_\theta \int g(x) T(x) e^{\theta^T(x)} dx = \\ &= \int T(x) f_\theta(x) dx \end{aligned}$$

Ricordiamo che il campione  $X_1 \dots X_n$  ha legge  $m_\theta$   
 $= \mathbb{E}[T(X_1)]$

$$\begin{aligned} \frac{d^2}{d\theta^2} \psi(\theta) &= \frac{d}{d\theta} \psi'(\theta) = \frac{d}{d\theta} \left( e^{-\psi(\theta)} \int g(x) T(x) e^{\theta^T(x)} dx \right) \\ &= -(\psi'(\theta))^2 + c_\theta \int g(x) T(x)^2 e^{\theta^T(x)} dx \\ &= -(\mathbb{E}_\theta[T(X_1)])^2 + \int T(x)^2 f_\theta(x) dx \\ &= -(\mathbb{E}_\theta[T(X_1)])^2 + \mathbb{E}[T(X_1)^2] \\ &= \text{Var}_\theta(T(X_1)) \end{aligned}$$

In particolare  $\psi''(\theta) = \text{Var}_\theta(T(X_1))$

Mostriamo che  $\psi''(\theta) > 0$  per ogni  $\theta$

Supponiamo invece che  $\psi''(\theta) = 0$  per un valore di  $\theta$ . Quindi

$$\text{Var}(T(X_1)) = 0 \longrightarrow T(X_1) \text{ è costante } m_\theta - \text{q.o.}$$

Ma le  $m_\theta$  sono mutuamente assolutamente continue e assolutamente continue con la misura di Lebesgue. Quindi  $T$  costante q.o. contro l'ipotesi che il modello sia una banda

Scriviamo la verosimiglianza

$$L_\theta(x_1, \dots, x_n) = f_\theta(x_1) f_\theta(x_2) \dots f_\theta(x_n)$$

$$\begin{aligned} \log L_\theta(x_1, \dots, x_n) &= \sum_{i=1}^n \log f_\theta(x_i) = \sum_{i=1}^n \log(e^{-\psi(\theta)} g(x_i) e^{\theta T(x_i)}) \\ &= -n\psi(\theta) + \sum_{i=1}^n g(x_i) + \theta \sum_{i=1}^n T(x_i) \end{aligned}$$

$$\frac{d}{d\theta} \log L_\theta(x_1, \dots, x_n) = -n\psi'(\theta) + \sum_{i=1}^n T(x_i)$$

$$\psi'(\hat{\theta}_n) = \frac{1}{n} \sum_{i=1}^n T(x_i)$$

$$\hat{\theta}_n = (\psi')^{-1}\left(\frac{1}{n} \sum_{i=1}^n T(x_i)\right)$$

Concludiamo: fissato  $\theta \in \Theta$

• per la legge dei grandi numeri (in  $(\Omega, \mathcal{F}, P_\theta)$ )

$$\frac{1}{n} \sum_{i=1}^n T(x_i) \xrightarrow{\text{in probabilità}} \mathbb{E}_\theta[T(x_1)] = \psi'(\theta)$$

Poiché  $(\psi')^{-1}$  è continua

$$\hat{\theta}_n = (\psi')^{-1}\left(\frac{1}{n} \sum_{i=1}^n T(x_i)\right) \xrightarrow{\text{in prob.}} (\psi')^{-1}(\psi'(\theta)) = \theta$$

□

Lemma Se  $(X_n)_{n \geq 1}$  v.a. e supponiamo  $X_n \rightarrow \ell$  in probabilità

dire Fissato  $\varepsilon > 0$  per continuità esiste  $\delta > 0$  tale che

$$|x - \ell| < \delta \Rightarrow |\varphi(x) - \varphi(\ell)| < \varepsilon$$

Quindi

$$\{|\varphi(x_n) - \varphi(\ell)| \geq \varepsilon\} \subset \{|x_n - \ell| \geq \delta\}$$

Quindi

$$P[|\varphi(x_n) - \varphi(\ell)| \geq \varepsilon] \leq P[|x_n - \ell| \geq \delta] \xrightarrow{n \rightarrow \infty} 0$$

□

lo stesso teorema nel caso discreto

Supponiamo  $(H) \subset \mathbb{R}$  e  $(m_\theta)_{\theta \in H}$  Sia un modello esponenziale  
ovvero

$$P_\theta(k) = C_\theta g(k) e^{\theta T(k)} \quad k \in \mathbb{N}$$

tale che

$$\sum_{k=0}^{\infty} C_\theta g(k) T(k)^2 e^{\theta(T(k))} < \infty \quad \text{per ogni } \theta \in (H)$$

Consideriamo per ogni  $n$  un campione  $X_1, \dots, X_n$  di  
dimensione  $n$  e legge  $m_\theta$  e supponiamo

- lo stimatore  $\hat{\theta}_n$  di massima verosimiglianza esiste in  $(H)$   
per ogni  $n$ , q.c.

Allora  $(\hat{\theta}_n)_{n \geq 1}$  è consistente

Osservazioni 1. lo stesso risultato vale nel caso  
multidimensionale  $(H) \subset \mathbb{R}^d$

2. lo stesso risultato vale se

$$f_\theta(x) = C_\theta g(x) e^{\ell(\theta)T(x)}$$

con  $\ell$  differenziabile e iniettiva

### REGIONI DI FIDUCIA

Regione di fiducia Dati un modello statistico  $(\Omega, \mathcal{F}, P_\theta)_{\theta \in (H)}$

e un numero  $d \in (0, 1)$ , un sottoinsieme **aleatorio**  $D \subset (H)$

è una regione di fiducia per il parametro  $\theta$  al livello  
di fiducia  $1-d$  se per ogni  $\theta$

$$P_\theta [\underbrace{\theta \in D}_{\{w: \theta \in D(w)\}}] \geq 1-d$$

Esempi 1. popolazione di Bernoulli  $X_1, \dots, X_n$

$$P_p \left[ \frac{|X_1 + \dots + X_n - np|}{\sqrt{np(1-p)}} \leq q \right] \approx 2\phi(q) - 1$$

(  $\Phi$  : funzione di ripartizione di una Gaussiana standard )

Dato  $\alpha$

$$2\Phi(q) - 1 = 1 - \alpha \longrightarrow \Phi(q) = 1 - \frac{\alpha}{2}$$

ovvero  $q = q_{1-\frac{\alpha}{2}}$  (quantile di una Gaussiana standard)

Osserviamo  $\sqrt{np(1-p)} \leq \frac{\sqrt{n}}{2}$ . Quindi

$$p \in \left( \bar{x} - \frac{\sqrt{n}}{2} q_{1-\frac{\alpha}{2}} \right)$$

con probabilità (approssimativamente)  $2\Phi(q_{1-\frac{\alpha}{2}}) - 1 = 1 - \alpha$

2. Popolazione Gaussiana di legge  $N(m, \sigma^2)$  e supponiamo di conoscere il valore di  $\theta$ . Allora

$$Z = \frac{X_1 + \dots + X_n - nm}{\sigma\sqrt{n}}$$

$$m = \bar{x} - \frac{\sigma}{\sqrt{n}} Z$$

$Z$  ha legge  $N(0, 1)$ . Come prima

$$m \in \left( \bar{x} - \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}, \bar{x} + \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} \right) \quad \text{con probabilità } 1 - \alpha$$

Quantità pivotale Una quantità pivotale è una funzione del campione e del parametro t.c.

1. invertibile rispetto a  $\theta$  (dato il campione)

2. la sua legge non dipende dal parametro

quantile Sia  $F$  una funzione di ripartizione. Se  $\alpha \in (0, 1)$ , il quantile di ordine  $\alpha$  per  $F$  è ogni numero  $q$  tale che

$$F(q) \geq \alpha$$

$$F(q^-) \leq \alpha$$

$$\text{dove } F(q^-) = \lim_{y \rightarrow q^-} F(y)$$

osservazione Ogni mediana è un quantile di ordine  $1/2$

esercizio Dato  $\alpha \in (0, 1)$

$$q_1 = \inf \{ x : F(x) \geq \alpha \} \quad q_2 = \sup \{ x : F(x^-) \leq \alpha \}$$

allora  $\cdot q_1 \leq q_2$

$\cdot$  ogni  $x \in [q_1, q_2]$  è un quantile di ordine  $\alpha$

## TEST STATISTICI

Qual'è l'idea? Abbiamo un fenomeno da studiare e lo parametrizziamo. Ad esempio nel lancio di una moneta definiamo la quantità che indica la probabilità che esca una faccia piuttosto che l'altra; poi possiamo formulare un modello in cui ogni lancio è indipendente dagli altri ma ha la stessa distribuzione, cioè che esca la tal faccia con tale probabilità, che è il nostro parametro.

Vogliamo mettere alla prova questo modello, in che modo?

Testando le ipotesi: noi ci concentreremo sul test statistico relativo ai parametri in gioco, ma è possibile mettere alla prova tutte le ipotesi.

Ad esempio: verificare una certa distribuzione (test del  $\chi^2$ ) ecc...

Cosa può andar bene? Cosa può andar male? Possiamo basarci solo sul campione, dunque l'idea di fondo è quella di decidere se il valore del parametro che ci è stato proposto è compatibile col campione oppure no. Il test ha un risultato significativo quando riusciamo a rifiutare l'ipotesi che ci è stata proposta. Al contrario, se non riusciamo a rifiutare l'ipotesi avremmo trovato una istanza del campione che funziona, ce ne potrebbero essere altre in cui non funziona. Non si rigetta la realtà ma il modello. Se non possiamo rigettare il modello, lo si tiene finché non si trova un'istanza per il quale non funziona. I test statistici sono la base del nostro approccio quantitativo alla realtà.

Problema 4

Sapendo che, per definizione di massimo, per un certo  $x$ ,  $P[\max\{X_1, \dots, X_n\} \leq x] = P[X_1 \leq x, \dots, X_n \leq x]$

si ha che la f. di distribuzione di  $U = \max\{X_1, \dots, X_n\}$  è  $F_U(u) = P[U \leq u] = P[\max\{X_1, \dots, X_n\} \leq u] = P[X_1 \leq u, \dots] =$   
 $= P[X_1 \leq u] \dots P[X_n \leq u] = F_1(u) \dots F_n(u)$

la densità di  $U$ ,  $f_U(u)$ , si trova facendo la derivata di  $F_U(u)$ :  $f_U(u) = F_U'(u) = \frac{d(F_1(u) \dots F_n(u))}{du}$

Per il minimo, invece, sapendo per def., per un certo  $x$ ,  $P[\min\{X_1, \dots, X_n\} > x] = P[X_1 > x, \dots, X_n > x]$

si ha che la funz. di distribuzione di  $W = \min\{X_1, \dots, X_n\}$  deriva da  $1 - F_W(w) = P[W > w] = P[X_1 > w] \dots P[X_n > w] =$   
 $= (1 - P[X_1 \leq w]) \dots = (1 - F_1(w)) \dots$

Perciò  $F_W(w) = 1 - (1 - F_1(w)) \dots (1 - F_n(w))$

la densità di  $W$ ,  $f_W(w)$ , si trova facendo la derivata di  $F_W(w)$ :  $f_W(w) = F_W'(w) = \frac{d((1 - F_1(w)) \dots (1 - F_n(w)))}{dw}$

Nel caso in cui  $f_1 = \dots = f_n = f$  (quindi  $F_1 = \dots = F_n = F$ ) si ha:

$$f_U(u) = F_U'(u) \ni f_U(u) = \frac{dF^n(u)}{du} = n F^{n-1}(u) F'(u) = n F^{n-1}(u) f(u)$$

$$f_W(w) = F_W'(w) \ni f_W(w) = -\frac{d(1 - F(w))^n}{dw} = n(1 - F(w))^{n-1} f(w)$$

In particolare se  $f$  è la densità di una v.a.

uniforme su  $(0, 1)$  si ha  $f(x) = 1$  e  $F(x) = x$

Perciò si ha  $f_U(u) = n u^{n-1}$  e  $f_W(w) = n(1-w)^{n-1}$

## Problema 5

$$\begin{aligned}
 \bullet \mathbb{P}[Y < 2X] &= \int_{\mathbb{R}^2} f_{(x,y)}(x,y) dy dx = \iint \frac{3}{x^3 y^2} \mathbb{1}_{\{x>1, y>x\}} dy dx = \\
 &= 3 \int_1^{+\infty} \frac{1}{x^3} \int_x^{2x} \frac{1}{y^2} dy dx = 3 \int_1^{+\infty} \frac{1}{x^3} \left[ -\frac{1}{y} \right]_x^{2x} dx = 3 \int_1^{+\infty} \frac{1}{2x^4} dx = \\
 &= \frac{1}{2} \left[ -\frac{1}{x^3} \right]_1^{+\infty} = \frac{1}{2}
 \end{aligned}$$

$\phi: A \rightarrow B$ ,  $\phi(x,y) = \left( \log x, \frac{x}{y} \right)$ , dove  $A = \{(x,y) \in \mathbb{R}^2 \mid x>1, y>x\}$   
 $B = \{(u,v) \in \mathbb{R}^2 \mid u>0, v \in (0,1)\}$  è un diffeomorfismo.  
 $\phi^{-1}(u,v) = (x(u,v), y(u,v)) = \left( e^u, \frac{e^u}{v} \right)$  perché  $\phi \circ \phi^{-1} = \text{Id}$

$$\begin{vmatrix} \frac{dx}{du}(u,v) & \frac{dx}{dv}(u,v) \\ \frac{dy}{du}(u,v) & \frac{dy}{dv}(u,v) \end{vmatrix} = \begin{vmatrix} e^u & 0 \\ \frac{e^u}{v} & -\frac{e^u}{v} \end{vmatrix} = \frac{e^{2u}}{v^2}$$

$\phi(x,y) = (u,v)$ , quindi, per teo. di cambio di variabile,

$$\begin{aligned}
 f_{(u,v)}(u,v) &= f_{(x,y)}(x(u,v), y(u,v)) |J\phi^{-1}(u,v)| = \frac{3v^2}{e^{3u} e^{2u}} \mathbb{1}_B \frac{e^{2u}}{v^2} = \\
 &= 3e^{-3u} \mathbb{1}_{\{u>0, v \in (0,1)\}} = 3e^{-3u} \mathbb{1}_{u>0} \mathbb{1}_{v \in (0,1)}
 \end{aligned}$$

Quindi  $f_u(u) = 3e^{-3u} \mathbb{1}_{u>0}$  e  $f_v(v) = \mathbb{1}_{v \in (0,1)}$ , cioè

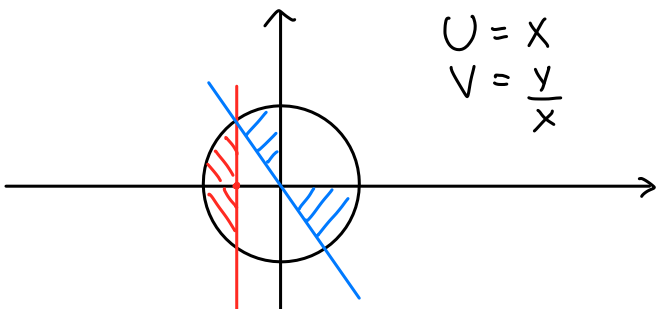
$U \sim \exp(3)$ ,  $V \sim \text{Unif}(0,1)$  e  $U \perp\!\!\!\perp V$

↳ indipendenti

## Problema 6

sol. di gdg per  
l'indipendenza

$$\begin{aligned}
 U &= X \\
 V &= \frac{y}{x}
 \end{aligned}$$



indipendenza  $\Leftrightarrow \forall u, v$

$$\mathbb{P}(U \leq u, V \leq v) \neq \mathbb{P}(U \leq u) \mathbb{P}(V \leq v)$$

$$\mathbb{P}(\emptyset) = 0$$

$$0$$

→ non sono indipendenti

$V \leq -1 \rightarrow$  punti che hanno  $\tan \leq -1$

$U \leq -\frac{\sqrt{2}}{2} \rightarrow$  punti che hanno  $\cos \leq -\frac{\sqrt{2}}{2}$

## Problema 2

- Verificare che  $p$  è una densità.

i) positività:  $p(x) = \frac{1}{\pi(1+x^2)} \geq 0 \quad \forall x \in \mathbb{R}$

ii) Dobbiamo vedere che  $\int_{\mathbb{R}} p(x) = 1$

$$\int_{-\infty}^{+\infty} \frac{1}{\pi(1+x^2)} dx = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1}{1+x^2} dx = \frac{1}{\pi} [\arctan(x)]_{-\infty}^{+\infty} = \frac{1}{\pi} \left( \frac{\pi}{2} + \frac{\pi}{2} \right) = 1$$

- Funzione di ripartizione

$$F_X(z) = \int_{-\infty}^z \frac{1}{\pi(1+x^2)} dx$$

Cauchy ammette densità dunque è assolutamente continua

$$= \dots = \frac{1}{\pi} \arctan(z) + \frac{1}{2}$$

- Media e Varianza

Non ha media perché  $\int_{-\infty}^{+\infty} |x| p(x) dx < +\infty$

**[NB]** Possiamo vedere "a occhio" che l'integrale improprio non converge perché  $\frac{1}{x^d}$  converge solo per  $d > 1$  (sia in  $\Sigma$  che in  $S$ )

**[NB2]** La LGN non vale per Cauchy  $\rightarrow$  perciò  $\nexists \mathbb{E}[X]$

$$\text{Var}[X] = \mathbb{E}[X^2] - \underbrace{\mathbb{E}[X]}_{\text{non esiste}}$$

$\Rightarrow$  Di conseguenza neanche la varianza esiste

- Calcolare la legge di  $Y = X^2$

Attenzione!  $\varphi: \mathbb{R} \rightarrow \mathbb{R}^+$

$$x \mapsto x^2$$

$$\varphi^{-1}: \mathbb{R} \rightarrow \mathbb{R}$$

$$y \mapsto \sqrt{y}$$

**NON È UN  
DIFEOMORFISMO**

( $\varphi$  non è bigettiva)

Per calcolare la legge di  $Y$  si potrebbe passare alla funzione di ripartizione e poi derivare.

- $Y$  ha media? No perché  $\mathbb{E}[Y] = \mathbb{E}[X^2]$  e poiché  $\mathbb{E}[X]$  non esiste, non ci può essere momento secondo.

**[NB]**

$$E[X] = 0 \neq \cancel{E[X]}$$

$$E[X] = 0 \neq E[X^2] = 0$$

$$\cancel{E[X]} \Rightarrow \cancel{E[X^2]}$$

• -  $Z = \frac{1}{X}$  è ben posta? Sì perché, essendo  $X$  una

variabile aleatoria assolutamente continua, si ha

$P[X=x]=0$ , in particolare  $P[X=0]=0$ , dunque l'espressione  $\frac{1}{x}$  è ben definita per q.o.  $\omega \in \Omega$ .

Poiché  $x$  è un punto a misura trascurabile

**[NB]**  $P[X \in A] = \int_A f_x(x) dx \longrightarrow P[X=a] = \int_a^a f_x(x) dx = 0$

↳ vale solo per v.a. assolutamente continue

- Calcolare la legge di  $Z$

$$\varphi: \mathbb{R} \setminus \{0\} \longrightarrow \mathbb{R} \setminus \{0\}$$

$$t \longmapsto \frac{1}{t}$$

$$\varphi^{-1}: \mathbb{R} \setminus \{0\} \longrightarrow \mathbb{R} \setminus \{0\}$$

$$u \longmapsto \frac{1}{u}$$

$\varphi$  È UN  
DIFFEOMORFISMO

$$\frac{d}{du} \varphi^{-1}(u) = -\frac{1}{u^2} \longrightarrow \text{con la formula del cambio di variabile si ottiene la densità}$$

Soluzione gdg:

$$\text{ii) } P[Y \leq y] = P[X^2 \leq y] = P[|X| \leq \sqrt{y}] = P[-\sqrt{y} \leq X \leq \sqrt{y}] =$$

con la derivata =  $\frac{1}{\pi} \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{1+x^2} dx = F_x(\sqrt{y}) - F_x(-\sqrt{y}) = \frac{2}{\pi} \arctan(\sqrt{y})$   
si trova la legge

iii)  $z > 0$

$$P[Z \leq z] = P\left[\frac{1}{X} \leq z\right] = P\left[\frac{1}{X} \leq z, X > 0\right] + P\left[\frac{1}{X} \leq z, X < 0\right] = P\left[X \geq \frac{1}{z}, X > 0\right] + P\left[X \leq \frac{1}{z}, X < 0\right] = P\left[X \geq \frac{1}{z}\right] + P[X < 0] \dots$$

## TEST STATISTICI

test  $(\Omega, \mathcal{F}, P_\theta)_{\theta \in \Theta}$  e un campione  $X_1, \dots, X_n$

1. Formulare l'ipotesi: partizionare  $\Theta = \Theta_0 \cup \Theta_1$ ,  $\Theta_0 \cap \Theta_1 = \emptyset$

$\Theta_0$ : valori di  $\theta$  compatibili con l'ipotesi

$\Theta_1$ : " " " incompatibili

Ad esempio: lancio di monete,  $p$  = probabilità di successo

$$\begin{aligned} * \left| \begin{aligned} \mathcal{H}_0: p &= \frac{1}{2} & (\text{ipotesi nulla}) & \Theta_0 = \left\{ \frac{1}{2} \right\} \\ \mathcal{H}_1: p &\neq \frac{1}{2} & (\text{ipotesi alternativa}) & \Theta_1 = [0, 1] \setminus \left\{ \frac{1}{2} \right\} \end{aligned} \right. \end{aligned}$$

$$\begin{aligned} \left| \begin{aligned} \mathcal{H}_0: p &\geq \frac{1}{2} & (\text{ipotesi nulla}) & \Theta_0 = \left[ \frac{1}{2}, 1 \right] \\ \mathcal{H}_1: p &< \frac{1}{2} & (\text{ipotesi alternativa}) & \Theta_1 = \left[ 0, \frac{1}{2} \right) \end{aligned} \right. \end{aligned}$$

2. pianificazione: stabilisce quali risultati portano a rigettare o meno l'ipotesi nulla

Identificare un evento: regione critica o regione di rifiuto

Ad esempio, nel contesto delle ipotesi (\*)

$$C = \left\{ w : \left| \bar{X} - \frac{1}{2} \right| > \text{"abbastanza grande"} \right\}$$

$p$

$\uparrow$  è dipendente dalla num. campione

ERRORI: Come si può sbagliare in un test

errore di 1<sup>a</sup> specie: rigettare un'ipotesi nulla corretta

errore di 2<sup>a</sup> specie: non rigettare un'ipotesi nulla non corretta

(contrapposti a esiti fausti del test:

$\downarrow$  insiste sui parametri dell'ipotesi alternativa

- rigettare un'ipotesi nulla non corretta

- non rigettare un'ipotesi nulla corretta)

livello di un test Dati un  $C$  regione critica e  $d \in [0, 1]$

Il test di regione critica  $C$  ha livello  $d$  se

se  $d$  è molto piccolo è improbabile compiere un errore di 1ª specie

$$\sup_{\theta \in \Theta_0} P_{\theta}[C] = d$$

il più grande errore di 1ª specie

( $\leq d$  va bene uguale)

Che succede se l'esito capita nella regione critica?  
Rigettiamo l'ipotesi nulla

L'errore di 2ª specie non si può controllare

potenza la potenza di un test di regione critica  $C$  è la funzione  $\pi_C: \Theta_1 \rightarrow [0, 1]$

definita come

$$\pi_C(\theta) = P_{\theta}[C]$$

capacità di accorgersi che l'ipotesi nulla non è corretta, noto il valore del parametro  $\theta$

errore di 2ª specie:  $1 - \pi_C(\theta)$

I  $\theta$  parametri sono pescati tra i parametri dell'ipotesi alternativa

③ Stabilire il livello  $d$  del test e, tra i test di livello  $d$ , scegliere quello di potenza massima

esempio Svolgere  $n$  lanci ( $n = 1000$  e otteniamo 545 teste,  $p_0 = \frac{1}{2}$ )

$X_1, \dots, X_n$  Bernoulli di parametro  $p$   $p \in [0, 1]$

$$H_0: p = p_0$$

$$H_1: p \neq p_0$$

$\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$  è uno stimatore (consistente) di  $p$

Cerchiamo una regione critica del tipo

$$C = \{ |\bar{X} - p_0| > \delta \}$$

determiniamo  $\delta$  in modo che il test di livello  $d$ , per un  $d$  prefissato

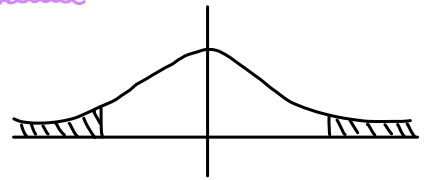
$$\Theta_0 = \{p_0\} \quad \Theta_1 = [0, 1] \setminus \{p_0\} \quad \text{ipotesi semplice}$$

Per stabilire che il test ha livello  $d$  è sufficiente

$$P_{p_0}[C] = d \quad (\leq d \text{ ma anche } \approx d)$$

$$P_{p_0}[C] = P_{p_0}[|\bar{X} - p_0| > \delta] = P_p \left[ \underbrace{\sqrt{n} \frac{|\bar{X} - p_0|}{\sqrt{p_0(1-p_0)}}}_{\text{approx. Gaussiana}} > \frac{\sqrt{n}}{\sqrt{p_0(1-p_0)}} \delta \right]$$

$$\approx 2 \left( 1 - \Phi \left( \sqrt{\frac{n}{p_0(1-p_0)}} \delta \right) \right)$$



$\Phi$ : funzione cumulativa di una Gaussiana standard

$$2 \left( 1 - \Phi \left( \sqrt{\frac{n}{p_0(1-p_0)}} \delta \right) \right) = \alpha$$

$$\Phi \left( \delta \sqrt{\frac{n}{p_0(1-p_0)}} \right) = 1 - \frac{\alpha}{2}$$

$$\delta \frac{\sqrt{n}}{\sqrt{p_0(1-p_0)}} = q_{1-\frac{\alpha}{2}} \longrightarrow \delta = \sqrt{\frac{p_0(1-p_0)}{n}} q_{1-\frac{\alpha}{2}}$$

La regione critica è

$$C = \left\{ |\bar{X} - p_0| > \sqrt{\frac{p_0(1-p_0)}{n}} q_{1-\frac{\alpha}{2}} \right\}$$

esempio 2 (stesso contesto)

$$H_0: p \leq p_0 \quad H_1: p > p_0$$

cerchiamo una regione critica del tipo  $\{\bar{X} > \delta\}$

Il livello  $\sup_{p \leq p_0} P_p[C]$

$$P[\bar{X} > \delta] = P_p \left[ \frac{\bar{X} - p}{\sqrt{p(1-p)}} \sqrt{n} > \frac{\delta - p}{\sqrt{p(1-p)}} \sqrt{n} \right]$$

$$\leq P_p \left[ \frac{\bar{X} - p}{\sqrt{p(1-p)}} \sqrt{n} > \frac{\delta - p_0}{\sqrt{p_0(1-p_0)}} \right]$$

$$\stackrel{CLT}{\approx} 1 - \Phi \left( \frac{\delta - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n} \right)$$



in quanto  $\frac{\delta - p}{\sqrt{p(1-p)}} \leq \frac{\delta - p_0}{\sqrt{p_0(1-p_0)}} \quad \text{se } p \leq p_0$

Determiniamo  $\delta$

$$1 - \Phi \left( \frac{\delta - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n} \right) = \alpha$$

ovvero

$$\hat{p} = p_0 + \sqrt{\frac{p_0(1-p_0)}{n}} q_{1-\alpha}$$

legame tra Regioni di fiducia e regioni critiche di test con ipotesi semplici

ipotesi semplice  $\mathcal{H}_0 = \{\theta_0\}$   $\mathcal{H}_1 = \mathcal{H} \setminus \{\theta_0\}$

• se  $D$  è una regione di fiducia al livello  $1-\alpha$ , allora

$$C = \{w : \theta_0 \notin D(w)\}$$

è una regione critica di livello  $\alpha$

$$P_{\theta_0}[C] = 1 - P_{\theta_0}[\theta_0 \in D] = 1 - (1-\alpha) = \alpha$$

• Viceversa, se  $(C_{\theta_0})_{\theta_0 \in \mathcal{H}}$  sono regioni critiche nel senso che per ogni  $\theta_0$ :  $C_{\theta_0}$  è regione critica di livello  $\alpha$  per il test  $H_0: \theta = \theta_0$   $H_1: \theta \neq \theta_0$

Allora

$$D(w) = \{\theta \in \mathcal{H} : w \notin C_{\theta}\}$$

è una regione di fiducia al livello  $1-\alpha$

Modelli a rapporto di verosimiglianza crescente

$(\Omega, \mathcal{F}, P_{\theta})_{\theta \in \mathcal{H}}$ ,  $(x_1, \dots, x_n)$ , supponiamo  $\mathcal{H} \subset \mathbb{R}$  intervallo

Il modello è a rapporto di verosimiglianza crescente

rispetto a una variabile aleatoria  $T$  se per ogni  $\theta_1 < \theta_2$

$$\frac{L_{\theta_2}(x_1, \dots, x_n)}{L_{\theta_1}(x_1, \dots, x_n)} \text{ è una funzione strettamente crescente di } T$$

esempio popolazione di Bernoulli

$$L_{\theta}(x_1, \dots, x_n) = \theta^{x_1 + \dots + x_n} (1-\theta)^{n-(x_1 + \dots + x_n)}$$

$$x_i \in \{0, 1\}$$

Allora  $\frac{L_{\theta_2}}{L_{\theta_1}} = \left(\frac{\theta_2}{\theta_1}\right)^{n\bar{x}} \left(\frac{1-\theta_2}{1-\theta_1}\right)^{n(1-\bar{x})}$   $\theta_1 < \theta_2$

è str. crescente rispetto a  $T = \bar{x}$

$\frac{L_{\theta_2}(X_1, \dots, X_n)}{L_{\theta_1}(X_1, \dots, X_n)}$  è funzione strett. crescente di  $T$   
 $\theta_1 < \theta_2$

teorema (test unilatero)

$(\Omega, \mathcal{F}, P_\theta)_{\theta \in \Theta}$   $(X_1, \dots, X_n)$  campione. Supponiamo  $\Theta \subset \mathbb{R}$   
modello a verosimiglianza crescente rispetto a una v.a.  $T$

Consideriamo il test

$$H_0: \theta \leq \theta_0 \quad H_1: \theta > \theta_0 \quad e \quad C = \{T > \delta\}$$

Allora  $\sup_{\theta \leq \theta_0} P_\theta(C) = P_{\theta_0}(C)$

- il test con regione critica  $C$  è più potente di ogni altro test a livello  $P_{\theta_0}(C)$

Per dimostrare questo, usiamo

lemma  $(\Omega, \mathcal{F}, P_\theta)_{\theta \in \Theta}$   $(X_1, \dots, X_n)$ , e  $\Theta = \{\theta_0, \theta_1\}$

Sia  $L$  la funzione di verosimiglianza Per  $c > 0$

$$C = \{L_{\theta_0} \leq c L_{\theta_1}\}$$

ed il test

$$H_0: \theta = \theta_0 \quad H_1: \theta = \theta_1$$

Allora

- $P_{\theta_0}(C) \leq P_{\theta_1}(C)$
- $C$  è la regione critica di un test più potente di ogni altro test di livello  $P_{\theta_0}(C)$

dim Sia  $C_*$  un'altra regione critica

$$(\mathbb{1}_C^{(w)} - \mathbb{1}_{C_*}^{(w)})(L_{\theta_0}^{(w)} - c L_{\theta_1}^{(w)}) \leq 0$$

$$\begin{array}{lll} w \in C & \geq 0 & \leq 0 \\ w \notin C & \leq 0 & \geq 0 \end{array}$$

Ad esempio nel caso di campione con legge ass. continua

$$L_{\theta_0}(x_1, \dots, x_n) = f_{\theta_0}(x_1) f_{\theta_0}(x_2) \dots f_{\theta_0}(x_n)$$

$$\int (\mathbb{1}_C - \mathbb{1}_{C_*}) L_{\theta_1} = P_{\theta_1}(C) - P_{\theta_1}(C_*)$$

In conclusione

$$P_{\theta_0}[C] - P_{\theta_0}[C_*] \leq c (P_{\theta_1}(C) - P_{\theta_1}(C_*))$$

Se il test con regione critica  $C_*$  ha livello (al più)  $P_{\theta_0}[C]$

$$P_{\theta_0}[C_*] \leq P_{\theta_0}[C]$$

quindi

$$P_{\theta_1}(C_*) \leq P_{\theta_1}(C)$$

e il test con regione critica  $C$  è più potente

Per la disuguaglianza  $P_{\theta_0}[C] \leq P_{\theta_1}[C]$ , usiamo la disuguaglianza

$$(\mathbb{1}_C - P_{\theta_0}[C])(L_{\theta_0} - c L_{\theta_1}) \leq 0$$

integrando

$$0 = P_{\theta_0}(C) - P_{\theta_0}[C] \leq c (P_{\theta_1}(C) - P_{\theta_0}(C))$$

dim (teorema)

□

Dimostriamo che  $\theta \mapsto P_{\theta}(C)$  è monotona crescente

per  $\theta \leq \theta_0$  (quindi  $\sup_{\theta \leq \theta_0} P_{\theta}(C) = P_{\theta_0}(C)$ )

Prendiamo  $\theta_1 \leq \theta_2$

$$C = \{T > \delta\}$$

$$\text{Su } C \quad \frac{L_{\theta_2}}{L_{\theta_1}} \geq c \quad \text{ovvero} \quad L_{\theta_2} \geq c L_{\theta_1}$$

Dal lemma  $P_{\theta_1}(C) \leq P_{\theta_2}(C)$

Mostriamo adesso che il test con regione critica  $C$  è più

potente di ogni altro test di livello  $P_{\theta_0}(C)$ , ovvero che, se  $C_*$  è una regione critica di livello  $P_{\theta_0}(C)$ , cioè

$$\sup_{\theta \leq \theta_0} P_{\theta}[C_*] \leq P_{\theta_0}[C]$$

allora per ogni  $\theta > \theta_0$  :  $P_{\theta}[C_*] \leq P_{\theta}[C]$

Ancora per il lemma precedente applicato a  $\theta_0 = \theta$  con ipotesi nulla  $\theta_0$ .

$$P_{\theta}[C_*] \leq P_{\theta_0}[C]$$

quindi  $P_{\theta}[C_*] \leq P_{\theta}[C]$  (test più potente) □

Osservazione Un analogo risultato vale per il test

$$H_0 : \theta \geq \theta_0$$

$$H_1 : \theta < \theta_0$$

p-value

Supponiamo di essere nelle seguenti condizioni

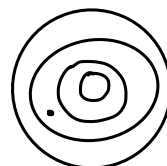
esiste una famiglia  $(C_d)_{d \in (0,1)}$  tale che

- $\bigcup C_d = \Omega$

- $\bigcap C_d = \emptyset$

- se  $d \leq d'$ , allora  $C_d \subset C_{d'}$

- $C_d$  è una regione critica di livello  $d$



$(X_1, \dots, X_n)$ , allora per ogni  $w \in \Omega$  esiste un unico  $p$  t.c.

- se  $d < p$  allora  $w \notin C_d$  (ipotesi nulla non rigettata)

- se  $d > p$  allora  $w \in C_d$  (ipotesi nulla rigettata)

$p$  = soglia di rigetto

Ad esempio nel lancio di monete :  $H_0 : p = \frac{1}{2}$

$$p\text{-value} = P_{\frac{1}{2}} \left[ \sqrt{n} \frac{|\bar{X} - \frac{1}{2}|}{\frac{1}{2}} > \sqrt{n} \frac{|\bar{x} - \frac{1}{2}|}{\frac{1}{2}} \right]$$

$x_1, \dots, x_n$   $\bar{x} \rightarrow$  i dati dell'esperimento realizzato

$X_1, \dots, X_n$   $\bar{X} \rightarrow$  campione nel modello

# POPOLAZIONI GAUSSIANE

$X_1, \dots, X_n$  hanno legge  $N(m, \sigma^2)$

stimatori di massima verosimiglianza

$$\mathcal{H} = \mathbb{R} \times (0, \infty) \quad f_{(m, \sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-m)^2}$$

$$L_{(m, \sigma^2)}(x_1, \dots, x_n) = (2\pi)^{-\frac{n}{2}} \sigma^{-n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2}$$

$$\log L_{(m, \sigma^2)} = -\frac{n}{2} \log 2\pi - n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2$$

$$\begin{cases} \frac{\partial}{\partial m} \log L(m, \sigma^2) = 0 \\ \frac{\partial}{\partial \sigma} \log L(m, \sigma^2) = 0 \end{cases} \quad \begin{cases} \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - m) = 0 \\ -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - m)^2 = 0 \end{cases}$$

$$\hat{m} = \bar{X}$$

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

Osserviamo però potremmo essere nel caso in cui il valore di  $m$  è noto. In tal caso lo stimatore di  $\sigma$  è  $\frac{1}{n} \sum_{i=1}^n (x_i - m)^2$

Modello Gaussiano è esponenziale

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-m)^2} = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}}_{C_\theta} e^{-\frac{m^2}{2\sigma^2}} \underbrace{e^{(x, -x^2)}}_{T(x)} \underbrace{\ell(m, \sigma^2)}_{\left(\frac{m}{\sigma^2}, \frac{1}{2\sigma^2}\right)}$$

$Q(m, \sigma^2)$  è invertibile e continua

Lemma  $X_1, \dots, X_n$  indipendenti e con legge  $N(0, 1)$

A matrice ortogonale. Allora  $A(X_1, \dots, X_n)$  è ancora un vettore di v.a. indipendenti e con legge  $N(0, 1)$

diue la densità di  $(X_1, \dots, X_n)$  è  $\frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}|x|^2}$

$y \mapsto Ay$   $\det(A) = \pm 1$ . Inoltre  $|Ax| = |A^{-1}x| = |x|$

$$\begin{aligned} f_{A(X_1, \dots, X_n)}(y_1, \dots, y_n) &= f_{X_1, \dots, X_n}(A^{-1}(y_1, \dots, y_n)) |\det A|^{-1} \\ &= f_{X_1, \dots, X_n}(y_1, \dots, y_n) \end{aligned}$$

teorema Se  $X_1, \dots, X_n$  indipendenti e con legge  $N(m, \sigma^2)$

$$\bar{X} = \frac{1}{n} (X_1 + \dots + X_n) \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Allora

- $\bar{X}, S^2$  sono indipendenti
- $\bar{X}$  ha distribuzione  $N(m, \frac{\sigma^2}{n})$
- $(n-1) \frac{S^2}{\sigma^2}$  ha distribuzione  $\chi^2(n-1)$
- $\frac{\bar{X} - m}{S/\sqrt{n}}$  ha distribuzione di Student  $t(n-1)$

dim  $X_i = m + \sigma z_i$   $z_i$  è  $N(0, 1)$

$$z_i = \frac{X_i - m}{\sigma}$$

$z_1, \dots, z_n$  indipendenti

Inoltre

$$\bar{X} = m + \sigma \bar{Z}$$

$$\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (z_i - \bar{Z})^2 =: U^2$$

Per dimostrare il teorema, basta far vedere

- $\bar{Z}, U^2$  indipendenti
- $U^2$  ha legge  $\chi^2(n-1)$
- $\bar{Z}$  ha legge  $N(0, \frac{1}{n})$
- $\sqrt{n(n-1)} \frac{\bar{Z}}{U}$  ha legge  $t(n-1)$

Consideriamo la matrice ortogonale  $A$  data dal cambio di base in  $\mathbb{R}^n$

$$e_1 = \frac{1}{\sqrt{n}} (1, \dots, 1) \quad e_2, \dots, e_n \text{ base di } e_1^\perp$$

Allora

$A(z_1, \dots, z_n)$  indipendenti e con legge  $N(0, 1)$

$$(AZ)_1 = \sqrt{n} \bar{Z}$$

$$\begin{aligned} |(A \cdot z)_2|^2 + \dots + |(A \cdot z)_n|^2 &= |A \cdot z|^2 - ((AZ)_1)^2 \\ &= |z|^2 - n(\bar{Z})^2 \end{aligned}$$

$$= U^2$$

- quindi
- $\bar{Z}$ ,  $U^2$  indipendenti (funzioni di v.a. indipendenti)
  - $U^2$  ha distribuzione  $\chi^2(n-1)$  in quanto somma dei quadrati di  $(n-1)$  v.a. Gaussiane standard indipendenti
  - $\sqrt{n(n-1)} \frac{\bar{Z}}{U}$  è Student con  $(n-1)$  gradi di libertà

Osservazione dedurre dal lemma precedente che se  $X, Y$  Gaussiane indipendenti allora  $X + Y$  è Gaussian

Esercitaz. 10  
12-05-2021  
(Di Gesù)

### Problema 1

$(X, Y)$  coord sul piano aleatorie

$X \sim N(0, 1)$     $Y \sim N(0, 1)$     $X$  e  $Y$  indep.

$(R, \theta)$  coord polari del punto  $(X, Y)$

$$R^2 = X^2 + Y^2 \sim \chi(2) = \text{Gamma}\left(1, \frac{1}{2}\right) = \exp\left(\frac{1}{2}\right)$$

$$R^2 \sim \exp\left(\frac{1}{2}\right)$$

$$P[R^2 \leq r] = P[X^2 + Y^2 \leq r]$$

$$\text{Gamma}(r, \lambda) = \frac{1}{\Gamma(r)} \lambda^r x^{r-1} e^{-\lambda x} \mathbb{1}_{[0, +\infty)}$$

$$\text{se } r=1 \rightarrow \exp(\lambda)$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$$

$$f_{(X,Y)}(x,y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} \quad , \quad f_{(R,\theta)} = \frac{\rho}{2\pi} e^{-\frac{1}{2}\rho^2}$$

$$f_\theta(\theta) = \int_{\mathbb{R}} f_{(R,\theta)}(\rho, \theta) d\rho = \int_0^\infty \frac{1}{2\pi} \rho e^{-\frac{1}{2}\rho^2} d\rho = \frac{1}{2\pi}$$

$$f_R(\rho) = \int_0^{2\pi} f_{(R,\theta)} d\theta = \int_{(R,\theta)} 2\pi = \rho e^{-\frac{1}{2}\rho^2}$$

$$R \text{ e } \theta \text{ indep } f_\theta(\theta) f_R(\rho) = f_{(R,\theta)}(\rho, \theta)$$

$$\begin{aligned} X &= R \cos \theta \\ Y &= R \sin \theta \end{aligned}$$

$$\mathbb{E}\left[\frac{X^2}{R^2}\right] = \mathbb{E}\left[\frac{R^2 \cos^2 \theta}{R^2}\right] = \mathbb{E}[\cos^2 \theta] = \mathbb{E}[g(\theta)] \\ = \int_0^\pi \cos^2 \theta f_\theta(\theta) d\theta = \frac{1}{2\pi} \int_0^{2\pi} \cos^2 \theta d\theta = \frac{1}{2}$$

$$\mathbb{E}\left[\frac{X^2}{R^2}\right] + \mathbb{E}\left[\frac{Y^2}{R^2}\right] = 1$$

$$\bullet \mathbb{E}\left[\frac{\min\{|X|, |Y|\}}{\max\{|X|, |Y|\}}\right] = \mathbb{E}\left[g(X, Y)(\mathbb{1}_{\{|X| \geq |Y|\}} + \mathbb{1}_{\{|Y| > |X|\}})\right] \\ g''(X, Y)$$

$$\mathbb{E}[g(X, Y)\mathbb{1}_{\{|X| \geq |Y|\}}] + \mathbb{E}[g(X, Y)\mathbb{1}_{\{|Y| > |X|\}}]$$

$$\mathbb{E}\left[\frac{|Y|}{|X|} \mathbb{1}_{\{|X| \geq |Y|\}}\right] + \mathbb{E}\left[\frac{|X|}{|Y|} \mathbb{1}_{\{|Y| > |X|\}}\right]$$

$$\mathbb{E}[|\tan \theta| \mathbb{1}_{\{-\frac{\pi}{4} \leq \theta \leq \frac{\pi}{4}, \frac{3}{4}\pi \leq \theta \leq \frac{5}{4}\pi\}}] + \mathbb{E}[|\cot \theta| \mathbb{1}_{\{\frac{\pi}{4} \leq \theta \leq \frac{3}{4}\pi, \frac{5}{4}\pi \leq \theta \leq \frac{7}{4}\pi\}}]$$

$$\int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \frac{|\tan \theta|}{2\pi} d\theta + \int_{\frac{3}{4}\pi}^{\frac{5}{4}\pi} \frac{|\tan \theta|}{2\pi} d\theta + \int_{\frac{\pi}{4}}^{\frac{3}{4}\pi} \frac{|\cot \theta|}{2\pi} d\theta + \int_{\frac{5}{4}\pi}^{\frac{7}{4}\pi} \frac{|\cot \theta|}{2\pi} d\theta = \frac{\log 4}{\pi} \\ \parallel \parallel \parallel \frac{1}{2\pi} \log 2 \parallel \parallel \parallel$$

### Problema 5

$X_1, \dots, X_n$  di legge unif. su  $[a-b, a+b]$  con  $f_{X_i} = \frac{1}{2b}$  con  $b > 0$   
 $L(a, b)(X_1, \dots, X_n) = \left(\frac{1}{2b}\right)^n \cdot \mathbb{1}_{\{a-b \in (-\infty, \min(X_i)\}, a+b \in (\max(X_i), +\infty)\}}$

$$\Rightarrow \begin{cases} b \geq a - \min(X_i) \\ b \geq \max(X_i) - a \end{cases} \Rightarrow \text{una delle due disuguaglianze} \\ \text{è sempre } > 0$$

$L(a, b)$  ha il sup per il più piccolo  $b$  t.c.  $\forall i \ X_i \in [a-b, a+b]$

quindi  $L(a, b)$  ha max per  $\bar{b} = \max_i (|X_i - a|)$

$$F_{\bar{b}}(K) = \mathbb{P}[\max_i |X_i - a| \leq K] = \prod \mathbb{P}[|X_i - a| \leq K] = \dots = \left(\frac{K}{b}\right)^n$$

$$d F_{\bar{b}}(K) = f_{\bar{b}}(K) = \frac{n K^{n-1}}{b^n} \quad \mathbb{E}[\bar{b}] = \int_0^b \frac{n K^n}{b^n} = \frac{b n}{n+1}$$

$$R_b(\bar{b}) = \mathbb{E}_b[(\bar{b} - b)^2] = \mathbb{E}_b[\bar{b}^2 + b^2 - 2b\bar{b}] = b^2 - \frac{2b^2 n}{n+1} +$$

$$+ \int_0^b \frac{K^{n+1} \cdot n}{b^n} dK = b^2 - \frac{2b^2 n}{n+1} + \frac{b^2 n}{n+2} = \frac{n(3b - 3b^2) + 2b}{(n+1)(n+2)}$$

$$\lim_{n \rightarrow \infty} \mathbb{P}[|\bar{b}_n - b| \geq \varepsilon] = \lim_{n \rightarrow \infty} \mathbb{P}[|\bar{b}_n - b|^2 \geq \varepsilon^2] \leq \frac{3(3b - 3b^2) + 2b}{\varepsilon^2(n+1)(n+2)} \rightarrow 0$$

## Problema 4

- $\mathbb{E}[\hat{\theta}_n] = \theta$
- $\mathbb{E}[\hat{\theta}_n^k] = \frac{n}{n+1} \theta$        $\mathbb{E}[\hat{\theta}_n^2] = \frac{n}{n+2} \theta^2$
- $R(\theta, \hat{\theta}_n) = \mathbb{E}[(\hat{\theta}_n - \theta)^2] = \frac{2\theta^2}{(n+1)(n+2)}$
- $\mathbb{P}[|\hat{\theta}_n - \theta| \geq \varepsilon] = \mathbb{P}[(\hat{\theta}_n - \theta)^2 \geq \varepsilon^2] \leq \frac{1}{\varepsilon^2} R_\theta(\hat{\theta}_n) = \frac{2\theta^2}{\varepsilon^2(n+1)(n+2)}$
- $L_\theta(x_1, \dots, x_n) = \theta^{-n} \mathbb{1}_{[\max(x_1, \dots, x_n), +\infty)}(\theta)$
- $\mathbb{E}[\bar{\theta}_n] = \theta$        $\mathbb{E}[X_1] = \frac{\theta}{2}$
- $\text{Var}(\theta, \bar{\theta}_n)$        $\text{Var}(X_1) = \frac{\theta^2}{12}$
- $R(\theta, \bar{\theta}_n) = \frac{\theta^2}{3n}$
- $\mathbb{P}[|\bar{\theta}_n - \theta| \geq \varepsilon] \leq \frac{1}{\varepsilon^2} \text{Var}(\bar{\theta}_n) = \frac{\theta^2}{3n\varepsilon^2}$

- $\frac{\theta^2}{3n} \geq \frac{2\theta^2}{(n+1)(n+2)}$        $n=1, 2 \rightarrow$  sono uguali  
" " " " " "  
 $R(\bar{\theta}_n, \theta)$        $R(\bar{\theta}_n, \theta)$        $n > 2 \rightarrow$  è preferibile lo stimatore di massima verosimiglianza

## Problema 2

$$f(x) = \frac{\theta}{x^2} \mathbb{1}_{(\theta, +\infty)}(x) \quad \theta > 0$$

$V = \min(x_1, \dots, x_n) \rightarrow$  perché? corrisponde allo stimatore di max. verosimiglianza

$$U = \frac{n-1}{n} V$$

$$\mathbb{E}[U] = \frac{n-1}{n} \mathbb{E}[V]$$

$$F_V(v) = \mathbb{P}[V \leq v]$$

$$= 1 - \mathbb{P}[V \geq v]$$

$$= 1 - (1 - F_{X_i}(x_i))^n$$

$$\mathbb{E}[U] = \theta$$

$$\lim_{n \rightarrow +\infty} \mathbb{P}[|U - \theta| \geq \varepsilon] = 0$$

$$P[|U - \theta| \geq \varepsilon] \leq \frac{\text{Var}(U)}{\varepsilon^2}$$

$$\text{Var}(U) = \theta^2 \left( \frac{(n-1)^2}{n(n-2)} - 1 \right)$$

$$\frac{\text{Var}(U)}{\varepsilon^2} \xrightarrow{n \rightarrow +\infty} 0$$

$$P[\theta \in D] \geq 1 - \alpha$$

↓

regione o intervallo di fiducia

lezione 20  
17-05-21  
(Romito)

## POPOLAZIONI GAUSSIANE

teorema Se  $X_1, \dots, X_n$  indipendenti e con legge  $N(m, \sigma^2)$

$$\bar{X} = \frac{1}{n} (X_1 + \dots + X_n) \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Allora

- $\bar{X}, S^2$  sono indipendenti
- $\bar{X}$  ha distribuzione  $N(m, \frac{\sigma^2}{n})$
- $(n-1) \frac{S^2}{\sigma^2}$  ha distribuzione  $\chi^2(n-1)$
- $\frac{\bar{X} - m}{S/\sqrt{n}}$  ha distribuzione di Student  $t(n-1)$

la somma di  
Gaussiane è una  
Gaussiana

## VARIANZA NOTA

Campione  $(X_1, \dots, X_n)$  con legge  $N(m, \sigma^2)$   $\sigma^2$  valore noto

- intervalli di fiducia

quantità pivotale:  $Z = \sqrt{n} \frac{\bar{X} - m}{\sigma}$  ha legge  $N(0, 1)$

regione di fiducia al livello  $1 - \alpha$ :

$$\{|Z| \leq q_{1-\frac{\alpha}{2}}\} \rightarrow m \in \left[ \bar{X} - \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}, \bar{X} + \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} \right]$$

nota bene:

- **Z-test**

- test bilatero

$$H_0: m = m_0$$

$$H_1: m \neq m_0$$

Una regione critica a livello  $\alpha$  è

$$\left\{ |Z| > q_{1-\frac{\alpha}{2}} \right\} \rightarrow \left\{ |\bar{X} - m| > \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} \right\}$$

Per il  $p$ -value, supponendo di aver raccolto i dati

$$x_1, \dots, x_n \quad \bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$$

$$p = P_{m_0} \left[ \sqrt{n} \frac{(\bar{X} - m)}{\sigma} > \sqrt{n} \frac{|\bar{x} - m|}{\sigma} \right]$$

• test unilatero: modello a rapporto di verosimiglianza crescente rispetto a  $\bar{X}$

$$\frac{L_{m_2}(X_1 \dots X_n)}{L_{m_1}(X_1 \dots X_n)} = \frac{e^{-\frac{1}{2\sigma^2} \sum (x_i - m_2)^2}}{e^{-\frac{1}{2\sigma^2} \sum (x_i - m_1)^2}} = e^{\frac{n}{2\sigma^2} (m_2 - m_1) \bar{X} - \frac{n}{2\sigma^2} (m_2^2 - m_1^2)}$$

dunque una regione critica al livello  $\alpha$

$$\left\{ \bar{X} > m_0 + \frac{\sigma}{\sqrt{n}} q_{1-\alpha} \right\}$$

### VARIANZA NON NOTA

$X_1, \dots, X_n$  con legge  $N(m, \sigma^2)$  parametri:  $(m, \sigma^2)$

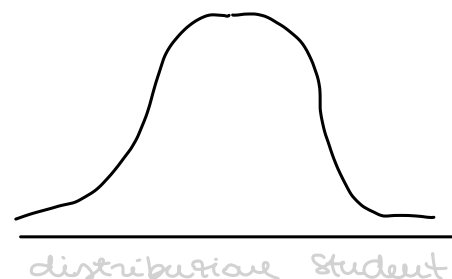
• intervalli di fiducia

quantità pivotale  $T = \sqrt{n} \frac{\bar{X} - m}{S}$  ha legge  $t(n-1)$

Detti  $t_{n,\alpha}$  i quantili della distribuzione di Student con  $n$  gradi di libertà

regione di fiducia  $\left\{ |T| \leq t_{n-1, 1-\frac{\alpha}{2}} \right\}$

densità Student  $c_n \left( 1 + \frac{x^2}{n} \right)^{-\frac{n+1}{2}}$



• test di Student

• test bilatero sulla media  $(m, \sigma^2) \in \mathbb{R} \times [0, \infty)$

$H_0: m = m_0, \sigma^2$  qualsiasi

$H_1: m \neq m_0, \sigma^2$  qualsiasi

una regione critica a livello  $\alpha$

$$T = \sqrt{n} \frac{\bar{X} - m_0}{S}$$

$$\{ |T| > t_{n-1, 1-\frac{\alpha}{2}} \}$$

• test unilatero

$H_0: m \leq m_0, \sigma^2$  qualsiasi

$H_1: m > m_0, \sigma^2$  qualsiasi

$T = \sqrt{n} \frac{\bar{X} - m_0}{S}$  Ognuna delle  $X_i$  del campione ha  $E[X_i] = m$

$T$  non ha distribuzione di Student  $\rightarrow$  non ha media nulla

$T$  ha distribuzione di Student "decentrata" di  $m - m_0$  che è a rapporto di verosimiglianza crescente rispetto a  $\bar{X}$

Una regione critica di livello  $\alpha$  è

$$\{ T > t_{n-1, 1-\alpha} \}$$

**VARIANZA**  $(X_1, \dots, X_n)$  con legge  $N(m, \sigma^2)$

• intervalli di fiducia per la media. Quantità pivotali:

•  $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - m)^2$  ha distribuzione  $\chi^2(n)$

(quantità pivotali se  $m$  ha valore noto)

•  $(n-1) \frac{S^2}{\sigma^2}$  ha legge  $\chi^2(n-1)$

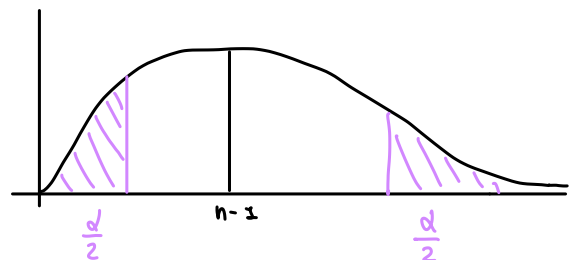
regione di fiducia

$$\{ \chi^2_{n-1, \frac{\alpha}{2}} \leq (n-1) \frac{S^2}{\sigma^2} \leq \chi^2_{n-1, 1-\frac{\alpha}{2}} \}$$

ovvero

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

$\rightarrow$  quale è meno significativo?



è meno significativo perché una varianza "piccola" non è preoccupante

meglio:

$$\left\{ (n-1) \frac{S^2}{\sigma^2} \geq \chi^2_{n-1, \alpha} \right\}$$

ovvero

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

• test bilatero

$$H_0: \sigma^2 = \sigma_0^2, \text{ m qualsiasi}$$

$$H_1: \sigma^2 \neq \sigma_0^2, \text{ m qualsiasi}$$

poco interessante

• test unilatero

$$H_0: \sigma^2 \leq \sigma_0^2, \text{ m qualsiasi}$$

$$H_1: \sigma^2 > \sigma_0^2, \text{ m qualsiasi}$$

densità di una  
somma di m  
Gaussiane indipend.  
di media 0 e  
varianza  $\sigma^2$

$$f(x) = \underbrace{C_m}_{\text{costante di normalizzazione}} \sigma^{-(m+1)} x^{\frac{m-3}{2}} e^{-\frac{x^2}{2\sigma^2}}$$

Voglio riconoscere che è un modello a rapporto di verosimiglianza crescente (rispetto a  $S^2$ )

$$(X_1 - m)^2 \dots (X_n - m)^2$$

Il quoziente tra le funzioni di verosimiglianza è

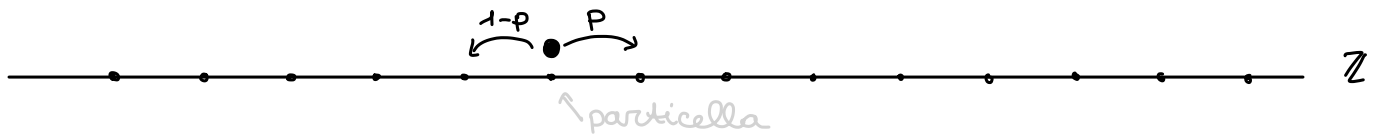
$$\frac{L_{\sigma_2^2}}{L_{\sigma_1^2}} = \left( \frac{\sigma_1}{\sigma_2} \right)^{(n+1)} e^{-\frac{n}{2} \left( \frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2} \right) S^2}$$

quindi una regione critica a livello  $\alpha$

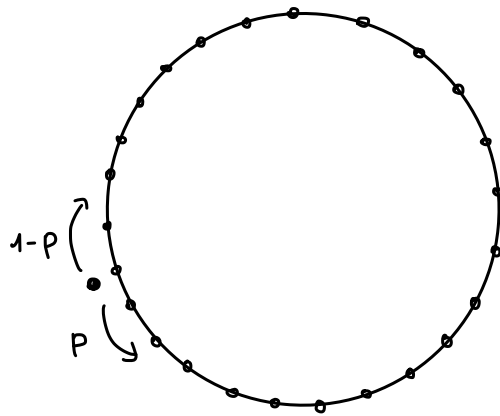
$$\left\{ (n-1) \frac{S^2}{\sigma_0^2} > \chi^2_{n-1, 1-\alpha} \right\}$$

finisce qui il programma di Statistica trattato nel corso

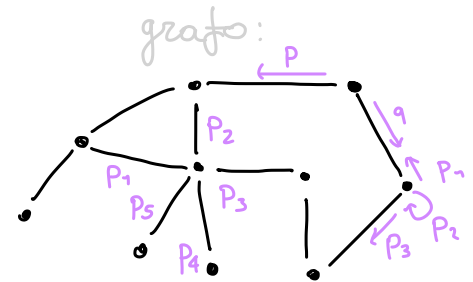
# CATENE DI MARKOV



descriviamo  
l'evoluzione  
elementare



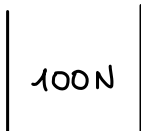
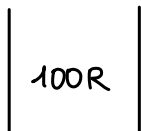
passeggiata  
aleatoria  
semplice



$$P_1 + P_2 + P_3 + P_4 + P_5 = 1$$

può essere orientato  
o non orientato  
Gli archi rappresentano  
le direzioni

modello di Ehrenfest



Il problema è parametrizzato  
dal numero di palle  
rosse del primo contenitore

riverscolamento di carte (40 carte)

sistema degli stati: possibili stati del nostro evento

spazio degli stati: permutazioni su  $\{1, \dots, 40\}$

evoluzione: a ogni passo estraiamo la prima carta del  
mazzo, e la reinsertiamo a caso all'interno del mazzo

catene di Markov: Una famiglia  $(X_n)_{n \geq 0}$  di v.a. a valori  
in un insieme  $S$  (detto spazio degli stati) è una catena di Markov  
se per ogni  $n \geq 1$  e ogni  $x, x_0, \dots, x_{n-1} \in S$   $P[X_n = x, x_0, \dots, x_{n-1}] > 0$   
allora

$$P[X_n = x | X_0 = x_0, \dots, X_{n-1} = x_{n-1}] = P[X_n = x | X_{n-1} = x_{n-1}]$$

"dimentica il passato"  
Proprietà  
di Markov

catene omogenee: Una catena di Markov è omogenea se per ogni  
 $n \geq 0, x, y \in S$ , con  $P[X_n = x] > 0$

$$P[X_{n+1} = y | X_n = x] = P[X_1 = y | X_0 = x]$$

matrice di transizione: per una catena di Markov omogenea

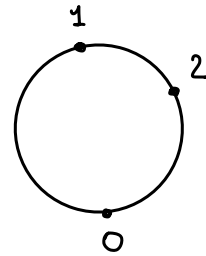
$$P_{xy} = P[X_1 = y | X_0 = x]$$

Proprietà (supponiamo  $S$  numerabile)

- $P_{x,y} \geq 0$  per ogni  $x, y \in S$
- $\sum_{y \in S} P_{xy} = 1$  (matrice stocastica)

esempio passeggiata aleatoria semplice su  $\{0, 1, 2\}$

	0	1	2	
0	0	$p$	$1-p$	$= 1$
1	$1-p$	0	$p$	$= 1$
2	$p$	$1-p$	0	$= 1$



proprietà equivalenti (alla proprietà di Markov)

1. per ogni  $m, n \geq 0$ , e ogni  $x, x_0, \dots, x_m$  t.c.  $P[X_0 = x_0, \dots, X_m = x_m] > 0$   
 $P[X_{n+m} = x | X_0 = x_0, \dots, X_m = x_m] = P[X_{m+n} = x | X_m = x_m]$
2. per ogni  $n$ , ogni  $K \leq n$ , e ogni  $n_1 < n_2 < \dots < n_K < n$ , ogni  $x, x_{n_1}, \dots, x_{n_K} \in S$  tali che  $P[X_{n_1} = x_{n_1}, \dots, X_{n_K} = x_{n_K}] > 0$   
 $P[X_{n+1} = x | X_{n_1} = x_{n_1}, \dots, X_{n_K} = x_{n_K}] = P[X_{n+1} = x | X_{n_K} = x_{n_K}]$

dim (caso) (1) formula di disintegrazione rispetto a ciò che è avvenuto nei tempi da  $m+1$  a  $n+m-1$

$$\begin{aligned} P[X_{n+1} = x | X_0 = x_0, \dots, X_m = x_m] &= \\ &= \sum P[X_{n+m} = x | X_0 = x_0 \dots X_{n+m-1} = X_{n+m-1}] \cdot \\ &\quad \cdot P[X_{n+1} = x_{n+1}, \dots, X_{n+m-1} = x_{n+m-1} | X_0 = x_0 \dots X_m = x_m] \end{aligned}$$

etc. etc.

□

CHIARIMENTO:

$X_1, \dots, X_n$  v.a. indipendenti, eguale distribuzione

$$\mu = E[X_1]$$

$$\sigma^2 = \text{Var}(X_1)$$

Per la CLT si ha

$$\mathbb{P} \left[ m + \frac{\sigma}{\sqrt{n}} a \leq \frac{X_1 + \dots + X_n}{n} \leq m + \frac{\sigma}{\sqrt{n}} b \right] \xrightarrow{n \rightarrow \infty} \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

Se ora supponiamo che  $X_1, \dots, X_n$  Gaussiane  $N(m, \sigma^2)$

$$\mathbb{P} \left[ m + \frac{\sigma}{\sqrt{n}} a \leq \frac{X_1 + \dots + X_n}{n} \leq m + \frac{\sigma}{\sqrt{n}} b \right] \stackrel{(\circledast)}{=} \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

||

$$\mathbb{P} \left[ a \leq \frac{X_1 + \dots + X_n - m}{\frac{n}{\sigma}} \leq b \right] = \Phi(b) - \Phi(a)$$

Esercizio 11  
18-05-2021  
(Di Gesù)

## Problema 1

$X_1, \dots, X_n$  campione con legge uniforme su  $[a-b, a+b]$

$$M = \max_i (X_i) \quad \hat{a} = \frac{M+m}{2} \quad \hat{b} = \frac{M-m}{2} \quad \text{STIMATORI DI MASSIMA VEROSIMILIANZA}$$

$$m = \min_i (X_i)$$

$$\mathbb{E}[\hat{a}] = \frac{\mathbb{E}[M] + \mathbb{E}[m]}{2} \quad \mathbb{E}[\hat{b}] = \frac{\mathbb{E}[M] - \mathbb{E}[m]}{2}$$

$$M, F_n(t) = \mathbb{P}[X_i \leq t \quad \forall i] = \prod_{i=1}^n \mathbb{P}[X_i \leq t] = F^n(t) \quad t \in [a-b, a+b]$$

$$F(t) = \frac{(t-a+b)}{2b} \quad t \in [a-b, a+b]$$

$$f_n(t) = [F_n(t)]' = n f(t) \cdot F(t)^{n-1} = n \left( \frac{t-a+b}{2b} \right)^{n-1} \cdot \frac{1}{2b} \mathbb{1}_{[a-b, a+b]}$$

$$\begin{aligned} \mathbb{E}[M] &= \int_{\mathbb{R}} t \cdot f_n(t) dt = \int_{a-b}^{a+b} \frac{nt}{2b} [t-a+b]^{n-1} dt = \\ &= t \left( \frac{t-a+b}{2b} \right)^n \Big|_{a-b}^{a+b} - \int_{a-b}^{a+b} \left( \frac{t-a+b}{2b} \right)^n dt = (a+b) - \left[ \frac{1}{n+1} \frac{(t-a+b)^{n+1}}{(2b)^n} \Big|_{a-b}^{a+b} \right] = (a+b) - \frac{2b}{n+1} \end{aligned}$$

$$\mathbb{E}[m] = \dots = (a-b) + \frac{2b}{n+1}$$

$$\mathbb{E}[\hat{a}] = \left( (a+b) - \frac{2b}{n+1} + (a-b) + \frac{2b}{n+1} \right) \frac{1}{2} = \frac{2a}{2} = a \rightarrow \text{corretto}$$

$$\mathbb{E}[\hat{b}] = \left( (a+b) - \frac{2b}{n+1} - (a-b) - \frac{2b}{n+1} \right) \frac{1}{2} = b - \frac{2b}{n+1} \rightarrow \text{asintoticamente corretto}$$

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} \xrightarrow{n} \begin{pmatrix} a \\ b \end{pmatrix}$$

$$\bullet \mathbb{P} \left[ \left\| \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} - \begin{pmatrix} a \\ b \end{pmatrix} \right\|_2 \geq \varepsilon \right] \xrightarrow{n} 0$$

$$\mathbb{P} \left[ \left\| \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} - \begin{pmatrix} a \\ b \end{pmatrix} \right\|_2^2 \geq \varepsilon^2 \right] = \mathbb{P} \left[ (\hat{a} - a)^2 + (\hat{b} - b)^2 \geq \varepsilon^2 \right]$$

$$R(\hat{a}, a) = \mathbb{E}[(\hat{a} - a)^2] = \mathbb{E}[\hat{a}^2] - 2a \mathbb{E}[\hat{a}] + a^2$$

$$R(\hat{b}, b) = \mathbb{E}[\hat{b}^2] - 2b \mathbb{E}[\hat{b}] + b^2$$

$$\mathbb{E}[\hat{a}^2] = \frac{\mathbb{E}[M^2] + \mathbb{E}[m^2] + 2\mathbb{E}[mM]}{4}$$

$$\mathbb{E}[\hat{b}^2] = \frac{\mathbb{E}[M^2] + \mathbb{E}[m^2] - 2\mathbb{E}[mM]}{4}$$

*M e m NON sono indipendenti*

$$\mathbb{E}[m^2] = \dots = (a-b)^2 + \frac{4b(a-b)}{n+1} + \frac{8b^2}{(n+1)(n+2)}$$

$$\mathbb{E}[M^2] = \dots = (a+b)^2 - \frac{4(a+b)b}{n+1} + \frac{8b^2}{(n+1)(n+2)}$$

$$\mathbb{E}[m^2] + \mathbb{E}[M^2] = 2a^2 + 2b^2 + \frac{16b^2}{(n+1)(n+2)} - \frac{8b^2}{n+1} \longrightarrow 0 \text{ è consistente}$$

### Problema 3

$X_1, \dots, X_n$  campione con densità esponenziale di parametro  $\lambda > 0$

• Trovare la stima di massima verosimiglianza  $\hat{\lambda}_n$  di  $\lambda$ :

$$f_x(x) = \lambda e^{-\lambda x} \mathbb{1}_{\{x \in [0, +\infty)\}} \quad \text{densità di } X_1, \dots, X_n$$

$$L_\lambda(X_1, \dots, X_n) = \lambda^n e^{-\lambda \sum_{i=1}^n X_i} \mathbb{1}_{\{x \in [0, +\infty)\}}$$

$$\log(L_\lambda(X_1, \dots, X_n)) = n \log(\lambda) - \lambda \sum_{i=1}^n X_i \Rightarrow \tilde{\lambda}_n = \frac{n}{X_1 + \dots + X_n}$$

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

Cerchiamo  $n \mathbb{E} [1/\hat{X}]$

$$\begin{aligned} n \mathbb{E} [1/\hat{X}] &= n \int_0^\infty \frac{1}{s} f_{\hat{X}}(s) ds = \frac{n \lambda^n}{\Gamma(n)} \int_0^\infty s^{n-2} e^{-\lambda s} ds = \frac{1}{s} e^{-\lambda s} \downarrow \substack{t=\lambda s \\ dt=ds} \\ &= \frac{n \lambda^n}{\Gamma(n)} \int_0^\infty \frac{t^{n-2}}{\lambda^{n-2}} e^{-t} \frac{dt}{\lambda} = \frac{n \lambda}{\Gamma(n)} \Gamma(n-1) = \frac{n \lambda}{(n-1)} \quad \text{per } n > 1 \end{aligned}$$

È corretto asintoticamente ma non per  $n$  fissato fissato  
Si poteva vedere senza calcoli?

$$\mathbb{E} [1/\hat{X}] > \frac{1}{\mathbb{E}[\hat{X}]} = \lambda \quad \text{disuguaglianza di Jensen}$$

$\mathcal{H}_0: \lambda \leq 1$  e  $\mathcal{H}_1: \lambda > 1$  al livello  $\alpha$

$$\{\hat{\lambda}_n > \delta\}$$

$$\mathbb{P}_\lambda [\hat{\lambda}_n > \delta] \leq \alpha \quad \forall \lambda \leq 1$$

$$\mathbb{P}_\lambda [\hat{\lambda}_n > \delta] = \mathbb{P}_\lambda \left[ \frac{1}{\bar{X}_n} > \delta \right] = \mathbb{P} \left[ \bar{X}_n < \frac{1}{\delta} \right]$$

$$\mathbb{P}_\lambda \left[ \bar{X}_n < \frac{1}{\delta} \right] = F_{n, \lambda} \left( \frac{n}{\delta} \right) = \alpha$$

$$\frac{n}{\delta} = F_{n, \lambda}^{-1}(\alpha)$$

$$\delta = \frac{n}{F_{n, \lambda}^{-1}(\alpha)}$$

$$\lambda_1 < \lambda_2$$

$\frac{L_{\lambda_2}}{L_{\lambda_1}}$  è strettamente crescente

$$\frac{L_{\lambda_2}}{L_{\lambda_1}} = \frac{\lambda_2^n}{\lambda_1^n} e^{(\sum_{i=1}^n x_i)(-\lambda_2 + \lambda_1)} = \left( \frac{\lambda_2}{\lambda_1} \right)^n e^{-\bar{X}_n(-\lambda_2 + \lambda_1)} = \left( \frac{\lambda_2}{\lambda_1} \right)^n e^{\frac{n}{\bar{X}_n}(-\lambda_2 + \lambda_1)}$$

È strettamente crescente rispetto a  $\hat{\lambda}_n$

Il test che abbiamo scelto è di potenza massima

Usando ie TLC ottengo:  $\delta \geq \frac{\sqrt{n}}{\sqrt{n} + q_d}$

$$P_2[|\lambda_n - 2| > \delta] = P_2[|\hat{\lambda}_n - 2|^2 > \delta^2] \leq \frac{E[(\hat{\lambda}_n - 2)^2]}{\delta^2} =$$

$$= \frac{n^2 4}{(n-1)(n-2)} - \frac{8n}{n-1} + 4 = \frac{4(n+2)}{(n-1)(n-2)} \cdot \frac{1}{\delta^2} \leq \alpha$$

$$\delta \geq \sqrt{\frac{4(n+2)}{(n-1)(n-2) \alpha}}$$

## Problema 2

$$\left. \begin{aligned} A_n &= \min_k \{X_k\} \\ \theta_n &= \bar{X}_n - A_n \end{aligned} \right\} \text{ si trova con il calcolo differenziale}$$

$$E[A_n] = \frac{\theta}{n} + a$$

$$\text{Var}(A_n) = \frac{\theta^2}{n^2}$$

$A_n$  è asintoticamente corretto

$\theta_n$  ie discorso è simile

$$E[\theta_n] = \frac{n-1}{n} \theta$$

Per la consistenza di  $\hat{\theta}$ :

Variante 1:  $| \text{Var}(\bar{X}_n) + \text{Var}(A_n) - 2 \text{Cov}(\bar{X}_n, A_n) | \leq$

$$\leq \underbrace{\text{Var}(\bar{X}_n)}_{\downarrow 0} + \underbrace{\text{Var}(A_n)}_{\downarrow 0} + 2 \underbrace{\sqrt{\text{Var}(\bar{X}_n)}}_{\downarrow 0} \underbrace{\sqrt{\text{Var}(A_n)}}_{\downarrow 0}$$

Variante 2:  $\{ |\hat{\theta}_n - E[\hat{\theta}_n]| \geq \varepsilon \} =$

$$= \{ |\underbrace{\bar{X}_n - E[\bar{X}_n] - (A_n - E[A_n])}_{\text{...}}| \geq \varepsilon \}$$

$$\leq |\bar{X}_n - E[\bar{X}_n]| + |A_n - E[A_n]|$$

$$\subset \{ |\bar{X}_n - E[\bar{X}_n]| \geq \frac{\varepsilon}{2} \} \cup \{ |A_n - E[A_n]| \geq \frac{\varepsilon}{2} \}$$

$S$  spazio degli stati

$(X_n)_{n \geq 0}$  a valori in  $S$  tali che

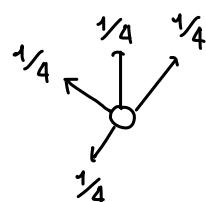
$$P[X_{n+1} = x \mid X_0 = x_0, X_1 = x_1, \dots, X_n = x_n] = P[X_{n+1} = x \mid X_n = x_n]$$

proprietà di  
Markov

per ogni  $n$   $x, x_0, \dots, x_n$  t.c.  $P[X_0 = x_0, \dots, X_n = x_n] > 0$

esempi:

- passeggiata aleatoria semplice su un grafo
- diffusione di Ehrenfest
- rimiscelamento di carte
- $\xi_1, \xi_2, \dots$  indipendenti e con la stessa distribuzione  
 $X_0$  dato indipendente da  $\xi_1, \xi_2, \dots$



$$f: \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$X_{n+1} = f(X_n, \xi_{n+1}) \quad n \geq 1$$

allora questo è una catena di Markov

Caso particolare: passeggiata aleatoria semplice su  $\mathbb{Z}$

$$\text{se } f(x, y) = x + y$$

$$\xi_i = \begin{cases} 1 & p \\ -1 & q \end{cases}$$

Supponiamo d'ora in poi  $(X_n)_{n \geq 0}$  sia una catena di Markov omogenea

$$P = (P_{ij})_{i, j \in S} \quad P_{ij} = P[X_{n+1} = j \mid X_n = i] \quad (\text{quale sia } n)$$

- $P_{ij} \geq 0$  per ogni  $i, j$
- $\sum_j P_{ij} = 1$  (matrice stocastica)

$P_{ij}$  probabilità di passare in un passo dallo stato  $i$  allo stato  $j$

Consideriamo la distribuzione dopo  $n$  passi

$$(P^{m, m+n})_{ij} = P[X_{m+n} = j \mid X_m = i]$$

In particolare  $P = P^{0,1} = P^{n, n+1}$

Equazione di Chapman-Kolmogorov

Se  $m \geq 0$ ,  $n, r \geq 1$

$$P^{m, m+n+r} = P^{m, m+n} P^{m+n, m+n+r}$$

e in particolare  $P^{m, m+n} = (P)^n$

dim

$$(P^{m, m+n+r})_{ij} = P[X_{m+n+r} = j \mid X_m = i]$$

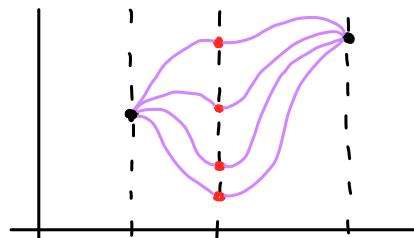
$$= \sum_{\ell \in S} P[X_{m+n+r} = j, X_{m+n} = \ell \mid X_m = i]$$

$$= \sum_{\ell \in S} P[X_{m+n+r} = j \mid X_{m+n} = \ell, X_m = i] \cdot P[X_{m+n} = \ell \mid X_m = i]$$

$$= \sum_{\ell \in S} P[X_{m+n+r} = j \mid X_{m+n} = \ell] \cdot P[X_{m+n} = \ell \mid X_m = i]$$

$$= \sum_{\ell \in S} (P^{m+n, m+n+r})_{\ell j} (P^{m, m+n})_{i \ell}$$

$$= [(P^{m, m+n}) \cdot (P^{m+n, m+n+r})]_{ij}$$



osservazione la legge di ogni v.a. della catena è univocamente determinata dalla matrice di transizione e dalla distribuzione iniziale

$\mu_0$  la densità discreta di  $X_0$ :  $\mu_0(i) = P[X_0 = i]$   $i \in S$

la distribuzione di  $X_1$

$$P[X_1 = j] = \sum_{i \in S} P[X_1 = j \mid X_0 = i] P[X_0 = i]$$

$$= \sum_{i \in S} \mu_0(i) P_{ij} = (\mu_0 P)_j$$

Iterando otteniamo  $P[X_n = j] = (\mu_0 P^n)_j$ , il passo di induzione

$$\begin{aligned}
 P[X_n = j] &= \sum_{i \in S} P[X_n = j \mid X_{n-1} = i] P[X_{n-1} = i] \\
 &= \sum_{i \in S} P[X_{n-1} = i] P_{ij} \\
 &= \sum_{i \in S} (\mu_0 P^{n-1})_i P_{ij} = (\mu_0 P^n)_j
 \end{aligned}$$

Problema: se la legge di  $X_n$  converge, al tendere di  $n$  a infinito, a qualcosa

Oss  $|S| = n$

$$S = \{x_1, \dots, x_N\} \leftrightarrow \{1, 2, \dots, N\}$$

$\mu_0$  densità discreta su  $\{1, \dots, N\}$

$$\mu_0(1), \dots, \mu_0(N)$$

$P$  è una matrice  $N \times N$

$\mu_0 \cdot P$  = prodotto tra il vettore  $\mu_0$  e la matrice  $P$

$$\begin{array}{c}
 (\mu_0 \cdot P)_i \\
 \underbrace{(1 \times N \quad N \times N)}_{1 \times N}
 \end{array}$$

La legge di  $X_n$  è  $\mu_0 \cdot P^n$

$\mu_0$ : densità discreta di  $X_0$

$P$ : matrice di transizione

Dire che la legge di  $X_n$  "converge", significa dire che la successione di vettori  $(\mu_0 \cdot P^n)_{n \geq 0}$  converge a qualcosa

Proviamo a scoprire com'è fatto il limite

$$\begin{array}{ccc}
 \mu_0 \cdot P^n & \searrow & \textcircled{?} \\
 \mu_0 \cdot P^{n+1} & \nearrow & \textcircled{?}
 \end{array}$$

in quanto sottosuccessione

$$(\mu_0 \cdot P^n) \longrightarrow \textcircled{?} \cdot P$$

Se c'è un limite  $\pi$ , questo limite deve necessariamente soddisfare la relazione  $\pi = \pi \cdot P$

osservazione:  $\pi$  deve essere un autovettore (sinistro)  
relativo all'autovalore 1

1 è certamente un autovalore

$$P \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

distribuzione stazionaria Una legge  $\pi = (\pi_i)_{i \in S}$  è una distribuzione stazionaria (invariante) per la matrice di transizione  $P$  se

- $\pi_i \geq 0$  per ogni  $i \in S$
  - $\sum_{i \in S} \pi_i = 1$
  - $\pi = \pi \cdot P$
- }  $\pi$  è una legge

osservazione Se  $\pi$  è stazionaria e  $X_0$  ha distribuzione  $\pi$   
allora  $X_n$  ha legge  $\pi$  per ogni  $n$

$$\pi P^n = \pi$$

### TEOREMA

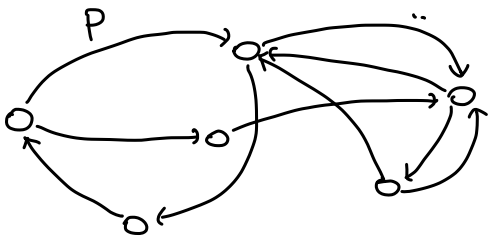
Data una catena di Markov <sup>omogenea</sup>  $(X_n)_{n \geq 0}$  con matrice di transizione  $P$

- irriducibile
- aperiodica

allora

- esiste un'unica distribuzione invariante  $\pi$
- per ogni  $i \in S$   $\lim_n P[X_n = i] = \pi_i$

osservazione Una catena di Markov omogenea su un insieme finito  $S$  può essere codificata attraverso un grafico orientato, i cui vertici sono gli elementi di  $S$  e gli archi  $\{(i, j) : P_{ij} > 0\}$



## Classificazione degli stati

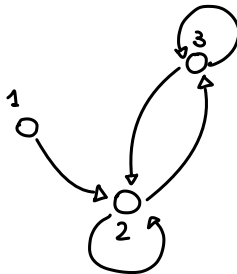
- uno stato  $i \in S$  è ricorrente / persistente se

$$P[X_n = i \text{ per un qualche } n \geq 1 | X_0 = i] = 1$$

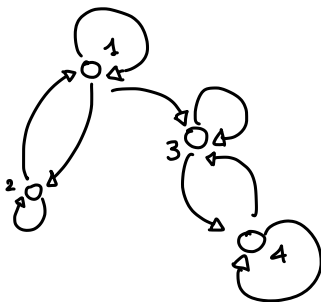
- uno stato  $i \in S$  è transiente se

$$P[X_n = i \text{ per un qualche } n \geq 1 | X_0 = i] < 1$$

## Esempio



1 : transiente  
2 : ricorrente  
3 : ricorrente



1 : transiente  
2 : "  
3 : ricorrente  
4 : "

Dato  $i$   $N_i = \text{cardinalità } \{n : X_n = i\}$

$$P[N_i = \infty | X_0 = i] = \begin{cases} 1 & i \text{ ricorrente} \\ 0 & i \text{ transiente} \end{cases}$$

teorema Se  $S$  è finito, allora esiste sempre uno stato ricorrente

## CLASSIFICAZIONE DI CATENE

comunicazione tra stati: lo stato "comunica" con  $j$  se esiste  $m \geq 0$  tale che  $P[X_m = j | X_0 = i] = (P^m)_{ij} > 0$

stati intercomunicanti:  $i, j$  sono mutuamente comunicanti

se  $i \rightarrow j$  e  $j \rightarrow i$  (ovvero  $i \leftrightarrow j$ )

osservazione:  $\leftrightarrow$  è una relazione d'equivalenza

$$i \leftrightarrow i \quad P[X_0 = i | X_0 = i] > 0$$

simmetria: ovvia

transitività:  $i \leftrightarrow j$  e  $j \leftrightarrow k$  allora  $i \leftrightarrow k$

infatti è sufficiente  $i \rightarrow j$  e  $j \rightarrow k$  allora  $i \rightarrow k$

$$(P^m)_{ij} > 0 \quad (P^{m'})_{jk} > 0$$

$$(P^{m+m'})_{ik} = \sum_{l \in S} (P^m)_{il} (P^{m'})_{lk} \\ \geq (P^m)_{ij} (P^{m'})_{jk} > 0$$

classe chiusa  $C \subset S$  è chiuso se

$$P_{ij} = 0 \text{ per ogni } i \in C \text{ e } j \notin C$$

classe irriducibile  $C \subset S$  se  $i \leftrightarrow j$  per ogni  $i, j \in C$

teorema di decomposizione

$$S = T \cup C_1 \cup \dots \cup C_k$$

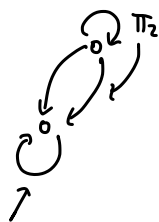
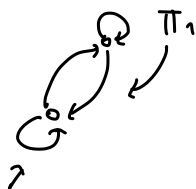
dove

$T$  = stati transienti

$C_1, \dots, C_k$  sono classi chiuse irriducibili

osservazione:  $C$  classe di equivalenza per  $\leftrightarrow$

esempio

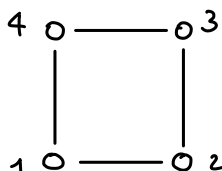
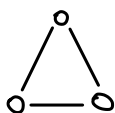


$$\lambda \pi_1 + (1-\lambda) \pi_2$$



Il solo fatto che il sistema è irriducibile non basta

grafi di  
catene  
irriducibili



$$P[X_n = 2 | X_0 = 1] = \begin{cases} > 0 & \text{se } n \text{ disp.} \\ = 0 & \text{se } n \text{ pari} \end{cases}$$