venerdì 4 novembre 2022 16:37

Nozioni di base di probabilità

Esperimento aleatorio: fenomeno (fisico, biologico, sociale, ...) il cui esito non è determinabile con certezza a priori.

Spazio campionario Ω = insieme di tutti i possibili esiti dell'esperimento aleatorio in esame (Ω insieme $\neq \emptyset$)

Nella teoria della probabilità, traduciamo affermazioni sui possibili esiti in sottoinsiemi di Ω (eventi)

l'esito di un esperimento soddisfa una proprietà $p \leftrightarrow \{\omega \in \Omega \mid p(\omega)$ è vera$\}$

Esempi:

- lancio di una moneta: $\Omega = \{T, C\}$
  esce testa $\leftrightarrow \{T\}$
- lancio di un dado: $\Omega = \{1, 2, \dots 6\}$
  esce 2 $\leftrightarrow \{2\}$
  esce un numero pari $\leftrightarrow \{2, 4, 6\}$
- due lanci di moneta (conta l'ordine dei lanci): $\Omega = \{(T,T), (T,C), (C,T), (C,C)\} = \{T,C\} \times \{T,C\}$
  esce T al 1° e C al 2° $\leftrightarrow \{(T,C)\}$
  esce T al 1° $\leftrightarrow \{T\} \times \{T,C\}$
- numero di figli di una data persona: $\Omega = \mathbb{N} = \{0, 1, 2, \dots\}$
  almeno 3 figli $\leftrightarrow \{n \in \mathbb{N} \mid n \geq 3\}$
- temperatura (in kelvin) di un dato gas: $\Omega = [0, +\infty)$
  temperatura tra 400 e 500 °K (estremi inclusi) $\leftrightarrow [400, 500]$

Corrispondenza tra operazioni logiche e operazioni insiemistiche: (A, B, A_i eventi)

si verificano A e B $\leftrightarrow$ $A \cap B$ (si verificano tutti gli $A_i \leftrightarrow \bigcap_i A_i$)

si verifica A o B $\leftrightarrow$ $A \cup B$ (si verifica almeno uno tra gli $A_i \leftrightarrow \bigcup_i A_i$)

non si verifica A $\leftrightarrow$ $A^c = \Omega \setminus A$

si verifica A ma non B $\leftrightarrow$ $A \setminus B$

si verifica o A o B $\leftrightarrow$ $A \Delta B = A \setminus B \cup B \setminus A$

non si verifica nulla $\leftrightarrow \emptyset$, si verifica qualcosa $\leftrightarrow \Omega$

Corrispondenza tra relazioni logiche e relazioni insiemistiche

$A \subseteq B \leftrightarrow$ se accade A, allora necessariamente accade B

$A \cap B = \emptyset \leftrightarrow$ non possono accadere sia A sia B

esempi: $\Omega = \{1, 2, \dots 6\}$

$A = \{1\} \subseteq B = \{$dispari$\}$

$A = \{$pari$\}$, $B = \{$dispari$\}$

Una probabilità P associa a ogni evento A di Ω un numero $P(A)$, che esprime il il grado di fiducia di A (quanto probabile si ritiene A)

$P(A) \in [0,1]$, più $P(A)$ è vicino a $1/0$, più/meno A è probabile.

Per motivi tecnici, non sempre è possibile definire "coerentemente" una probabilità su ogni sottoinsieme di Ω. Per questo, si introduce una classe di sottoinsiemi "ammissibili".

Def: Una σ-algebra $\mathcal{F}$ su Ω è un insieme $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ di sottoinsiemi di Ω t.c.

- $\Omega \in \mathcal{F}$
- se $A \in \mathcal{F}$, allora $A^c \in \mathcal{F}$
- se $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}$, allora $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$.

Evento: sottoinsieme di Ω in $\mathcal{F}$

(evento elementare: evento del tipo $\{\omega\}$, $\omega \in \Omega$)

$(\Omega, \mathcal{F})$ spazio misurabile

Definizione assiomatica di probabilità (Kolmogorov):

$\Omega \neq \emptyset$, $\mathcal{F}$ σ-algebra su Ω.

Una probabilità P su $(\Omega, \mathcal{F})$ è una funzione $P: \mathcal{F} \to \mathbb{R}$ t.c.

i) $0 \le P(A) \le 1 \quad \forall A \in \mathcal{F}$

ii) $P(\Omega) = 1$

iii) $\forall (A_n)_{n\in\mathbb{N}} \in \mathcal{F}$ eventi a due a due disgiunti (cioè $A_i \cap A_j = \emptyset \quad \forall i \neq j$), vale

$$P\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} P(A_n) \qquad (\sigma\text{-additività})$$

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

(Oss: non sempre una prob. su $\mathcal{F}$ si può estendere a una probabilità su tutto $\mathcal{P}(\Omega)$; per questo si devono introdurre le σ-algebre.)

Oss: • Se P è probabilità, allora soddisfa $P(\emptyset) = 0$: infatti

$$1 = P(\Omega) = P(\Omega \cup \emptyset \cup \emptyset \dots) = P(\Omega) + \sum_{n=2}^{\infty} P(\emptyset), \text{ quindi deve essere } P(\emptyset) = 0$$

• Inoltre, P soddisfa:

iii') $\forall A_1, \dots A_N \in \mathcal{F}$ eventi a due a due disgiunti, vale

$$P\Big(\bigcup_{n=1}^{N} A_n\Big) = \sum_{n=1}^{N} P(A_n) \qquad (\text{additività})$$

come si ottiene da (iii) prendendo $A_{N+1} = A_{N+2} = \dots = \emptyset$

Diciamo che P è una prob. finitamente additiva se valgono (i)+(ii)+(iii')

· Inoltre, se Ω è finito, la def. di probabilità è equivalente a prob. finitamente additiva.

Significato della definizione (cioè perché è ragionevole):

· (i), (ii): la prob di un evento è in $[0,1]$, la prob che accada un qualche esito (Ω) è 1

· (iii') (additività): se A e B sono incompatibili, la prob che accada A o B è la

la somma delle prob. di A e di B

es: numero figli: $\Omega = \mathbb{N}$, $A = \{0,1\}$ (al più un figlio), $B = \{2\}$ (2 figli)

$P\{\text{al più due figli}\} = P\{\text{al più un figlio}\} + P\{\text{due figli}\}$

· (iii) (σ-additività) esempio numero figli: $\Omega = \mathbb{N}$: $\{n° \text{ pari di figli}\} = \bigcup_{n \in \mathbb{N}} \{2n\}$

$P[n° \text{ pari di figli}] = \sum_{n \in \mathbb{N}} P\{2n\}$

Esempio fondamentale: modello uniforme (o equiprobabile):

$\Omega \neq \emptyset$ finito, $\mathcal{F} = \mathcal{P}(\Omega)$

P uniforme: $\forall A \in \mathcal{P}(\Omega)$

$P(A) = \frac{\# A}{\# \Omega}$ (con $\# A$ = cardinalità di A = numero di elementi di A)

" $= \frac{\# \text{casi favorevoli a } A}{\# \text{casi possibili}}$ "

in particolare, $\forall \omega \in \Omega$, $P(\{\omega\}) = \frac{1}{\# \Omega}$

Oss: P uniforme verifica la def. assiomatica di probabilità

Esempi:

· lancio di una moneta equilibrata

($\Omega = \{T, C\}$ $\mathcal{F} = \mathcal{P}(\Omega)$ P uniforme $P\{T\} = P\{C\} = \frac{1}{2}$)

· lancio di un dado equilibrato

($\Omega = \{1, 2 .. 6\}$ $\mathcal{F} = \mathcal{P}(\Omega)$ P uniforme $P\{\omega\} = \frac{1}{6}$ $\forall \omega \in \Omega$)

· n lanci di moneta equilibrata

$\Omega = \{T, C\}^n$ $\mathcal{F} = \mathcal{P}(\Omega)$ P uniforme: "ogni sequenza di T e C ha la stessa probabilità di uscire"

$P\{\omega\} = \frac{1}{2^n}$

· estrazione (da un'urna, una popolazione, ...)

$\Omega = \{\text{oggetti nell'urna/popolazione/..}\}$ $\mathcal{F} = \mathcal{P}(\Omega)$ P uniforme

Caso Ω finito o numerabile ($\Omega = \{\omega_1, \omega_2 ..\}$), $\mathcal{F} = \mathcal{P}(\Omega)$:

In questo caso, P è univocamente determinata dalla funzione

$\Omega \ni \omega_i \mapsto p(\omega_i) := P\{\omega_i\}$ (densità discreta)

Infatti, per additività, $\forall A \in \mathcal{P}(\Omega)$

$P(A) = \sum_{\omega_i \in A} P\{\omega_i\}$

Esempi:

- 4 lanci di moneta equilibrata, n° teste: $\Omega = \{0,1,2,3,4\}$, $\mathcal{F} = \mathcal{P}(\Omega)$

k	{esattamente k teste}	P{esattamente k teste} =
0	{CCCC}	1/16
1	{TCCC, CTCC, CCTC, CCCT}	4/16
2	{TTCC, TCTC, TCCT, CTTC, CTCT, CCTT}	6/16
3	{TTTC, TTCT, TCTT, CTTT}	4/16
4	{TTTT}	1/16

- moneta truccata, prob di testa = 2/3

  $\Omega = \{T, C\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, P t.c.

  $P\{T\} = 2/3$, $P\{C\} = 1 - P\{T\} = 1/3$

Esempio non numerabile: punto scelto a caso in $[0,1]$

$\Omega = [0,1]$ P = "uniforme", cioè $P(A)$ = "lunghezza di A"

$\mathcal{F} = ?$ come vedremo, non possiamo prendere $\mathcal{F} = \mathcal{P}(\Omega)$

Nota: come vedremo, $P\{x\} = 0$ $\forall x \in [0,1]$, perciò P non è determinata da $P\{x\}$

Proprietà di una σ-algebra $\mathcal{F}$:

- $\phi \in \mathcal{F}$: $\phi = \Omega^c$
- $(A_n)_n \subseteq \mathcal{F} \Rightarrow \bigcap_n A_n \in \mathcal{F}$: $\bigcap_n A_n = \left(\bigcup_n A_n^c\right)^c$, $A_n \in \mathcal{F}\ \forall n \Rightarrow A_n^c \in \mathcal{F}\ \forall n \Rightarrow \bigcup_n A_n^c \in \mathcal{F} \Rightarrow \left(\bigcup_n A_n^c\right)^c \in \mathcal{F}$
- $A, B \in \mathcal{F} \Rightarrow A \cup B,\ A \cap B,\ A \setminus B \in \mathcal{F}$

  $A \cup B = A \cup B \cup \phi \cup \phi \cup \ldots \in \mathcal{F}$

  $A \cap B = A \cap B \cap \Omega \cap \Omega \cap \ldots \in \mathcal{F}$

  $A \setminus B = A \cap B^c \in \mathcal{F}$

Proprietà di una probabilità P:

a) $P(\phi) = 0$

b) $\forall A \in \mathcal{F}$, $P(A^c) = 1 - P(A)$

c) $\forall A, B \in \mathcal{F}$, se $A \subseteq B$ allora $P(A) \leq P(B)$ e $P(B \setminus A) = P(B) - P(A)$

d) $\forall A, B \in \mathcal{F}$, $P(B \setminus A) = P(B) - P(A \cap B)$

e) $\forall A, B \in \mathcal{F}$, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

f) $\forall A_1, \ldots A_n \in \mathcal{F}$,

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n+1} P(A_1 \cap A_2 \cap \ldots \cap A_n)$$

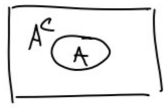
(formula di inclusione - esclusione)

Dimostrazione:

a) Già visto

b) $\Omega = A \dot{\cup} A^c$ ($\dot{\cup}$ = unione disgiunta), quindi

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

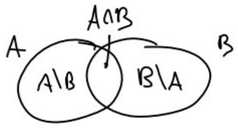


c) se $A \subseteq B$, allora $B = A \dot{\cup} (B \setminus A)$, quindi

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

d) Segue da (c) rimpiazzando A con $A \cap B$

e) $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$ a due a due disgiunti



$$\begin{aligned} P(A \cup B) &= P(A \setminus B) + P(B \setminus A) + P(A \cap B) \\ &= P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

f) Esercizio: per induzione su (n)

Idea intuitiva per f: usare modello unif

Oss (esercizio): • Se P soddisfa (i'), (ii), (iii) delle def di probabilità, con
(i'): $P(A) \geq 0 \quad \forall A \in \mathcal{F}$,
allora $P(A) \leq 1 \ \forall A \in \mathcal{F}$ e quindi P è misura di probabilità
(segue da $P(\Omega) = 1$ e monotonia)

• Se Ω è finito e P soddisfa (i), (ii), (iii') con
(iii'): $\forall A, B \in \mathcal{F}$ disgiunti, $P(A \cup B) = P(A) + P(B)$ (additività)
allora P soddisfa (iii) e quindi è misura di probabilità
(poiché $\mathcal{F}$ è finito)

Oss: La probabilità si riduce alla combinatoria? non proprio:

- esistono esempi di modelli non riconducibili a uniforme:
  - moneta truccata, "modelli continui"
  - numero di teste in n lanci di moneta "per n grande" (risultati asintotici)
- alcuni concetti e metodi (prob condizionata, valore atteso di v.a., ...), anche se in modelli uniformi si possono scrivere in termini combinatorici, hanno un loro valore e una loro utilità oltre la combinatoria

Lemma: $(\Omega, \mathcal{F}, P)$ sp di probabilità, A_n, $n \in \mathbb{N}$, A eventi.

a) Se $A_n \nearrow A$, cioè $A_n \subseteq A_{n+1}$ e $\bigcup_n A_n = A$, allora $P(A_n) \uparrow P(A)$

b) Se $A_n \searrow A$, cioè $A_n \supseteq A_{n+1}$ e $\bigcap_n A_n = A$, allora $P(A_n) \downarrow P(A)$

Dim:

a) $B_n := A_n \setminus A_{n-1}$, $B_1 = A_1$, B_k sono a due a due disg., $\bigcup_{k=1}^{\infty} B_k = A$, $\bigcup_{k=1}^{n} B_k = A_n$, quindi

$$P(A) = \sum_{k=1}^{\infty} P(B_k) = \lim_n \sum_{k=1}^{n} P(B_k) = \lim_n P(A_n)$$

b) Da (a) prendendo A_n^c.

Oss: Se P è prob finitamente additiva, allora abbiamo (iii) $\Leftrightarrow$ (a) $\Leftrightarrow$ (b) (esercizio)

Lemma: $(\Omega, \mathcal{F}, P)$ sp di probabilità, A_n, $n \in \mathbb{N}$ eventi. Allora

$P(\bigcup_n A_n) \leq \sum_n P(A_n)$ (σ-subadditività)

In particolare (prendendo $A_{N+1} = A_{N+2} = \ldots = \emptyset$) $P(\bigcup_{n=1}^{N} A_n) \leq \sum_{n=1}^{N} P(A_n)$

Dim:

$B_1 = A_1$, $B_n = A_n \setminus \bigcup_{k=1}^{n-1} A_k$, $n \geq 2$, allora B_n disgiunti a due a due, $B_n \subseteq A_n$, $\bigcup_n B_n = \bigcup_n A_n$, quindi

$$P(\bigcup_n A_n) = P(\bigcup_n B_n) = \sum_n P(B_n) \leq \sum_n P(A_n)$$

Def. Un evento $A \in \mathcal{F}$ si dice $\begin{cases} \text{trascurabile: se } P(A)=0 \\ \text{quasi certo: se } P(A)=1 \end{cases}$

Si dice che una data proprietà q si verifica quasi certamente (q.c.):
se $\exists\, A \in \mathcal{F}$ quasi certo tale che q è vera su A

Esercizio: se $\#\Omega = \#\mathbb{N}$ e $\mathcal{F} = \mathcal{P}(\Omega)$, $\nexists$ P prob su $(\Omega, \mathcal{P}(\Omega))$ t.c.
$P\{\omega\} = c$ (indipendente da ω) $\forall \omega \in \Omega$

Altre definizioni della probabilità:

Storicamente, prima della definizione assiomatica di Kolmogorov (data negli anni Trenta del Novecento), sono state date altre definizioni, relative al concetto di probabilità e alla modellizzazione di esperimenti aleatori. Ne riportiamo due:

- Definizione classica: considera la probabilità come rapporto tra numero di casi favorevoli e numero di casi possibili, è espressa rigorosamente nel modello uniforme
  - → adatta per descrivere situazioni come estrazioni o lanci di dado
  - → conveniente per alcuni calcoli elementari (si riconduce al modello uniforme)
  - → conveniente per l'intuizione (se $P(A) = 1/4$, possiamo pensare a un'estrazione da una popolazione in cui A si verifica in $\frac{1}{4}$ dei casi)

* meno adatta per fenomeni come esperimenti fisici (effetti da errore) o biologici (ad es. effetto della somministrazione di un farmaco, effetto da variabilità dei pazienti)

- Definizione frequentista: considera la probabilità di un evento come il limite, per $n \to \infty$, della frequenza relativa dell'evento in n prove ripetute dell'esperimento.

→ più adatta per fenomeni come esperimenti fisici o biologici

* richiede di sapere a priori che il limite esiste, e non è chiara la def. rigorosa di limite

* operativamente, non sempre è possibile effettuare tante prove di un esperimento

L'introduzione della definizione assiomatica permette sia di dotare la probabilità di un impianto matematico rigoroso, sia di separare il problema matematico da quello modellistico:

- problema matematico: dato uno spazio di probabilità $(\Omega, \mathcal{A}, P)$, studiarne le proprietà
- " modellistico: dato un problema reale, scegliere il modello probabilistico più adeguato (cioè scrivere le informazioni date del problema in termini di P)

Sequenze di esperimenti

sabato 14 marzo 2020 13:38

Modellizzare sequenza ordinate (finita o infinita) di esperimenti

- $\Omega = \{ (\omega_1, \omega_2, \omega_3, \dots) \mid \omega_1 = \text{esito del } 1^\circ \text{ esperimento}$
  $\omega_2 = $ " " 2° "
  ...
  $\}$

  $= \{ (\omega_1, \omega_2, \omega_3, \dots) \mid \omega_1 \in \Omega_1, \omega_2 \in \Omega_{2,\omega_1}, \omega_3 \in \Omega_{3,\omega_1,\omega_2}, \dots \}$

  dove $\Omega_1 = \{$esiti del 1° esperimento$\}$

  $\Omega_{2,\omega_1} = \{$esiti del 2° " $\}$ (possibilmente dipendenti da ω_1)

  $\Omega_{3,\omega_1,\omega_2} = \{$esiti del 3° " $\}$ (" " " ω_1, ω_2)

  ecc.

  Se Ω_k non dipende da $\omega_1, \dots \omega_{k-1}$, $\forall k$, allora

  $\Omega = \times_k \Omega_k$

- Data una sequenza di due esperimenti, con esiti rispettivamente Ω_1, Ω_2 (quindi $\Omega = \Omega_1 \times \Omega_2$)

  dato A_k evento relativo all'esperimento k-simo ($A_k \subseteq \Omega_k$), $k = 1, 2$

  l'evento $\{$accade A_1 al primo esperimento$\}$ è

  $\{(\omega_1, \omega_2) \mid \omega_1 \in A_1, \omega_2 \in \Omega_2\} = A_1 \times \Omega_2$

  e analogamente $\{$accade A_2 al secondo esperimento$\}$ è

  $\{(\omega_1, \omega_2) \mid \omega_1 \in \Omega_1, \omega_2 \in A_2\} = \Omega_1 \times A_2$

  L'evento $\{A_1$ al primo esperimento, A_2 al secondo$\}$ è

  $\{(\omega_1, \omega_2) \mid \omega_1 \in A_1, \omega_2 \in A_2\} = A_1 \times A_2$

  Analogamente per tre o più esperimenti

- Esempi

  - Estrazione ordinata di 3 biglie con rimpiazzo da un'urna di 10

    $\Omega = \{ (x_1, x_2, x_3) \mid x_i \in \{1, \dots 10\}\} = \{1, \dots 10\}^3$

    $\{3^a$ biglia estratta $\in \{5, \dots 10\}\}$

    $= \{ (x_1, x_2, x_3) \mid x_3 \geq 5\} = \{1, 2, \dots 10\}^2 \times \{5, 6, \dots 10\}$

    $\{1^a$ oppure 2^a biglia $\in \{5, \dots 10\}\}$

    $= \{ (x_1, x_2, x_3) \mid x_1 \geq 5$ o $x_2 \geq 5\} = \{5, 6, \dots 10\} \times \{1, 2, \dots 10\}^2 \cup \{1, 2, \dots 10\} \times \{5, 6, \dots 10\} \times \{1, 2, \dots 10\}$

- Estrazione ordinata di 3 biglie senza rimpiazzo da un'urna di 10

  $\Omega = \{(x_1, x_2, x_3) \mid x_i \in \{1, .. 10\},\ x_i \text{ tutti distinti}\}$

- Prove di Bernoulli:

  sequenza infinita di esperimenti, ciascun esperimento ha come esiti successo (1) e insuccesso (0)

  $\Omega = \{(a_1, a_2, a_3, ..) \mid a_i \in \{0,1\}\ \forall i \in \mathbb{N}^+\} = \{0,1\}^{\mathbb{N}^+}$

  $A_i = \{\text{successo all' } i\text{-sima prova}\} = \{a \mid a_i = 1\}$

  ($\mathcal{F} = \sigma\{A_1, A_2, A_3, \dots\}$ "la più piccola σ-algebra contenente $A_1, A_2, \dots$")

  $\{$1° successo alla 3ª prova$\} = A_1^c \cap A_2^c \cap A_3 = \{a \mid a_1 = a_2 = 0, a_3 = 1\}$

  $\{$solo successi dalla 10ª prova$\} = \bigcap_{n \geq 10} A_n$

  $\{$solo successi da una certa prova in poi$\} = \bigcup_{m \in \mathbb{N}^+} \bigcap_{n \geq m} A_n$

- (Date due urne ciascuna con 10 biglie, scelta dell'urna e successiva estrazione

  $\Omega = \{(\omega_1, \omega_2) \mid \omega_1 \in \{A, B\}, \omega_2 \in \{1, 2, ..10\}\} = \{A, B\} \times \{1, 2, .. 10\}$)

- Consegna urgente/non urgente, in/fuori città

  $\Omega = \{(\omega_1, \omega_2) \mid \omega_1 \in \{U, NU\},\ \omega_2 \in \{I, F\}\}$

- Modellizzare possibili esiti di un'estrazione di k oggetti da un insieme U, senza rimpiazzo senza considerare l'ordine

  $\Omega = \{\{x_1 ... x_k\} \mid x_i \in U\ \forall i,\ x_i \text{ tutti distinti}\}$

  $= \{A \subseteq U \mid \# A = k\}$

3
giovedì 29 dicembre 2022 12:29

Probabilità condizionata e indipendenza

Esempio: Lancio di un dado equilibrato ($\Omega = \{1, 2, \dots 6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, P uniforme)

Supponiamo, dopo aver lanciato il dado (e prima di aver visto il risultato) di ricevere l'info che l'esito è pari, come cambia la probabilità?

La nuova probabilità sarà allora uniforme sugli esiti pari $\{2, 4, 6\}$, cioè, per $A \subseteq \Omega$,

$$\text{prob di } A \text{ dato } \underbrace{\text{esito pari}}_{B} = \frac{\#\text{ esiti in } A \text{ tra quelli pari}}{\#\text{ esiti pari}} = \frac{\#(A \cap B)}{\#B}$$

$$= \frac{\#(A \cap B)/\#\Omega}{\#B/\#\Omega} = \frac{P(A \cap B)}{P(B)}$$

Def. Dato $B \in \mathcal{F}$ con $P(B) > 0$, probabilità condizionata di $A \in \mathcal{F}$ dato B,

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

$P(A|B)$ esprime la prob. che accada A sapendo che accade B

Lemma: Fissato $B \in \mathcal{F}$ con $P(B) > 0$, la funzione

$$P(\cdot|B) : \mathcal{F} \to \mathbb{R}, \quad A \mapsto P(A|B)$$

è una probabilità (cioè soddisfa (i), (ii), (iii) della def. assiomatica di probabilità)

Dim: esercizio

Oss: $B \mapsto P(A|B)$ <u>non</u> è una probabilità

Lemma: Dati $A, B \in \mathcal{F}$ con $P(B) > 0$, vale $P(A \cap B) = P(A|B) P(B)$

Più in generale, dati $A_1 \dots A_n \in \mathcal{F}$ con $P(A_1 \cap \dots \cap A_n) > 0$,

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

Dim: . $P(A \cap B) = P(A|B) P(B)$ da def.

. Se $P(A_1 \cap \dots \cap A_n) > 0$, allora $P(A_1 \cap \dots \cap A_i) > 0 \;\; \forall i$ e

$$P(A_1 \cap \dots \cap A_n) = P(A_n|A_1 \cap \dots \cap A_{n-1}) P(A_1 \cap \dots \cap A_{n-1}) = P(A_n|A_1 \cap \dots \cap A_{n-1}) P(A_{n-1}|A_1 \cap \dots \cap A_{n-2})$$

$$= \dots \quad \text{(per induzione)}$$

Diciamo che $B_1 \dots B_n$ formano un sistema di alternative se $(B_1 \dots B_n)$ è una partizione di Ω (cioè $B_i \cap B_j = \emptyset \;\; \forall i \neq j$, $\bigcup_{i=1}^n B_i = \Omega$) e i B_k sono eventi non trascurabili (cioè $B_k \in \mathcal{F}$, $P(B_k) > 0$)

Prop (Formula della partizione o delle prob. totale): $B_1 \dots B_n$ sistema di alternative. Allora, $\forall A \in \mathcal{F}$,

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

Dim: $A = \bigcup_{i=1}^n (A \cap B_i)$ unione di insiemi a due a due disgiunti, quindi

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

Oss: La formula della partizione si generalizza facilmente al caso $n=\infty$ (cioè $(B_n)_{n\in\mathbb{N}}$ partizione numerabile di Ω in eventi non trascurabili).

Questa formula serve per ricavare la probabilità di un evento A a partire dalle probabilità condizionate a un sistema di alternative B_i ($P(A|B_i)$).

Esempio tipico: Due urne, l'urna (1) contiene 5 biglie rosse e 5 blu, l'urna (2) contiene 8 biglie rosse e 2 blu. Esperimento: scegliamo casualmente un'urna e dall'urna scelta estraiamo una biglia e ne osserviamo il colore. Prob di estrarre una biglia rossa?

$\Omega = \{1,2\}\times\{R,B\}$ $\quad$ (oppure $\Omega = \{(x,y) \mid x\in\{1,2\},\ y\in\{1_R,\dots 5_R, 1_B,\dots 5_B\}$ se $x=1$, $y\in\{1_R,\dots 8_R, 1_B, 2_B\}$ se $x=2\}$)

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

$P\{\text{urna } 1\} = P(\{1\}\times\{R,B\}) = \frac{1}{2} = P\{\text{urna } 2\}$

$P(\{\text{rossa}\}|\{\text{urna } 1\}) = P(\{1,2\}\times\{R\} \mid \{1\}\times\{R,B\}) = \frac{5}{10} = \frac{1}{2}$

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

$P\{\text{rossa}\} \underset{\text{formula partizione}}{=} P(\{\text{rossa}\}|\{\text{urna } 1\})P\{\text{urna } 1\} + P(\{\text{rossa}\}|\{\text{urna } 2\})P\{\text{urna } 2\} = \frac{13}{20}$

Prop (formula di Bayes): Siano $A, B\in\mathcal{F}$ con $P(A)>0$, $P(B)>0$. Allora

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

Cor: $B_1,\dots B_n$ sistema di alternative. Allora, $\forall i$

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

Dim formula di Bayes:

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

La formula di Bayes è utile per "invertire" il condizionamento (ricavare $P(B|A)$ da $P(A|B)$)

Esempio tipico: Due urne, urna (1) con 5 rosse e 5 blu, urna (2) con 8 rosse e 2 blu. Se la biglia estratta è rossa, qual è la prob che venga dall'urna (1)?

$$P(\{\text{urna } 1\}|\{\text{rossa}\}) \underset{\text{Bayes}}{=} \frac{P(\{\text{rossa}\}|\{\text{urna } 1\})\cdot P\{\text{urna } 1\}}{P\{\text{rossa}\}} = \frac{\frac{1}{2}\cdot\frac{1}{2}}{13/20} = \frac{5}{13}$$

Altro esempio tipico: falsi positivi: dati:
- prob di essere malato
- prob di test positivo se la persona è sana
- " " " negativo " " " malata

calcolare prob. che persona con test positivo sia malata

Oss: stesso problema: sano/malato ↔ urna, test ↔ biglia

Oss: la risposta dipende dalla prob (non condizionata) di essere malato!

Se si pensa ai B_i come eventi "causa" e ad A come evento "osservato", la formula di Bayes fornisce la probabilità della "causa" dato l'evento "osservato".

Def: Due eventi A e B si dicono (stocasticamente) indipendenti se
$P(A \cap B) = P(A)\,P(B)$

Oss: A, B eventi con $P(B)>0$. Allora
A, B sono indipendenti $\iff P(A|B) = P(A)$

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

Significato dell'indipendenza: A e B indipendenti $\iff$ la probabilità di A non cambia sapendo che accade B

Esempio: Estrazione di una carta da un mazzo di 40 carte napoletane
$[\Omega = \{1_B, \dots 10_B, 1_C, \dots 10_C, 1_O, \dots 10_O, 1_S, \dots 10_S\},\ \mathcal{F} = \mathcal{P}(\Omega),\ P \text{ uniforme}]$
$A = \{asso\} = \{1_B, 1_C, 1_O, 1_S\}$, $B = \{oro\} = \{1_O, \dots 10_O\}$ sono indipendenti.

Oss: • Se A e B sono indipendenti, allora A^c e B sono indipendenti
A e B^c " "
A^c e B " "

• $P(A) \in \{0,1\} \iff A$ è indipendente da ogni evento $\iff A$ è indipendente da sé stesso

• Due eventi incompatibili A e B ($A\cap B = \emptyset$) sono indipendenti $\iff$ almeno uno è trascurabile ($P(A)=0$ o $P(B)=0$)

Dim: per esercizio.

Def: Sia $(A_i)_{i\in I}$ una famiglia di eventi. (A_i) è una famiglia di eventi indipendenti (o A_i, $i\in I$, sono collettivamente indipendenti) se
$\forall J \subseteq I$ finito ($\neq\emptyset$), $P(\bigcap_{i\in J} A_i) = \prod_{i\in J} P(A_i)$

Oss: Se $(A_i)_{i\in I}$ è famiglia di eventi indipendenti, allora $(A_i)_{i\in I'}$ è famiglia di eventi indipendenti $\forall I' \subseteq I$

Oss: Invece, l'indipendenza a due a due non implica l'indipendenza della famiglia.

Esempio 1:

• $\Omega = \{1,2,3,4\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, P uniforme
$A = \{1,2\}$, $B = \{1,3\}$, $C = \{1,4\}$ sono a due a due indipendenti, ma non sono collettivamente indip.

Esempio 2:

- Due lanci indipendenti di moneta equilibrata + un lancio di moneta truccata, con esito testa se i primi due lanci hanno esito concorde, croce altrimenti

$\Omega = \{T, C\}^3$, $\mathcal{F} = \mathcal{P}(\Omega)$, P t.c. $P\{T \text{ al } 1°\} = P\{C \text{ al } 1°\} = \frac{1}{2}$

$P\{T \text{ al } 2°\} = P\{C \text{ al } 2°\} = \frac{1}{2}$

$\{T \text{ al } 1°\}$, $\{T \text{ al } 2°\}$ indipendenti

$P\{T \mid \text{primi due lanci concordi}\} = 1$

$P\{C \mid$ " " discordi$\} = 1$

$A_i = \{\text{Testa al lancio } i\text{-simo}\}$ sono a due a due indipendenti ma non collettivamente indipendenti

Oss: L'indipendenza stocastica non è necessariamente l'assenza di causa-effetto:

- Due eventi A e B possono non essere (direttamente) legati da rapporto causa-effetto, pur essendo dipendenti:

Esempio: ...

- Due eventi A e C possono essere indipendenti, pur essendo legati da rapporto causa effetto

Esempio: Nell'esempio 2 precedente (due lanci di moneta equilibrati + uno truccato)

$A = \{\text{Testa al } 1°\}$, $C = \{\text{testa al } 3°\}$ sono indipendenti

ma l'esito del primo lancio, assieme all'esito del 2°, influenza l'esito del 3° lancio

Lemma: Se $(A_i)_{i \in I}$ è famiglia di eventi indipendenti, allora $(A_i^{\alpha_i})_{i \in I}$ è famiglia di eventi indipendenti, $\forall (\alpha_i) \in \{1, c\}^I$, con $A_i^1 = A_i$, $A_i^c = \Omega \setminus A_i$,

Lemma: Dati $A_1, \dots A_n \in \mathcal{F}$ eventi, allora sono equivalenti

- $A_1 \dots A_n$ indipendenti
- $P(A_1^{\alpha_1} \cap \dots \cap A_n^{\alpha_n}) = P(A_1^{\alpha_1}) \cdot \dots \cdot P(A_n^{\alpha_n}) \quad \forall (\alpha_i)_{i=1,\dots n} \in \{1, c\}^n$

Prove ripetute: consideriamo un esperimento ripetuto n volte (o anche infinite volte) nelle stesse condizioni di partenza, ci aspettiamo che l'esito di una o più ripetizioni non influenzi l'esito nelle altre ripetizioni: quindi, in n ripetizioni di un esperimento eventi riferiti a una o più prove sono indipendenti da eventi riferiti ad altre prove; (se A_i è un evento riferito all'i-sima ripetizione, allora $A_1, \dots A_n$ sono indipendenti)

Questo si applica ad esempio a

- n lanci di moneta o di dado (non necessariamente equilibrato)
  es: $\{1°\text{ lancio} = 6\}$, $\{4°\text{ lancio pari}\}$, $\{5°\text{ lancio} \le 2\}$ sono indipendenti
  $\{1°\text{ e } 3°\text{ lancio pari}\}$, $\{2°\text{ lancio} = 3\}$ sono indipendenti
- n estrazioni con ordine e con rimpiazzo
- ripetizioni di un esperimento "di Bernoulli", cioè con esito successo o insuccesso
  $\Omega = \{0, 1\}^n$, ad es: {successo al 1° lancio}, {insuccesso al 3° lancio} sono indipendenti
- ...

Da un punto di vista rigoroso, possiamo prendere
$\Omega = S^n$ dove S è l'insieme degli esiti dell'esperimento, supponiamo S al più numerabile, $\mathcal{F} = \mathcal{P}(\Omega)$,
P t.c., $\forall B_1, \dots B_n \subseteq S$, gli eventi $A_i = S \times \dots \times S \times \underset{\text{i-sima posiz.}}{B_i} \times S \dots \times S = \{B_i \text{ all'i-sima prova}\}$
siano indipendenti

Si dimostra (come vedremo):
- P esiste ed è unica
- se la prob sulla singola prova è uniforme, allora P è uniforme (si ritrova modello uniforme per lanci ed estrazioni con rimpiazzo)
- se A e B sono eventi riferiti a gruppi disgiunti di prove (ad es $A = A_1 \times A_2 \times \Omega \times \Omega$, $B = \Omega \times \Omega \times B_1 \times B_2$), allora A e B sono indipendenti; analogamente per più di due eventi.

Probabilità su spazi discreti

Consideriamo Ω discreto, cioè finito o numerabile, e $\mathcal{F}=\mathcal{P}(\Omega)$, scriviamo $\Omega=\{\omega_1,\omega_2,\dots\}$

Def: Funzione di densità discreta p di P: $p:\Omega\to\mathbb{R}$

$p_i = p(\omega_i) = P\{\omega_i\}, \quad \omega_i\in\Omega$

Lemma: Ω discreto

a) Data P prob su $(\Omega,\mathcal{P}(\Omega))$, la sua funzione di densità discreta p soddisfa

i) $p(\omega_i)\geq 0 \ \forall i$

ii) $\sum_{i\in\Omega} p(\omega_i)=1$

b) Viceversa, data $p:\Omega\to\mathbb{R}$ funzione che soddisfa (i) e (ii), $\exists!\, P$ prob su $(\Omega,\mathcal{P}(\Omega))$ avente p per densità discreta, e P è data da

$$P(A)=\sum_{\omega_i\in A} p\{\omega_i\} \quad (*)$$

Oss: In particolare, da p possiamo calcolare $P(A)$ con $(*)$

Dim:

a) $p(\omega_i)=P\{\omega_i\}\geq 0$

$\Omega=\bigcup_{\omega_i\in\Omega}\{\omega_i\}$ unione al più numerabile, quindi

$1=P(\Omega)=\sum_{\omega_i\in\Omega}P\{\omega_i\}=\sum_{\omega_i\in\Omega}p(\omega_i)$

b) Unicità: se P ha p come densità disreta, allora, $\forall A\in\mathcal{P}(\Omega)$,

$A=\bigcup_{\omega_i\in A}\{\omega_i\}$ unione disgiunta al più numerabile, quindi

$P(A)=\sum_{\omega_i\in A}P\{\omega_i\}=\sum_{\omega_i\in A}p(\omega_i)$, cioè $(*)$

Esistenza: Data p, definiamo $P:\mathcal{P}(\Omega)\to\mathbb{R}$ tramite $(*)$, dobbiamo verificare che P è probabilità:

1) $P(A)\geq 0 \ \forall A\in\mathcal{P}(\Omega)$: ok da (i)

2) $P(\Omega)=1$: ok da (ii)

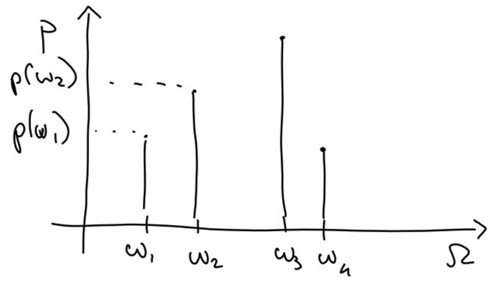
3) σ-additività: $A=\bigcup_{n\in\mathbb{N}}A_n$, A_n a due a due disgiunti

$\sum_n P(A_n)=\sum_n\sum_{i\,|\,\omega_i\in A_n}p(\omega_i)$

A_n a due a due disgiunti $\Rightarrow \{i\,|\,\omega_i\in A_n\}_n$ è partizione di $\{i\,|\,\omega_i\in\bigcup_n A_n\}$

$=\sum_{i,\,\omega_i\in\bigcup_n A_n}p(\omega_i)=P(\bigcup_n A_n)$

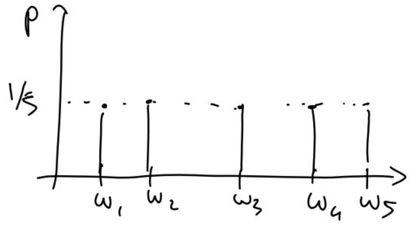
Rappresentazione della densità discreta tramite grafico a barre



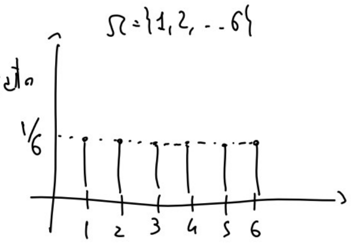
(ω_i non necessariamente ordinati)

Il grafico a barre indica dove P assegna più "massa"

Esempio: distribuzione uniforme su Ω finito $= \{\omega_1, \omega_2 - \omega_n\}$



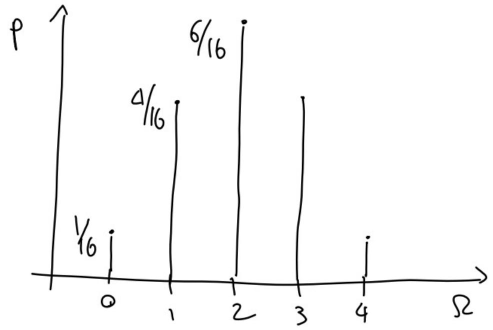
esito lancio dado equilibrato



Esempio: n° teste in 4 lanci di moneta:

$\Omega = \{$possibili n° di teste$\} = \{0, 1, 2, 3, 4\}$ $\quad \mathcal{F} = \mathcal{P}(\Omega)$

densità di probabilità $p(\omega) = P\{\omega \text{ teste}\}$



$P\{$al massimo 2 teste$\} = P\{0, 1, 2\} = p(0) + p(1) + p(2) = \frac{11}{16}$

Si può estendere la def. di probabilità discreta e densità discreta al caso di Ω più che numerabile:

Def: Dato $(\Omega, \mathcal{F})$ sp. misurabile con $\{\omega\} \in \mathcal{F}$ $\forall \omega \in \Omega$, P (prob. su $(\Omega, \mathcal{F})$) si dice discreta:

se $\exists\, \Omega_0 \subseteq \Omega$, Ω_0 al più numerabile, t.c.

$P(\Omega_0) = 1$ (P è concentrata su Ω_0 al più numerabile)

In questo caso, definiamo funzione di densità discreta $p: \Omega_0 \to \mathbb{R}$, $p(\omega) = P\{\omega\}$,

e la estendiamo a $p: \Omega \to \mathbb{R}$ ponendo $p(\omega) = P\{\omega\} = 0$ $\forall \omega \notin \Omega_0$.

Esempio: $\Omega_0 = \{0, 1, 2, \dots 10\}$, $\Omega = \mathbb{R}$

Oss tecnica (esercizio): $(\Omega, \mathcal{F})$ sp. misurabile t.c. $\{\omega\} \in \mathcal{F}$ $\forall \omega \in \Omega$

a) Se P è prob discreta su $(\Omega, \mathcal{F})$, concentrata su Ω_0 al più numerabile,

$P|_{\Omega_0} : \mathcal{P}(\Omega_0) \to \mathbb{R}$, $\mathcal{P}(\Omega_0) \ni A \mapsto P(A) \in \mathbb{R}$

è una probabilità (discreta) su $(\Omega_0, \mathcal{P}(\Omega_0))$.

b) Viceversa, se $\Omega_0 \subseteq \Omega$ è al più numerabile e Q è una probabilità su $(\Omega_0, \mathcal{P}(\Omega_0))$,

$P : \mathcal{F} \to \mathbb{R} \mid A \mapsto Q(A \cap \Omega_0)$

è probabilità discreta su $(\Omega, \mathcal{F})$

c) Se P è prob discreta su $(\Omega, \mathcal{F})$, concentrata su Ω_0 al più numerabile, con p densità discreta, allora

$$P(A) = \sum_{\omega \in A \cap \Omega_0} p(\omega) \text{"}=\text{"} \sum_{\omega \in A} p(\omega) \qquad \forall A \in \mathcal{F}$$

Def: Data P probabilità discreta su $(\Omega, \mathcal{P}(\Omega))$ (o $(\Omega, \mathcal{F})$ con $\mathcal{F} \ni \{\omega\}$ $\forall \omega \in \Omega$), con densità p, chiamiamo range di P

$$R_p := \{\omega \in \Omega \mid p(\omega) = P\{\omega\} > 0\}$$

Oss: $P(R_p) = 1$ e $P|_{R_p} : \mathcal{P}(R_p) \to \mathbb{R}$, $A \mapsto P(A)$ è prob. su $(R_p, \mathcal{P}(R_p))$

inoltre $P(A) = \sum_{x \in A \cap R_p} p(x)$

5

venerdì 30 dicembre 2022 19:38

Esempi notevoli di probabilità discrete

- Premessa: Sequenza di Bernoulli di parametro $p \in [0,1]$:

  successione, finita o infinita, di prove indipendenti, dove ciascuna prova ha esito successo (1) con probabilità p o insuccesso (0) con probabilità $1-p$ (prova di Bernoulli)

  Modello corrispondente.

  $\bar{\Omega}_n = \{0,1\}^n$ (caso n prove)

  $\bar{\mathcal{F}}_n = \mathcal{P}(\bar{\Omega}_n)$ $A_k := \{$successo alla k-sima prova$\} = \{a \in \Omega \mid k = 1\}$

  $\left.\begin{array}{l} \cdot\ \bar{P}_n(A_k) = p \quad \forall k = 1, \dots n \\ \cdot\ A_k \text{ indipendenti sotto } \bar{P}_n \end{array}\right\} (*)$

  Nel caso infinite prove: $\bar{\Omega}_\infty = \{0,1\}^{\mathbb{N}^+}$ $\bar{\mathcal{F}}_\infty = \sigma\{A_1, A_2, \dots\}$ "la più piccola σ-alg contenente $A_1, A_2, \dots$"

  (contiene tutti gli eventi rilevanti)

(Prop: $\exists !$ misura di probabilità $\bar{P}_n$ su $(\bar{\Omega}_n, \bar{\mathcal{F}}_n)$ che soddisfa $(*)$, $n \in \mathbb{N}^+ \cup \{+\infty\}$)

Esempi:

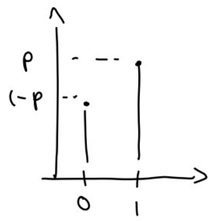
- lanci di moneta equilibrata, successo = "testa", $p = \frac{1}{2}$
- lanci di dado equilibrato; successo = "5", $p = \frac{1}{6}$
- estrazioni con ordine con rimpiazzo da un'urna, successo = "biglia rossa", $p = \frac{\#\text{rosse}}{\#\text{biglie}}$
- risposte casuali a un test a crocette, in cui ogni domanda ha quattro opzioni, con una sola giusta:

  prova = risposta a singola domanda, successo = risposta giusta, $p = \frac{1}{4}$
- ...

- Distribuzione di Bernoulli di parametro $p \in [0,1]$ $(B(p))$

  $\Omega = \{0,1\}$ $(\mathcal{F} = \mathcal{P}(\Omega))$

  $p(k) = \begin{cases} 1-p & k = 0 \\ p & k = 1 \end{cases}$



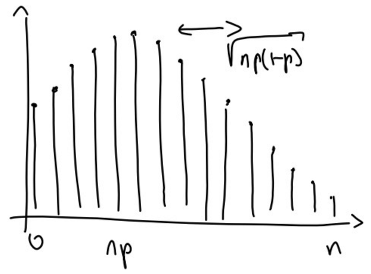
$B(p)$ rappresenta un esperimento con esito successo (1) o insuccesso (0), in cui il successo ha probabilità p (prova di Bernoulli) $(P\{1\} = \bar{P}_n\{\omega \mid \omega_i = 1 \ \forall i\})$

· Distribuzione binomiale di parametri $n \in \mathbb{N}^+$ e $p \in [0,1]$ $(B(n,p)$ o $Bin(n,p))$

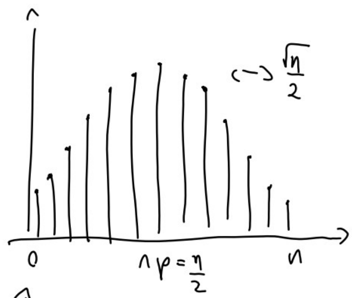
$\Omega = \{0, 1, \dots n\}$, $(\mathcal{F} = \mathcal{P}(\Omega))$

$p(k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k \in \Omega = \{0, 1, \dots n\}$

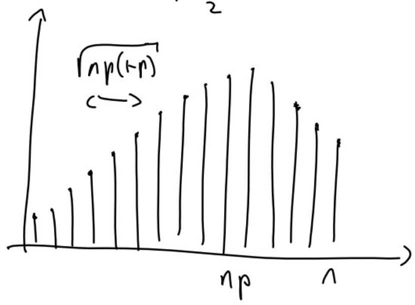
$p < \frac{1}{2}$



$p = \frac{1}{2}$



$p > \frac{1}{2}$



$B(n,p)$ è la distribuzione del n° di successi $(0, 1, \dots n)$ in una sequenza di n prove di Bernoulli di parametro p

cioè $\quad p(k) = \bar{P}_n\{k \text{ successi (su } n \text{ prove)}\} = \bar{P}_n\{X = k\} \quad k \in \{0, 1, \dots n\}$

con $X : \bar{\Omega}_n \to \Omega$, $X(\omega_1, \dots \omega_n) = \#$ successi dell'esito $(\omega_1, \dots \omega_n) = \omega_1 + \dots + \omega_n$

Dim:

$\{k \text{ successi}\} = \bigcup_{i_1, \dots i_k \text{ indici distinti in } \{1, \dots n\}} \{\text{successo alle prove } i_1, \dots i_k, \text{ insuccesso alle altre}\}$

$= \bigcup_{I \subseteq \{1, \dots n\}, \#I = k} \underbrace{\bigcap_{i \in I} A_i}_{\text{successo alle prove di indice } i} \cap \underbrace{\bigcap_{i \notin I} A_i^c}_{\text{insuccesso alle altre}}$

unione disgiunta

$\forall I \subseteq \{1, \dots n\}$, $\#I = k$, abbiamo $\bar{P}_n\left(\bigcap_I A_i \cap \bigcap_{I^c} A_i^c\right) = \prod_{i \in I} \bar{P}_n(A_i) \cdot \prod_{i \notin I} \bar{P}_n(A_i^c) = p^k (1-p)^{n-k}$

Il numero di $I \subseteq \{1, \dots n\}$ con $\#I = k$ è $\binom{n}{k}$

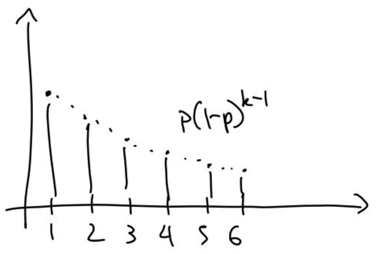
Quindi

$$\bar{P}_n\{k \text{ successi}\} = \sum_{I, \#I = k} \bar{P}_n\left(\bigcap_I A_i \cap \bigcap_{I^c} A_i^c\right) = \binom{n}{k} p^k (1-p)^{n-k}$$

· Distribuzione geometrica di parametro $p \in (0,1)$ $\quad (G(p))$

$\Omega = \mathbb{N}^+ = \{1, 2, 3, \dots\}$ $\quad (\mathcal{F} = \mathcal{P}(\Omega))$

$p(k) = p(1-p)^{k-1} \qquad k \in \Omega = \mathbb{N}^+$



$G(p)$ rappresenta l'istante (cioè il n° della prova) del primo successo, in una sequenza di Bernoulli di parametro p

cioè $\quad p(k) = \overline{P}_\infty(\text{primo successo alla } k\text{-sima prova}) = \overline{P}_\infty(T = k) \qquad k \in \mathbb{N}^+ = \{1, 2, 3, \dots\}$

con $T : \overline{\Omega}_\infty \to \mathbb{N}^+ \cup \{+\infty\}$, $T(\omega_1, \omega_2, \dots) =$ istante del primo successo nell'esito $(\omega_1, \omega_2, \dots)$

$= \inf\{i \in \mathbb{N}^+ \mid \omega_i = 1\}$, con $\inf \emptyset = +\infty$

Dim:

$$\begin{aligned}
\overline{P}_\infty(1^\circ \text{ successo alla } k\text{-sima prova}) &= \overline{P}_\infty(A_1^c \cap A_2^c \dots \cap A_{k-1}^c \cap A_k) \\
&= \overline{P}_\infty(A_1^c) \cdot \overline{P}_\infty(A_2^c) \cdots \overline{P}_\infty(A_{k-1}^c)\, \overline{P}_\infty(A_k) \\
&= (1-p)^{k-1} p
\end{aligned}$$

$$\left(\overline{P}_\infty(\text{nessun successo}) = 1 - \sum_{k=1}^{\infty} \overline{P}_\infty(\text{primo successo alla } k\text{-sima prova}) = 1 - \sum_{k=1}^{\infty} p(1-p)^{k-1} = 1 - 1 = 0 \right)$$

· Distribuzione binomiale negativa di parametri $h \in \mathbb{N}^+$ e $p \in (0,1)$ $\quad (BinNeg(h,p))$

$\Omega = \{h, h+1, h+2, \dots\}$ $\quad (\mathcal{F} = \mathcal{P}(\Omega))$ $\quad (BinNeg(1,p) = G(p))$

$$p(k) = \binom{k-1}{h-1} p^h (1-p)^{k-h}$$

$BinNeg(h,p)$ rappresenta l'istante dell'h-simo successo, in una sequenza di Bernoulli di parametro p

cioè $\quad p(k) = \overline{P}_\infty\{h\text{-simo successo alla } k\text{-sima prova}\} = \overline{P}_\infty\{T_h = k\}$

con $T_h : \overline{\Omega}_\infty \to \Omega \cup \{+\infty\}$, $T_h(\omega_1, \omega_2, \dots) =$ istante dell'h-simo successo in $(\omega_1, \omega_2, \dots)$

$= \inf\{i \in \mathbb{N}^+ \mid \omega_1 + \dots + \omega_i \geq h\}$

Dim: esercizio

- Distribuzione ipergeometrica di parametri $N \in \mathbb{N}^+$, $N_1, n \in \mathbb{N}$ con $1 \le N_1, n \le N$ $\quad (H(N, N_1, n))$

  $\Omega = \{$numeri naturali tra $0 \vee (n-(N-N_1))$ e $n \wedge N_1\}$, $\quad (\mathcal{F} = \mathcal{P}(\Omega))$

$$p(k) = \frac{\binom{N_1}{k}\binom{N-N_1}{n-k}}{\binom{N}{n}} \qquad k \in \Omega$$

Significato modellistico:

- in un'estrazione senza ordine senza rimpiazzo di n oggetti da un'urna di N oggetti
- dove gli N oggetti sono divisi in un gruppo (a) di N_1 oggetti e un gruppo (b) di $N-N_1$ oggetti
- $H(N, N_1, n)$ è la distribuzione del numero di oggetti del gruppo (a) tra quelli estratti

Dim:

Sia $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ il modello per estrazioni senza ordine e senza rimpiazzo di n oggetti tra N, così definito: $\tilde{\Omega} = \{S \subseteq \{1, .. N\} \mid \#S = n\}$, $\tilde{\mathcal{F}} = \mathcal{P}(\tilde{\Omega})$, $\tilde{P}$ uniforme su $\tilde{\Omega}$

$$\#\tilde{\Omega} = \binom{N}{n}$$

L'evento $A = \{k$ oggetti del gruppo (a), $n-k$ del gruppo (b)$\}$ ha cardinalità

$\#A = \#$ modi di estrarre k oggetti all'interno del gruppo (a) $\cdot$

$\quad \cdot \; \#$ " " " $n-k$ " " " " " (b)

$$= \binom{N_1}{k}\binom{N-N_1}{n-k}$$

Quindi $P(A) = \dfrac{\#A}{\#\Omega} = p(k)$.

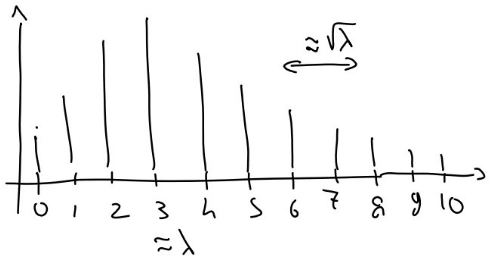
Più in generale, se gli N oggetti di un'urna sono divisi in gruppi $(a_1), .. (a_m)$, con $N_1, .. N_m$ elementi rispettivamente $(N_1 + .. + N_m = N)$, ed estraiamo n oggetti, allora

$$P\{k_1 \text{ elementi di } (a_1), .. k_m \text{ elementi di } (a_m)\} = \frac{\binom{N_1}{k_1} \cdot \ldots \cdot \binom{N_m}{k_m}}{\binom{N}{n}} \qquad \forall (k_1, .. k_m) \text{ con } \begin{array}{l} 0 \le k_i \le N_i \; \forall i \\ k_1 + .. + k_m = n \end{array}$$

- Distribuzione di Poisson di parametro $\lambda > 0$ $\quad (\mathcal{P}(\lambda)$ o $\mathrm{Poisson}(\lambda))$

  $\Omega = \mathbb{N} = \{0, 1, 2, ..\}$ $\quad (\mathcal{F} = \mathcal{P}(\Omega))$

$$p(k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad k \in \mathbb{N}$$



$P(\lambda)$ è la distribuzione "del n° di successi in una sequenza di n prove di Bernoulli di parametro p, con $n \gg 1$, $p \ll 1$ e $np \approx \lambda$" (distribuzione degli eventi rari)

Prop: Sia $(p_n)_{n \in \mathbb{N}}$ successione in $(0,1)$ con $np_n \to \lambda > 0$ per $n \to \infty$

Sia $p_n(k) = \binom{n}{k} p_n^k (1-p_n)^{n-k}$ la densità discreta $Bin(n,p)$ (estesa a $k \in \mathbb{N}$ ponendo $p_n(k) = 0 \;\; \forall k > n$).

Allora, $\forall k \in \mathbb{N}$, $p_n(k) \to p(k) = \frac{\lambda^k}{k!} e^{-\lambda}$ densità Poisson (λ).

Dim: esercizio

Esempi di applicazione:

- numero di particelle α emesse da una sorgente radioattiva in un'unità di tempo
  in questo caso
  - la singola prova è l'emissione o la non emissione di una particella da parte di un nucleo
  - le prove sono indipendenti
  - n° di prove = n° di nuclei $\gg 1$
  - prob di emissione per singolo nucleo $\ll 1$
- numero di utenti che accedono a uno sportello/sito web/...
  in questo caso:
  - singola prova: singolo utente accede/non accede al servizio
  - le prove sono indip
  - n° di prove = n° di utenti $\gg 1$
  - prob di accedere per singolo utente $\ll 1$

6
lunedì 2 gennaio 2023 17:41

Variabili aleatorie discrete

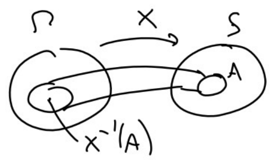
Def: Dato Ω spazio discreto (cioè finito o numerabile), con $\mathcal{F} = \mathcal{P}(\Omega)$, S insieme $\neq \emptyset$, una variabile aleatoria (v.a.) da Ω a S è una funzione
$$X: \Omega \to S$$
Tipicamente $S = \mathbb{R}$: v.a. reale
$S = \mathbb{R}^n$: vettore aleatorio

Esempio:
Data una sequenza di n lanci di moneta, $\Omega = \{T, C\}^n$, sono esempi di v.a. reali
- $X_i := \begin{cases} 1 & \text{se esce T alla i-sima estrazione} \\ 0 & \text{altrimenti} \end{cases}$

  $X_i((\omega_1, \dots \omega_n)) = \begin{cases} 1 & \text{se } \omega_i = T \\ 0 & \text{se } \omega_i = C \end{cases}$
- $S_n =$ n° di teste $= X_1 + X_2 + \dots + X_n$
- $(X_1, X_2, \dots X_n): \Omega \to \mathbb{R}^n$ è esempio di vettore aleatorio

Oss: Significato e utilità delle v.a:
- Le v.a. rappresentano delle quantità aleatorie, cioè delle caratteristiche quantitative degli esiti dell'esperimento aleatorio.
- È possibile effettuare operazioni algebriche (come somma, prodotto) e analitiche (come limiti) con le v.a.
- Ciò che conta sono le distribuzioni delle v.a. e le loro "relazioni" (indipendenza, distribuzione congiunta)

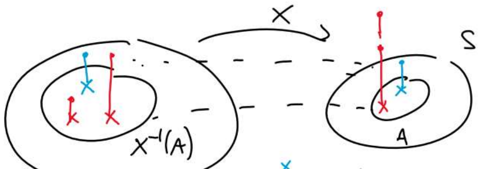
Notazione: Dato $A \subseteq S$, $X^{-1}(A) := \{X \in A\} := \{\omega \in \Omega \mid X(\omega) \in A\}$



Fatto: X^{-1} commuta rispetto alle operazioni insiemistiche:
dati $A, A_i, i \in I$, sottoinsiemi di S,
$$X^{-1}\left(\bigcup_{i \in I} A_i\right) = \bigcup_{i \in I} X^{-1}(A_i)$$
$$X^{-1}(A^c) = X^{-1}(A)^c \qquad \text{(facile verifica per esercizio)}$$
$$X^{-1}\left(\bigcap_{i \in I} A_i\right) = \bigcap_{i \in I} X^{-1}(A_i)$$

Def. Dati: Ω spazio discreto, P probabilità su $(\Omega, \mathcal{P}(\Omega))$, $X: \Omega \to S$ v.a., $S_X := X(\Omega) = \{X(\omega) \mid \omega \in \Omega\}$,
legge (o distribuzione) di X su S_X (o probabilità immagine di P tramite X)
P^X (o $X_\# P$ o $P \circ X^{-1}$ o $X(P)$): misura di probabilità su $(S_X, \mathcal{P}(S_X))$ definita da

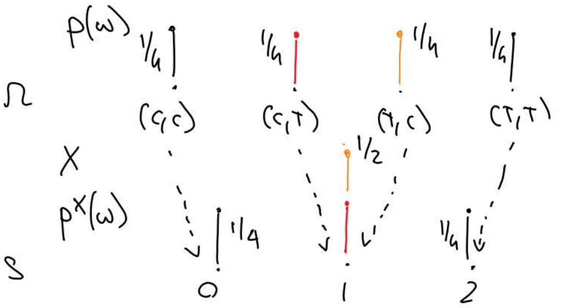
$P^X(A) = P\{X\in A\} = P(X^{-1}(A))$, $A\subseteq S_X$



(blu è controimmagine di blu
rosso è " " rosso)

Esempio: $\Omega = \{$due lanci di moneta equilibrata$\} = \{(T,T),(T,C),(C,T),(C,C)\}$, P uniforme

X = n° teste : $\Omega \to S = \{0,1,2\}$, $X(\omega_1,\omega_2) = 1_{\{T\}}(\omega_1) + 1_{\{T\}}(\omega_2)$



P^X è la prob associata all'esperimento "n° di teste in 2 lanci di moneta equilibrata"

In generale, se (Ω, P) modellizza un certo esperimento, che chiamiamo "esp", e $X:\Omega\to S$, (S_X, P_X) modellizza l'esperimento "eseguiamo "esp" e misuriamo X dell'esito".

Lemma: P^X è una probabilità sullo spazio discreto $(S_X, \mathcal{P}(S_X))$

Dim: . $P^X(A) = P\{X\in A\}\in[0,1]$ $\forall A\in\mathcal{P}(S_X)$

. $P^X(S_X) = P\{X\in S_X\} = P(\Omega) = 1$

. $\forall (A_n)_{n\in\mathbb{N}}$ a due a due disgiunti, $A_n\subseteq S_X$,

$$P\{X\in \bigcup_n A_n\} = P\Big(\bigcup_n \underbrace{\{X\in A_n\}}_{\text{a due a due disgiunti}}\Big) = \sum_n P\{X\in A_n\}$$

Oss: In quanto prob. discreta, si può estendere la legge anche a $\mathcal{P}(S)$

$P^X(A) = P\{X\in A\} = P^X(A\cap S_X)$, $A\subseteq S$

e in generale a $\mathcal{G}$, con $\mathcal{G}$ σ-algebra su S t.c. $\{x\}\in\mathcal{G}$ $\forall x\in S$. Chiameremo anche questa estensione legge di X su S.

Oss: Si può anche restringere P^X a una prob. su $(R_{P^X}, \mathcal{P}(R_{P^X}))$, $R_{P^X} = \{x\in S \mid P^X\{x\} = P\{X=x\}>0\}$

Poiché P_X è probabilità sullo spazio discreto $(\underset{=\{x_1,x_2,\dots\}}{S_X}, \mathcal{P}(S_X))$, possiamo definire la funzione di densità discreta

$p^X: S_X\to\mathbb{R}$, $p^X(x_i) = P\{X=x_i\}$, $x_i\in S_X$

ed estenderla a S ponendo $p^X(x) = 0$ $\forall x\in S\setminus S_X$

La relazione tra legge e densità dà

$$P\{X\in A\} = P^X(A) = \sum_{x\in A} p^X(x) \quad \forall A\subseteq S, \quad \text{dove } \sum_{x\in A} = \sum_{x\in A\cap S_X} = \sum_{x\in A\cap R_{P^X}}$$

Se una v.a. X ha distribuzione binomiale / Poisson / ...., diciamo che X è binomiale / Poisson / ...

Esempio: v.a. indicatrice: data Ω sp discreto, P prob su $(\Omega, \mathcal{P}(\Omega))$, $A \subseteq \Omega$,

$$\mathbb{1}_A : \Omega \to \mathbb{R}, \quad \mathbb{1}_A(\omega) = \begin{cases} 1 & \text{se } \omega \in A \\ 0 & \text{se } \omega \notin A \end{cases}$$

$\mathbb{1}_A \sim B(p)$, con $p = P(A)$ ("$\sim$" = ha legge)

$S_{\mathbb{1}_A} = \{0, 1\}$, $P\{\mathbb{1}_A = 1\} = p$, $P\{\mathbb{1}_A = 0\} = 1 - p$

Esempio: schema di Bernoulli di n prove, def di binomiale

$(\Omega = \{0,1\}^n \quad \mathcal{F} = \mathcal{P}(\Omega))$ $A_i = \{a \in \Omega \mid a_i = 1\}$, P t.c. $P(A_i) = p$, A_i indip.

$X_i = \mathbb{1}_{A_i} \sim B(p)$

$X = \#\text{ successi} = \sum_{i=1}^{n} \mathbb{1}_{A_i}$

Esempio: 5 lanci di dado equilibrato, prob che il 4 esca al max 2 volte?

(5 prove ripetute indip, successo = "4" di prob. $1/6$)

$(\Omega = \{1, 2, .. 6\}^5, \mathcal{F} = \mathcal{P}(\Omega))$ $\qquad A_i = \{\text{"4" all'i-sima prova}\}$ $\cdot\ P(A_i) = 1/6$

$\cdot\ A_i$ indip.

$X = \#\text{"4" nelle 5 prove} = \#\text{ successi} = \sum_{i=1}^{5} \mathbb{1}_{A_i} \sim B(5, \frac{1}{6})$

$$P\{\text{al massimo 2 volte "4"}\} = P\{X \le 2\} = p_X(0) + p_X(1) + p_X(2)$$
$$= \binom{5}{0}\left(\frac{5}{6}\right)^5 + \binom{5}{1} \cdot \frac{1}{6} \cdot \left(\frac{5}{6}\right)^4 + \binom{5}{2}\left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right)^3$$

Esempio: estrazione di un individuo da una popolazione e misurazione di una data quantità X discreta

$\Omega = \{\text{individui della popolazione}\}$, P uniforme, $X(\omega)$ = una certa caratteristica di ω

es: $\Omega = \{1_R, 2_R, .. 6_R, 1_B, 2_B, .. 4_B\}$, con k_R biglia rossa, k_B biglia blu

$$X(\omega) = \begin{cases} 1 & \text{se } \omega \in \{1_R, .. 6_R\} \\ 0 & \text{altrimenti} \end{cases}$$

$$p^X(x) = P\{X = x\} = \frac{\#\text{ individui } \omega \text{ con } X(\omega) = x}{\#\text{individui}} = \text{frequenza relativa di } X = x \text{ sulla popolazione}$$

es: $p^X(1) = P\{X = 1\} = \dfrac{\#\text{ biglie rosse}}{\#\text{ biglie}}$ = frequenza relativa di biglie rosse su tutte le biglie

Problema: Data una probabilità Q su $(S, \mathcal{P}(S))$, S discreto, esistono uno spazio discreto Ω una probabilità P su $(\Omega, \mathcal{P}(\Omega))$ e una v.a. $X : \Omega \to S$ che ha Q come legge?

Sì: costruzione canonica:

$\Omega = S$, $P = Q$, $X : \Omega \to S$, X = identità $(X(\omega) = \omega \ \forall \omega \in \Omega)$

$\forall A \in \mathcal{P}(S)$, $P\{X \in A\} = P(A) = Q(A)$

Composizione di v.a.: $\quad \Omega$ spazio discreto, P prob su $(\Omega, \mathcal{P}(\Omega))$

- Se $X : \Omega \to S$ è v.a. e $f : S \to S'$, allora

  $f(X) : \Omega \to S'$, $f(X)(\omega) = f(X(\omega))$ è v.a.

· Se $X:\Omega\to S_1$, $Y:\Omega\to S_2$ sono v.a., allora

$(X,Y):\Omega\to S_1\times S_2$, $(X,Y)(\omega)=(X(\omega),Y(\omega))$, $\omega\in\Omega$, è v.a. a valori in $S_1\times S_2$ (blocco o coppia di v.a.)

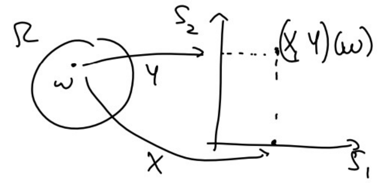
Attenzione: $(X,Y):\Omega\to S_1\times S_2$, non $(X,Y):\Omega^2\to S_1\times S_2$

Analogamente, se $X_i:\Omega\to S_i$, $i=1,\dots,n$, sono v.a., allora

$(X_1,\dots,X_n):\Omega\to S_1\times\dots\times S_n$, $(X_1,\dots,X_n)(\omega)=(X_1(\omega),\dots,X_n(\omega))$, $\omega\in\Omega$, è v.a.

· Se $X:\Omega\to\mathbb{R}$, $Y:\Omega\to\mathbb{R}$ sono v.a. reali, allora

$X+Y$, XY, ... sono v.a.



Infatti $X+Y=f(X,Y)$ con $f:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$, $(x,y)\mapsto x+y$

$XY=f(X,Y)$ $\quad f:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$, $(x,y)\mapsto xy$

Uguaglianza q.c. e in legge di v.a.: Ω spazio discreto, P prob. su $(\Omega,\mathcal{P}(\Omega))$

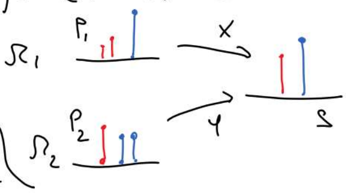
Def: Date $X:\Omega\to S$, $Y:\Omega\to S$ v.a. discrete, $X=Y$ q.c.: se

$P\{X=Y\}=1$

Def: Dati $(\Omega_1,\mathcal{P}(\Omega_1),P_1)$, $(\Omega_2,\mathcal{P}(\Omega_2),P_2)$ spazi di probabilità con Ω,Ω' discreti, date $X:\Omega_1\to S$, $Y:\Omega_2\to S$ v.a., X e Y hanno la stessa legge ($X\overset{(d)}{=}Y$): se

$P_1^X=P_2^Y$ (come probabilità su $(S,\mathcal{P}(S))$)

cioè $P_1\{X\in A\}=P_2\{Y\in A\}$ $\forall A\in\mathcal{P}(S)$.



Oss: $P_1^X=P_2^Y \iff p_1^X=p_2^Y$ su S, con p_1^X, p_2^Y densità discrete

$\iff R_{p_1^X}=R_{p_2^Y}$ e $p_1^X=p_2^Y$ su $R_{p_1^X}$, con $R_{p_1^X}=\{x\in S \mid p_1^X(x)>0\}$

Oss: Data $\mathcal{G}$ σ-algebra su S, con $\{x\}\in\mathcal{G}$ $\forall x\in S$, vale

$P_1^X=P_2^Y$ su $(S,\mathcal{P}(S)) \iff P_1^X=P_2^Y$ su $(S,\mathcal{G}) \iff p_1^X=p_2^Y$

Oss: Se $X=Y$ q.c., allora $X\overset{(d)}{=}Y$, ma il viceversa non vale

Dim: · Se $X=Y$ q.c., allora, $\forall A\in\mathcal{P}(S)$,

$$P\{X\in A\}=P\{X\in A, X=Y\}+\underbrace{P\{X\in A, X\neq Y\}}_{\leq P\{X\neq Y\}=0}$$

$$=P\{Y\in A, X=Y\}+P\{Y\in A, X\neq Y\}$$

$$=P\{Y\in A\}$$

Esempio di X,Y v.a. non uguali q.c. ma con $X\overset{(d)}{=}Y$

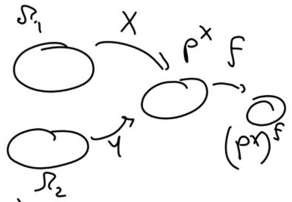
$X\sim B(\frac{1}{2})$, $Y=1-X$, allora

· $Y\sim B(\frac{1}{2})$: Y ha valori in $0,1$ q.c (cioè $P\{Y\in\{0,1\}\}=1$) e $P\{Y=1\}=P\{X=0\}=\frac{1}{2}$

· $P\{X\neq Y\}=1$: $P\{X=Y\}=P\{2X-1=0\}=P\{X=\frac{1}{2}\}=0$

Stabilità per composizione:

- Se $X:\Omega\to S$ e $Y:\Omega\to S$ sono uguali q.c. e $f:S\to S'$, allora $f(X)=f(Y)$ q.c.

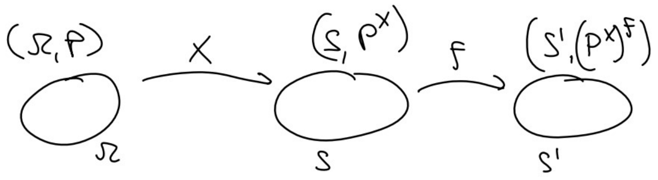


- Se $X:\Omega_1\to S$ e $Y:\Omega_2\to S$ hanno la stessa legge e $f:S\to S'$, allora $f(X)$ e $f(Y)$ hanno la stessa legge:

infatti $P_1\{f(X)\in A\} = P_1\{X\in f^{-1}(A)\}$ $\quad\forall A\in\mathcal{P}(S')$

$$\underset{X\overset{(d)}{=}Y}{=} P_2\{Y\in f^{-1}(A)\} = P_2\{f(Y)\in A\}$$

Notiamo che $P^{f(X)} = (P^X)^f$



7

mercoledì 4 gennaio 2023 10:41

Distribuzioni congiunte e indipendenza di v.a.

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

Siano $X_1 : \Omega \to S_1, \dots X_n : \Omega \to S_n$ v.a., cioè $X = (X_1, \dots X_n) : \Omega \to S_1 \times \dots \times S_n$ v.a.

Chiamiamo

- legge congiunta di $(X_1, \dots X_n)$: la legge P^X di X su $S_1 \times \dots \times S_n$
- leggi marginali: le leggi $P^{X_1} \dots P^{X_n}$ rispettivamente di $X_1 \dots X_n$ su $S_1, \dots S_n$

  e più in generale la legge di $(X_j)_{j\in J}$ su $\times_{j\in J} S_j$, per $J \subsetneq I$

Oss: $\forall A_i \subseteq S_i$, $\{X_1 \in A_1, \dots X_n \in A_n\} = \{(X_1, \dots X_n) \in A_1 \times \dots \times A_n\}$

e quindi $P\{X_1 \in A_1, \dots X_n \in A_n\} = P\{(X_1, \dots X_n) \in A_1 \times \dots \times A_n\} = P_{(X_1, \dots X_n)}(A_1 \times \dots \times A_n)$

Analogamente chiamiamo:

- densità discreta congiunta: la densità di X su $S_1 \times \dots \times S_n$
- densità discrete marginali: le densità di $X_1, \dots X_n$ risp. su $S_1, \dots S_n$

La densità congiunta

$$p_{X_1 \dots X_n}(x_1, \dots x_n) = P\{X_1 = x_1, \dots X_n = x_n\} \quad (x_1, \dots x_n) \in S_1 \times \dots \times S_n$$

è la prob. che congiuntamente $X_1 = x_1, \dots X_n = x_n$, mentre la densità marginale

$$p_{X_i}(x) = P\{X_i = x\} \quad x \in S_i$$

è, a i fissato, la prob che $X_i = x_i$

Prop: La legge congiunta determina le leggi marginali, e precisamente, dette p_X la densità congiunta, p_{X_i} le densità marginali, $i = 1, \dots n$,

$$p_{X_i}(x) = \sum_{x_j \in S_j,\, j \neq i} p_X(x_1, \dots x_{i-1}, x, x_{i+1}, \dots x_n) \qquad x \in S_i$$

$$P\{X_i \in A\} = P\{(X_1, \dots X_n) \in S_1 \times \dots \times S_{i-1} \times A \times S_{i+1} \times \dots \times S_n\} \qquad A \in \mathcal{P}(S_i)$$

Dim:

La seconda formula è immediata dall'osservazione

$$\{X_i \in A\} = \{(X_1, \dots X_n) \in S_1 \times \dots \times S_{i-1} \times A \times S_{i+1} \times \dots \times S_n\}$$

La prima formula segue dalla seconda, prendendo $A = \{x\}$ e osservando che

$$\sum_{(x_1, \dots x_n) \in S_1 \times \dots \times S_{i-1} \times \{x\} \times S_{i+1} \times \dots \times S_n} \dots = \sum_{x_j \in S_j,\, j \neq i,\, x_i = x} \dots$$

Esempio: ($\alpha \in (0,1)$ parametro)

$(X,Y): \Omega \to \{-1,1\}^2$, $\quad p_{(X,Y)}(1,1) = p_{(X,Y)}(-1,1) = \frac{\alpha}{2}$, $\quad p_{(X,Y)}(1,-1) = p_{(X,Y)}(-1,1) = \frac{1-\alpha}{2}$

$p_X(1) = p_{(X,Y)}(1,1) + p_{(X,Y)}(1,-1) = \frac{\alpha}{2} + \frac{1-\alpha}{2} = \frac{1}{2} \quad p_X(-1) = 1 - p_X(1) = \frac{1}{2}$

$p_Y(1) = p_{(X,Y)}(1,1) + p_{(X,Y)}(-1,1) = \frac{\alpha}{2} + \frac{1-\alpha}{2} = \frac{1}{2} \quad p_Y(-1) = 1 - p_Y(1) = \frac{1}{2}$

Oss: La legge congiunta contiene informazioni più ricche su $(X_1 .. X_n)$ rispetto alle leggi marginali, poiché codifica anche le relazioni tra le X_i:
infatti le leggi marginali non determinano la legge congiunta:

Esempio:

a) Due lanci di moneta equilibrata, X_i := esito lancio i-simo, $i=1,2$ (con 1=testa, 0=croce)

b) Un lancio di moneta equilibrata, uno con moneta truccata che ripete il 1° lancio, Y_i := esito lancio i-simo, $i=1,2$

$X_i \sim B(\frac{1}{2})$ $i=1,2$, $\quad Y_1 \sim B(\frac{1}{2})$, $\quad Y_2 = Y_1$ q.c. e quindi $Y_2 \sim B(\frac{1}{2})$

Ma $P\{(X_1,X_2) = (1,0)\} = \frac{1}{4}$, $P\{(Y_1,Y_2) = (1,0)\} = 0$, quindi (X_1,X_2) e (Y_1,Y_2) hanno diverse leggi congiunte.

Indipendenza di v.a. ($(\Omega, \mathcal{P}(\Omega), P)$ spazio di prob discreto)

Def: Date $X_1: \Omega \to S_1$, $X_2: \Omega \to S_2$ v.a., X_1 e X_2 si dicono indipendenti se

$$P\{X_1 \in A_1, X_2 \in A_2\} = P\{X_1 \in A_1\} P\{X_2 \in A_2\} \quad \forall A_1 \in \mathcal{P}(S_1), A_2 \in \mathcal{P}(S_2)$$

Def: Date $X_1: \Omega \to S_1, .. X_n: \Omega \to S_n$ v.a., $(X_1, .. X_n)$ si dice famiglia di v.a. indipendenti se

$$P\{X_1 \in A_1, .. X_n \in A_n\} = P\{X_1 \in A_1\} \cdot ... \cdot P\{X_n \in A_n\} \quad \forall A_1 \in \mathcal{P}(S_1), .. A_n \in \mathcal{P}(S_n)$$

equivalentemente, $P_{(X_1,..X_n)}(A_1 \times .. \times A_n) = P_{X_1}(A_1) \cdot ... \cdot P_{X_n}(A_n) \quad \forall A_1 \in \mathcal{P}(S_1), .. A_n \in \mathcal{P}(S_n)$

Oss: $X_1, .. X_n$ sono indipendenti $\iff \forall A_1 \in \mathcal{P}(S_1), .. A_n \in \mathcal{P}(S_n)$, $\{X_1 \in A_1\} .. \{X_n \in A_n\}$ sono indip.

Dim: $\Leftarrow$: da def.

$\Rightarrow$: dobbiamo dim: $\forall 1 \le i_1 < i_2 < .. < i_k \le n$, $P\{X_{i_1} \in A_{i_1}, .. X_{i_k} \in A_{i_k}\} = P\{X_{i_1} \in A_{i_1}\} ... P\{X_{i_k} \in A_{i_k}\}$

ma questo segue prendendo $A_j = S_j \ \forall j \neq i_1, .. i_k$

Significato modellistico dell'indipendenza: $X_1 .. X_n$ sono indipendenti se informazioni sull'esito di una X_i (del tipo $X_i \in A_i$) non modificano le probabilità relative alle altre X_j.

Oss: L'indipendenza di $X_1 .. X_n$ è una proprietà della legge congiunta di $(X_1 .. X_n)$:
se $(X_1, .. X_n) \overset{(d)}{=} (Y_1, .. Y_n)$ e $(X_1 .. X_n)$ sono indipendenti, allora $(Y_1, .. Y_n)$ sono indipendenti

Invece, $(X_1 .. X_n)$ indipendenti e $X_i \overset{(d)}{=} Y_i \ \forall i \not\Rightarrow (Y_1 .. Y_n)$ indipendenti

Oss: $A_1 \dots A_n$ sono indipendenti $\iff \mathbb{1}_{A_1} \dots \mathbb{1}_{A_n}$ sono indipendenti (esercizio)

Oss: Se $(X_1, \dots X_n)$ è famiglia di v.a. indip., allora $(X_j)_{j\in J}$ è famiglia di v.a. indip $\forall J \subseteq \{1, \dots n\}$

Def: Sia $X_i : \Omega \to S_i$, $i \in I$, una famiglia di v.a. (con I possibilmente infinito), $(X_i)_{i\in I}$ si dice famiglia di v.a. indipendenti: se, $\forall J \subseteq I$ finito, $(X_j)_{j\in J}$ è famiglia di v.a. indipendenti, cioè $P\left(\bigcap_{j\in J} \{X_j \in A_j\}\right) = \prod_{j\in J} P\{X_j \in A_j\}$ $\quad \forall A_j \in \mathcal{P}(S_j),\ j \in J$

Lemma: Siano $X_i : \Omega \to S_i$, $i = 1, \dots n$, v.a., $X = (X_1, \dots X_n) : \Omega \to S_1 \times \dots \times S_n$, p_X la densità discreta congiunta, p_{X_i} le densità discrete marginali. Allora

$(X_i)_{i=1,\dots n}$ è famiglia di v.a. indipendenti se e solo se

$$p_X(x_1, \dots x_n) = p_{X_1}(x_1) \cdot \dots \cdot p_{X_n}(x_n) \quad \forall (x_1, \dots x_n) \in S_1 \times \dots \times S_n$$

In particolare, se (X_i) è famiglia di v.a. indipendenti, dalle leggi marginali si ricava la legge congiunta

Dim:

$\Rightarrow$) Basta prendere $A_1 = \{x_1\}, \dots A_n = \{x_n\}$ nella def. di indipendenza

$\Leftarrow$) Per $A_1 \in \mathcal{P}(S_1), \dots A_n \in \mathcal{P}(S_n)$, abbiamo:

$$P\{X_1 \in A_1, \dots X_n \in A_n\} = \sum_{(x_1, \dots x_n) \in A_1 \times \dots \times A_n} p_X(x_1, \dots x_n) = \sum_{x_1 \in A_1, \dots x_n \in A_n} p_X(x_1, \dots x_n)$$

$$= \sum_{x_1 \in A_1, \dots x_n \in A_n} p_{X_1}(x_1) \cdot \dots \cdot p_{X_n}(x_n)$$

$$= \left(\sum_{x_1 \in A_1} p_{X_1}(x_1)\right) \cdot \dots \cdot \left(\sum_{x_n \in A_n} p_{X_n}(x_n)\right) = P\{X_1 \in A_1\} \cdot \dots \cdot P\{X_n \in A_n\}$$

Esempio:

a) n estrazioni con ordine con reinserimento da urna U di N oggetti

X_i = esito dell'i-sima estrazione, $i = 1, \dots n$

$\Omega = U^n$, P uniforme, $X_i : U^n \to U$ i-sima proiezione canonica ($X_i(\omega_1, \dots \omega_n) = \omega_i$)

- ciascuna X_i ha legge uniforme U: $P\{X_i = x\} = \frac{1}{N}$ $\forall x \in U$
- $X_1, \dots X_n$ sono indipendenti

b) n estrazioni con ordine senza reinserimento da urna U di N oggetti ($n \leq N$)

Y_i = esito dell'i-sima estrazione, $i = 1, \dots n$

- ciascuna Y_i è v.a. uniforme su U $\qquad P\{Y_i = x\} = \frac{1}{N}$ $\forall x \in U$
- $Y_1, \dots Y_n$ non sono indipendenti: infatti

<u>dato</u> $\{Y_1 = y\}$, Y_2 ha legge uniforme su $U \setminus \{y\}$, quindi

$$P\{Y_2 = y' \mid Y_1 = y\} = \begin{cases} \frac{1}{N-1} & \text{se } y' \neq y \\ 0 & \text{se } y' = y \end{cases} \neq \frac{1}{N} = P\{Y_2 = y'\}$$

Esempio: modello di tipo Ising

n particelle con spin ± 1: $\Omega = \{-1,1\}^n$, $\mathcal{F} = \mathcal{P}(\Omega)$

hamiltoniana $H: \Omega \to \mathbb{R}$, $H(\sigma) = \sum_{i,j=1}^{n} J_{ij} \sigma_i \sigma_j$ $\quad J \in \mathbb{R}^{n\times n}$ simmetrica

$$P_\beta\{\sigma\} = \frac{1}{Z_\beta} e^{-\beta H(\sigma)} \qquad \sigma \in \Omega$$

$\beta \in \mathbb{R}$ parametro $\quad Z_\beta$ costante di rinormalizzazione: $Z_\beta = \sum_{\sigma\in\Omega} e^{-\beta H(\sigma)}$ (t.c. P_β sia prob.)

$X_i: \Omega \to \{-1,1\}$, $X_i(\omega) = \omega_i$ spin dell'i-esima particella, $i = 1, \dots n$

Ogni marginale $(P_\beta)_{X_i}$ è uniforme (esercizio): idea: P_β è invariante per flipping degli spin, cioè se $F: \Omega \to \Omega$, $F(\sigma) = -\sigma$, $(P_\beta)_F = P_\beta$.

- per $\beta = 0$: P_0 uniforme, X_i indipendenti (spin indipendenti)
- per $\beta > 0$, $J_{ij} = \begin{cases} 1 & |i-j| = 1 \\ 0 & \text{altrimenti} \end{cases}$: X_i non indipendenti, favorita la repulsione tra vicini
- per $\beta < 0$, J_{ij} come sopra: X_i non indipendenti, favorita l'attrazione tra vicini

Oss: Come per gli eventi, l'indipendenza a due a due delle X_i non implica l'indipendenza (collettiva) delle X_i: basta prendere $X_i = \mathbb{1}_{A_i}$, A_i a due a due indip. ma non (collettivamente) indip.

Lemma (stabilità per composizione): Se $X_i: \Omega \to S_i$, $i \in I$, sono indipendenti e $f_i: S_i \to S_i'$, $i \in I$, allora

$(f_i(X_i))_{i\in I}$ è famiglia di v.a. indipendenti

Dim: Segue da $\{f_i(X_i) \in A_i\} = \{X_i \in f_i^{-1}(A_i)\}$.

Lemma: Se $X_i: \Omega \to S_i$, $i \in I$ sono indipendenti e $J_1, J_2, \dots J_m \subseteq I$ sono disgiunti e finiti, allora $X_{J_1} = (X_i)_{i\in J_1}: \Omega \to \times_{i\in J_1} S_i$, $\dots$ $X_{J_m} = (X_i)_{i\in J_m}: \Omega \to \times_{i\in J_m} S_i$ sono indipendenti

(gruppi disgiunti di v.a. indipendenti sono indipendenti)

Dim: [chiamiamo $x_J = (x_i)_{i\in J}$ per $J \subseteq I$ finito]

Per indip., $\quad p_{X_J}(x_J) = \prod_{i\in J} p_{X_i}(x_i)$ $\forall J \subseteq I$ finito, quindi

$$P_{(X_{J_1},\dots X_{J_m})}(x_{J_1},\dots x_{J_m}) = \prod_{i\in J_1\cup\dots\cup J_m} p_{X_i}(x_i) = \prod_{k=1}^{m} \prod_{i\in J_k} p_{X_i}(x_i) = \prod_{k=1}^{m} p_{X_{J_k}}(x_{J_k})$$

Esempio: In una sequenza di lanci di dado, detti X_i gli esiti dei lanci, sono indip.

esito del primo lancio $= X_1$

somma degli esiti del 2° e 3° lancio $= X_2 + X_3 = g(X_2, X_3)$

massimo degli esiti del 4°, 5°, 6° lancio $= \max\{X_4, X_5, X_6\} = h(X_4, X_5, X_6)$

Date leggi marginali discrete P_i su $(S_i, \mathcal{P}(S_i))$, $i=1,..n$, è possibile costruire uno spazio discreto di probabilità $(\Omega, \mathcal{P}(\Omega), P)$ e v.a. $X_i: \Omega \to S_i$, $i=1,..n$, t.c la legge marginale di X_i sia P_i e le X_i, $i=1,..n$, siano indipendenti

basta prendere $\Omega = S_1 \times .. \times S_n$, P associata alla densità discreta (vedi prop. successiva)

$$p(x_1..x_n) = p_1(x_1) \cdot \ldots \cdot p_n(x_n), \quad (x_1..x_n) \in \Omega \qquad (*)$$

(dove p_i è la densità discreta di P_i)

e $X_i = \pi_i$ proiezione canonica di $\Omega = S_1 \times .. \times S_n$ su S_i

Prop: a) p data da $(*)$ è una densità discreta su $S_1 \times .. \times S_n$

b) Detta P la corrispondente probabilità su $(S_1 \times .. \times S_n, \mathcal{P}(S_1 \times .. \times S_n))$, P è l'unica probabilità su $(S_1 \times .. \times S_n, \mathcal{P}(S_1 \times .. \times S_n))$ che soddisfa

$$P(A_1 \times .. \times A_n) = P_1(A_1) \cdot \ldots \cdot P_n(A_n) \quad \forall A_1 \subseteq S_1, \ldots A_n \subseteq S_n$$

Dim: a) esercizio

b) come per dimostrazione indipendenza $\Leftrightarrow$ $p_{(X_1..X_n)}(x_1..x_n) = p_{X_1}(x_1) \cdot \ldots \cdot p_{X_n}(x_n)$

Def: La prob P data nella proposizione precedente si chiama prob. prodotto

$$P = P_1 \otimes \ldots \otimes P_n$$

Oss: Modello probabilistico per n esperimenti indipendenti

$(S_1 \times .. \times S_n, \mathcal{P}(S_1 \times .. \times S_n), P_1 \otimes .. \otimes P_n)$ è uno sp. di prob. per una sequenza di n esperimenti:

- con spazio campionario S_i (discreto) e prob. P_i, $i=1,..n$
- indipendenti (cioè con X_i indipendenti, dove $X_i: S_1 \times .. \times S_n \to S_i$, $X_i(\omega) = \omega_i$ è l'esito dell'i-simo esperimento, $i=1,..n$)

In particolare, dato un esperimento di spazio delle prob. $(S, \mathcal{P}(S), P)$ (discreto), n ripetizioni di questo esperimento (nelle stesse condizioni di partenza) si rappresentano con $(S^n, \mathcal{P}(S^n), P^{\otimes n})$ e le v.a. $X_i: S^n \to S$, $X_i(\omega) = \omega_i$ (esiti delle ripetizioni)

- sono indipendenti
- hanno la stessa legge P

Esercizio: Se P_i sono uniformi, allora $P_1 \otimes .. \otimes P_n$ è uniforme.

Esempio

n estrazioni con ordine con reinserimento da una popolazione/un'urna U di N oggetti
$X : U \to S$ v.a. che rappresenta una data caratteristica

(es: urna di N biglie di cui N_1 blu, $X: U \to \mathbb{R}$, $X(x) = \mathbb{1}_{\{blu\}}(x)$)

$X_i :=$ data caratteristica dell'i-simo oggetto estratto, $i = 1, \dots n$

$\Omega = U^n$, P uniforme, $X_i : U^n \to S$, $X_i = X \circ \pi_i$, π_i i-sima proiezione canonica

- ogni X_i ha la stessa legge di X
- le X_i sono indipendenti

Esempio: sequenza di Bernoulli di n esperimenti

$\Omega = \{0,1\}^n$ $\quad \mathcal{F} = \mathcal{P}(\Omega)$ $\quad A_i := \{a \in \Omega \mid a_i = 1\}$ (successo all'i-sima prova)

P prob su $(\Omega, \mathcal{P}(\Omega))$ t.c. :
- $P(A_i) = p \ \forall i$
- A_i indipendenti

Allora $X_i := \mathbb{1}_{A_i}$ sono v.a. $\sim B(p)$ e indipendenti

In particolare $X = \#$ successi $= \sum_{i=1}^{n} X_i \sim B(n,p)$ è somma di n v.a. $\sim B(p)$ indipendenti

Stessa cosa in un generico spazio Ω, con P che soddisfi $P(A_i) = p \ \forall i$, A_i indipendenti.

8

domenica 5 marzo 2023 13:21

Valore atteso, momenti e varianza

Ω spazio discreto, $\mathcal{F} = \mathcal{P}(\Omega)$, P prob su $(\Omega, \mathcal{P}(\Omega))$ con densità discreta p.

Def: Sia $X : \Omega \to \mathbb{R}$ v.a. con $X \geq 0$ (cioè $X(\omega) \geq 0 \ \forall \omega \in \Omega$). Valore atteso di X.

$$E[X] = \sum_{\omega \in \Omega} X(\omega)\, p(\omega) \in [0, +\infty]$$

Def. Sia $X : \Omega \to \mathbb{R}$ v.a. Diciamo che X è integrabile: se

$$E[|X|] = \sum_{\omega \in \Omega} |X(\omega)|\, p(\omega) < \infty.$$

In tal caso, definiamo valore atteso (o speranza o momento primo o valor medio o integrale) di X: il numero

$$E[X] = \sum_{\omega \in \Omega} X(\omega)\, p(\omega) \in \mathbb{R}$$

Significato del valore atteso:

- baricentro della distribuzione di X:

  $E[X] = \sum_i x_i\, p_X(x_i)$ (vedi sotto)

  media dei valori x_i assunti da X, pesata con la densità discreta p_X di X
- nell'esempio di estrazione da una popolazione Ω (P uniforme su Ω) e misurazione di una caratteristica quantitativa X,

  $E[X] = \frac{1}{\# \Omega} \sum_{\omega \in \Omega} X(\omega)$ è la media aritmetica della caratteristica su tutta la popolazione
- come vedremo (legge dei grandi numeri), dato un esperimento con esito quantitativo X, la media aritmetica dei risultati di n ripetizioni dell'esperimento converge, per $n \to \infty$, al valore atteso (se questo è finito)

Prop (calcolo di $E[X]$ tramite densità discreta):

a) $X : \Omega \to \mathbb{R}$ v.a., $\mathbb{R}_X = X(\Omega)$. Allora

- X è integrabile $\iff \sum_{x \in \mathbb{R}_X} |x|\, p_X(x) < \infty$
- Se X è integrabile o non-negativa,

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

b) Più in generale, date $X: \Omega \to S$ v.a., $S_X = X(\Omega)$, p_X densità discreta di X, $\varphi: S \to \mathbb{R}$, allora

- $\varphi(X)$ è integrabile $\iff \sum_{x \in S_X}' |\varphi(x)| p_X(x) < \infty$
- Se $\varphi(X)$ è integrabile o non-negativa,

$$E[\varphi(X)] = \sum_{x \in S_X} \varphi(x) p_X(x) = \sum_{x \in R_{p_X}}' \varphi(x) p_X(x) \qquad (\text{where } R_{p_X} = \{x \in S \mid p_X(x) > 0\})$$

Dim: (a) segue da (b) con $\varphi(x) = x$.

(b): Per $x \in S_X$, sia $A_x = \{X = x\} \subseteq \Omega$, $(A_x)_{x \in S_X}$ è una partizione al più numerabile di Ω.

Quindi, se $\varphi(X) \geq 0$,

$$E[\varphi(X)] = \sum_{\omega \in \Omega}' \varphi(X) p(\omega) = \sum_{x \in S_X} \sum_{\omega \in A_x}' \varphi(X) p(\omega)$$

$$= \sum_{x \in S_X}' \varphi(x) \sum_{\omega \in A_x} p(\omega) = \sum_{x \in S_X}' \varphi(x) \underbrace{P(A_x)}_{\overset{\shortparallel}{p_X(x)}}$$

Applicando la formula a $|\varphi(X)|$, otteniamo $\varphi(X)$ integrabile $\iff \sum_{x \in S_X} |\varphi(x)| p_X(x) < \infty$.

Per $\varphi(X)$ integrabile, ripetiamo i passaggi precedenti (grazie alla convergenza assoluta di $\sum_{x \in S_X} \varphi(x) p_X(x)$).

Oss importante: $E[\varphi(X)]$, se esiste, dipende solo dalla densità p_X di X, quindi dipende solo dalla legge di X.

- Esempio: Lancio di un dado equilibrato, vinciamo 2 se esce 6, -1 se esce 1 o 2, 0 altrimenti

X = vincita

$E[|X|] < \infty$ perché somma finita

$$E[X] = 2 \cdot P\{X = 2\} - 1 \cdot P\{X = -1\} + 0 \cdot P\{X = 0\} = 2 \cdot P\{6\} - 1 \cdot P\{1, 2\} = 2 \cdot \frac{1}{6} - 1 \cdot \frac{2}{6} = 0$$

- Esempio: $X \sim G(p)$, $p \in (0,1)$, $\alpha > 0$, $E[e^{\alpha X}] = ?$

$\exists\, E[e^{\alpha X}] \in [0, +\infty]$ perché $e^{\alpha X} \geq 0$

$$E[e^{\alpha X}] = \sum_{k=1}^{\infty} e^{\alpha k} P\{X = k\} = \sum_{k=1}^{\infty} e^{\alpha k} p (1-p)^{k-1} = e^{\alpha} p \sum_{h=0}^{\infty} \left(e^{\alpha}(1-p)\right)^h = \begin{cases} \dfrac{e^{\alpha} p}{1 - e^{\alpha}(1-p)} & \text{se } \alpha < \log \frac{1}{1-p} \\ +\infty & \text{se } \alpha \geq \log \frac{1}{1-p} \end{cases}$$

Oss: La def di $E[X]$ si estende in modo naturale al caso in cui $X \geq 0$ q.c. e anche al caso in cui $X \leq 0$ q.c.

Oss: Si può estendere la def. di valore atteso al caso in cui $E[X^+] < \infty$ oppure $E[X^-] < \infty$, dove $X^+ = \max\{X, 0\}$, $X^- = \max\{-X, 0\}$: in questo caso, si pone

$$E[X] := E[X^+] - E[X^-] \in [-\infty, +\infty]$$

Lemma (proprietà del valore atteso):

a) $X = c$ q.c. $\Rightarrow E[X] = c$

b) X v.a integrabile / ≥ 0 q.c, $a \in \mathbb{R} \Rightarrow aX$ integrabile / ≥ 0 q.c. o ≤ 0 q.c,

$E[aX] = aE[X]$ (con la convenzione $0 \cdot \infty = 0$)

c) X v.a. ≥ 0 q.c., $E[X] = 0 \Rightarrow X = 0$ q.c

d) $X \overset{(d)}{=} Y$, X integrabile / ≥ 0 q.c. $\Rightarrow Y$ integrabile / ≥ 0 q.c. e $E[Y] = E[X]$

in particolare, vero se $X = Y$ q.c.

e) X, Y integrabili / ≥ 0 q.c, $X \leq Y$ q.c. $\Rightarrow E[X] \leq E[Y]$; in particolare $E[X] \leq E[|X|]$

f) X, Y integrabili / ≥ 0 q.c. $\Rightarrow E[X+Y] = E[X] + E[Y]$

Dim:

a) facile esercizio

b) $E[|aX|] = \sum_{\omega \in \Omega} |aX(\omega)|\, p(\omega) = |a| \sum_{\omega \in \Omega} |X(\omega)| p(\omega) < \infty$ se X è integrabile, in questo caso

$E[aX] = \sum_{\omega \in \Omega} aX(\omega) p(\omega) = aE[X]$.

c) Se $X \geq 0$ q.c, allora $X(\omega)p(\omega) \geq 0$ $\forall \omega$. Se $E[X] = \sum_\omega X(\omega)\, p(\omega) = 0$, allora $X(\omega)p(\omega) = 0$ $\forall \omega$,

cioè $X = 0$ q.c.

d) da oss precedente

e) Se $X \leq Y$ q.c, allora $X(\omega)p(\omega) \leq Y(\omega)\, p(\omega)$ $\forall \omega \in \Omega$, quindi

$E[X] = \sum_\omega X(\omega) p(\omega) \leq \sum_\omega Y(\omega) p(\omega) = E[Y]$

f) $E[|X+Y|] = \sum_\omega |X(\omega) + Y(\omega)|\, p(\omega) \leq \sum_\omega (|X(\omega)| + |Y(\omega)|)\, p(\omega) = \sum_\omega |X(\omega)| p(\omega) + \sum_\omega |Y(\omega)|\, p(\omega)$

$< \infty$ se X e Y sono integrabili, in questo caso

$E[X+Y] = \sum_\omega (X(\omega) + Y(\omega))\, p(\omega) = \sum_\omega X(\omega)\, p(\omega) + \sum_\omega Y(\omega) p(\omega) = E[X] + E[Y]$.

Esempi notevoli:

- $X \sim B(p)$, $X = \mathbb{1}_A$ q.c con $A = \{X = 1\}$, $P(A) = p$

  $E[X] = E[\mathbb{1}_A] = 1 \cdot P(A) + 0 \cdot P(A^c) = P(A) = p$

- $X \sim B(n, p)$

  Prendiamo $X_1, \ldots X_n \sim B(p)$ indipendenti, $X := X_1 + \ldots + X_n \sim B(n, p)$

  $E[X] = \sum_{i=1}^n E[X_i] = np$

  Oss: Non è restrittivo supporre X somma di Bernoulli indip: se $Y \sim B(n, p)$, allora

  $E[Y] = E[X] = np$

- $X \sim G(p)$

  $X \geq 0$ q.c. quindi $\exists\, E[X] \in [0,+\infty]$

  $$E[X] = \sum_{k=1}^{\infty} k\, p (1-p)^{k-1} = \dots = \frac{1}{p}$$

- $X \sim \text{Poisson}(\lambda)$, $\lambda > 0$

  $X \geq 0$ q.c. quindi $\exists\, E[X] \in [0,+\infty]$

  $$E[X] = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \lambda \sum_{h=0}^{\infty} \frac{\lambda^h}{h!} e^{-\lambda} = \lambda$$

Oss: Se $X_1, \dots, X_n \sim B(p)$ anche non indipendenti, allora $E[X_1 + \dots + X_n] = np$

Esempio: "numero medio di punti fissi di una permutazione (di N elementi)"

$\Omega = S_N = \{\sigma : \{1,\dots,N\} \circlearrowleft \text{ bigettive}\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, P uniforme

$X = \#$ pt fissi $\quad X(\sigma) = \#\{i \in \{1,\dots,N\} \mid \sigma(i) = i\}$

$X = \sum_{i=1}^{N} \mathbb{1}_{C_i}$, con $C_i = \{\sigma \in S_N \mid \sigma(i) = i\}$

$P(C_i) = \frac{\# C_i}{\# \Omega} = \frac{(N-1)!}{N!} = \frac{1}{N}$ $\quad$ C_i non indipendenti (esercizio)

$E[X] = \sum_{i=1}^{N} P(C_i) = N \cdot \frac{1}{N} = 1$

Indici di centralità: valori di sintesi che indicano il centro della distribuzione P^X di X, per X v.a. reale

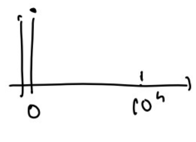
- valore atteso $E[X]$
- mediana: ogni valore $m_X \in \mathbb{R}$ t.c. $P\{X \leq m_X\} \geq \frac{1}{2}$, $P\{X \geq m_X\} \geq \frac{1}{2}$ (esercizio: esiste almeno una mediana)
- moda: ogni valore $mo_X \in \mathbb{R}$ di massimo per p_X (esercizio: esiste almeno una moda)

Oss: La mediana è un indicatore più "robusto" della media rispetto a valori estremi

Esempio: $X = \begin{cases} 10^4 & \text{con prob } 1/100 \\ 0 & \text{" " } 1 - 1/100 \end{cases}$ $\quad$ $E[X] = 10^4 \cdot \frac{1}{100} = 10^2$

mediana di X è 0

la mediana vede solo l'ordine dei valori di X, non la loro grandezza



Esempio: moda di Poisson(λ), $\lambda > 0$

max di $\mathbb{N} \ni k \mapsto p(k) = \frac{\lambda^k}{k!} e^{-\lambda}$

$$\frac{\lambda^k}{k!} e^{-\lambda} \geq \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \iff \lambda \geq k$$

quindi
- se $\lambda \notin \mathbb{N}$, l'unica moda è $[\lambda]$
- se $\lambda \in \mathbb{N}$, le mode sono $\lambda - 1$ e λ

Def: Dati X v.a. reale, $p \in \mathbb{R}$, momento assoluto di X di ordine p:

$$E[|X|^p]$$

Se $p \in \mathbb{Z}$ e $E[|X|^p] < \infty$, momento di X di ordine p:

$$E[X^p]$$

Lemma: Dati X v.a. reale, $1 \le p \le q < \infty$, se $E[|X|^q] < \infty$, allora $E[|X|^p] < \infty$

Dim:

$$E[|X|^p] = E[\underbrace{|X|^p \overbrace{\mathbb{1}_{|X| \le 1}}^{\mathbb{1}_{[-1,1]}(x)}}_{\le 1} + \underbrace{|X|^p \mathbb{1}_{|X|>1}}_{\le |X|^q \mathbb{1}_{|X|>1} \le |X|^q}]$$

$$\le 1 + E[|X|^q]$$

Oss. In realtà vale dis di Hölder: $E[|X|^p]^{1/p} \le E[|X|^q]^{1/q}$ per $1 \le p \le q$

Oss: Date X, Y v.a. reali, $1 \le p < \infty$, $a \in \mathbb{R}$, se $E[|X|^p] < \infty$ e $E[|Y|^p] < \infty$, allora

$$E[|aX + Y|^p] < \infty$$

come segue da $|aX+Y|^p \le 2^{p-1}(|a|^p |X|^p + |Y|^p)$.

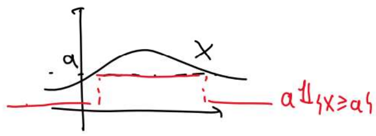
Prop (disuguaglianza di Markov):

Sia X v.a. reale ≥ 0. Allora, $\forall a > 0$,

$$P\{X \ge a\} \le \frac{E[X]}{a}$$

Dim:

$X \ge a \mathbb{1}_{\{X \ge a\}}$, quindi $E[X] \ge E[a \mathbb{1}_{\{X \ge a\}}] = a P\{X \ge a\}$



Cor: Sia X v.a. reale, sia $p > 0$. Allora

$$P\{|X| \ge a\} \le \frac{1}{a^p} E[|X|^p]$$

Dim:

$$P\{|X| \ge a\} = P\{|X|^p \ge a^p\} \le \frac{1}{a^p} E[|X|^p]$$

($\{|X| > a\}$, a grande)

Quindi i momenti controllano le code della distribuzione: più alto è l'ordine del momento assoluto finito di X, più "leggera" è la coda (più piccolo è $P\{|X| \ge a\}$)

Esercizio: X v.a. con densità $p_X : \mathbb{N}^+ \to \mathbb{R}$, $p_X(k) = c k^{-\alpha}$, $\alpha > 1$ parametro, $c = \left(\sum_{k=1}^{\infty} k^{-\alpha}\right)^{-1}$

Determinare per quali $p > 0$, $E[|X|^p] < \infty$ (sol. $p < \alpha - 1$)

Oss: Se X è limitata q.c., allora X ammette momenti di ogni ordine positivo ($p > 0$):

se $|X| \le M$, allora $E[|X|^p] \le M^p < \infty$

Esercizio: dati X v.a. reale, $a \in \mathbb{R}$, $\lambda > 0$, dimostrare $P\{X > a\} \leq e^{-\lambda a} E[e^{\lambda X}]$

Def: Sia X v.a. reale con $E[|X|^2] < \infty$. Varianza di X: il numero

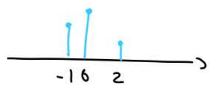
$$Var(X) = E[(X - E[X])^2] \in [0, +\infty)$$

Deviazione standard di X: $\sigma(X) = st(X) := \sqrt{Var(X)}$

Significato della varianza: la varianza è un indice della dispersione dei valori della distribuzione P^X rispetto al valore atteso, in quanto è la media dei quadrati degli scarti da $E[X]$: "i dati si discostano mediamente di $\sigma(X)$ da $E[X]$"

Oss. importante: la varianza dipende solo dalla legge P_X di X.

Esempio: lancio di un dado equilibrato

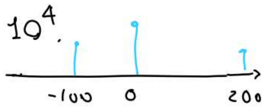


1) supponiamo di vincere 2 se esce 6, -1 se esce 1 o 2, 0 altrimenti

se X = vincita, $E[X] = 0$, $Var(X) = E[X^2] = 2^2 \cdot P\{X=2\} + (-1)^2 \cdot P\{X=-1\} = 4 \cdot \frac{1}{6} + 1 \cdot \frac{2}{6} = 1$

2) supponiamo ora di vincere 200 se esce 6, -100 se esce 1 o 2, 0 altrimenti

se Y = vincita, ancora $E[Y] = 0$, ma $Var(Y) = E[Y^2] = E[(100X)^2] = 10^4$.



Esempio: estrazione di un individuo da una popolazione e misurazione di una sua caratteristica quantitativa X ($\Omega = \{popol\}$, P uniforme, $X: \Omega \to \mathbb{R}$)

$$Var(X) = \sum_{\omega \in \Omega} (X(\omega) - E[X])^2 p(\omega) = \frac{1}{\#\Omega} \sum_{\omega \in \Omega} (X(\omega) - E[X])^2$$

media aritmetica dei quadrati degli scarti dalla media $E[X]$.

Proprietà della varianza: data X con $E[|X|^2] < \infty$,

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

b) Dati $a, b \in \mathbb{R}$, $Var(aX + b) = a^2 Var(X)$

c) $Var(X) \geq 0$, $Var(X) = 0 \Leftrightarrow X$ = costante q.c.

Dim:

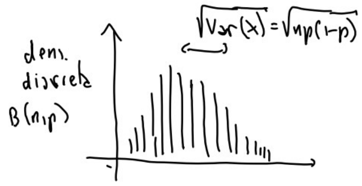
a) $Var(X) = E[X^2 - 2XE[X] + E[X]^2] = E[X^2] - 2E[X]E[X] + E[X]^2 = E[X^2] - E[X]^2$.

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

c) Poiché $(X - E[X])^2 \geq 0$, $Var(X) \geq 0$ e $= 0 \Leftrightarrow (X - E[X])^2 \geq 0 \Leftrightarrow X = E[X]$ = cost. q.c.

Esempi notevoli:

- $X \sim B(p)$
  $E[X^2] = 1^2 \cdot P\{X=1\} + 0^2 \cdot P\{X=0\} = p$, $Var(X) = p - p^2 = p(1-p)$
- $X \sim Bin(n,p)$
  $Var(X) = np(1-p)$ (vedi Var per somma di indip.)



- $X \sim G(p)$
  $Var(X) = \frac{1-p}{p^2}$
- $X \sim Poisson(\lambda)$
  $Var(X) = \lambda$

Prop (disuguaglianza di Chebyshev):
Sia X v.a. reale con $E[|X|^2] < \infty$. Allora, $\forall a > 0$,
$$P\{|X - E[X]| \geq a\} \leq \frac{1}{a^2} Var(X)$$
In particolare, se $Var(X) = 0$, allora $X =$ costante $(= E[X])$ q.c.

Dim: segue da Cor di dis. di Markov, applicato a $X - E[X]$

Oss: • $E[X]$ minimizza $\mathbb{R} \ni m \mapsto E[(X-m)^2]$; infatti $E[(X-m)^2] = Var(X) + (E[X]-m)^2$
• ogni mediana minimizza $\mathbb{R} \ni m \mapsto E[|X-m|]$ (esercizio)

Richiami su serie numeriche:

- Sia $(a_n)_{n \in \mathbb{N}^+}$ una successione con $a_n \geq 0$ $\forall n$, allora $\exists \sum_{n=1}^{\infty} a_n \in [0, +\infty]$
- Sia $(a_n)_{n \in \mathbb{N}^+}$ una successione con $a_n \geq 0$ $\forall n$ oppure $\sum_{n=1}^{\infty} |a_n| < \infty$
  - Se $v: \mathbb{N}^+ \to \mathbb{N}^+$ è biunivoca, allora $\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} a_{v(n)}$
  - Se $I_1, I_2, \dots$ è una partizione di $\mathbb{N}^+$ (finita o infinita, con I_j finiti o infiniti), allora
    $$\sum_{n=1}^{\infty} a_n = \sum_j \sum_{n \in I_j} a_n$$

9

domenica 12 marzo 2023 17:16

Indipendenza, covarianza e correlazione

Oss: Date $X:\Omega\to S_1$, $Y:\Omega\to S_2$ v.a. con densità congiunta $p_{(X,Y)}$, data $\varphi: S_1\times S_2\to\mathbb{R}$ con $\varphi(X,Y)\geq 0$ q.c. o $\varphi(X,Y)$ integrabile, vale

$$E[\varphi(X,Y)] = \sum_{(x,y)\in R_{p_{(X,Y)}}} \varphi(x,y)\, p_{(X,Y)}(x,y) \qquad (\text{con } R_{p_{(X,Y)}} = \{(x,y)\in S_1\times S_2 \mid p_{(X,Y)}(x,y)>0\})$$

Lemma (dis. di Schwarz): Date X, Y v.a. reali, se $E[|X|^2]<\infty$ e $E[|Y|^2]<\infty$, allora

$$E[|XY|] \leq E[X^2]^{1/2} E[Y^2]^{1/2}$$

Dim:

$$0\leq E[(X-tY)^2] = E[X^2]+t^2E[Y^2]-2tE[XY] \qquad \forall t\in\mathbb{R},$$

quindi $\frac{\Delta}{4} = E[XY]^2 - E[X^2]E[Y^2]\leq 0$

Prop: Siano X, Y v.a reali indipendenti. Se X, Y sono entrambe integrabili (o entrambe ≥ 0), allora XY è integrabile (o ≥ 0) e

$$E[XY] = E[X]E[Y]$$

Dim:

I caso: $X,Y\geq 0$. Allora, usando $E[\varphi(X,Y)] = \sum_{i,j}\varphi(x_i,y_j)\,p_{(X,Y)}(x_i,y_j)$

$$E[XY] = \sum_{i,j} x_i y_j\, p_{(X,Y)}(x_i,y_j) \underset{\text{indip: } p_{(X,Y)}(x,y) = p_X(x)p_Y(y)}{=} \sum_{i,j} x_i y_j\, p_X(x_i)\,p_Y(y_j) = \sum_i x_i\, p_X(x_i)\cdot\sum_j y_j\, p_Y(y_j) = E[X]E[Y]$$

II caso: X, Y integrabili: $E[|XY|] = E[|X|]E[|Y|]<\infty$ e l'uguaglianza $E[XY]=E[X]E[Y]$ segue come sopra

Oss: Non vale il viceversa (vedi esempi sotto)

Cor: Siano $X:\Omega\to S_1$, $Y:\Omega\to S_2$ v.a. Allora

X, Y sono indipendenti $\iff E[g(X)h(Y)] = E[g(X)]E[h(Y)]$ $\forall g: S_1\to\mathbb{R}$, $h:S_2\to\mathbb{R}$ con $g(X), h(Y)$ entrambe ≥ 0 o entrambe integrabili

Dim:

$\Rightarrow$: Se X e Y sono indip., $g(X)$ e $h(Y)$ lo sono e si applica prop precedente.

$\Leftarrow$: $\forall A_1\subseteq S_1$, $A_2\subseteq S_2$, prendendo $g=\mathbb{1}_{A_1}$, $h=\mathbb{1}_{A_2}$,

$$P\{X\in A_1, Y\in A_2\} = E[\mathbb{1}_{A_1\times A_2}(X,Y)] = E[\mathbb{1}_{A_1}(X)\mathbb{1}_{A_2}(Y)]$$
$$= E[\mathbb{1}_{A_1}(X)]E[\mathbb{1}_{A_2}(Y)] = P\{X\in A_1\}\cdot P\{Y\in A_2\}$$

Più in generale vale

Lemma: Siano $(X_i)_{i\in I}$, $X_i : \Omega \to S_i$, famiglia di v.a.

$(X_i)_{i\in I}$ è famiglia di v.a. indipendenti

$\iff \forall J \subseteq I$ finito, $\forall f_i : S_i \to \mathbb{R}$ con $f_i(X_i)$ integrabile, $i \in J$, vale $E[\prod_{i\in J} f_i(X_i)] = \prod_{i\in J} E[f_i(X_i)]$

Def: Siano X, Y v.a. con $E[|X|^2]<\infty$, $E[|Y|^2]<\infty$. Covarianza di X, Y: il numero

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

(ben definito per dis di Schwarz)

Proprietà di Cov:

- $Cov(X,X) = Var(X)$
- $Cov(X,Y) = E[XY] - E[X]E[Y]$ (esercizio)
- Cov è bilineare simmetrica:
  - $Cov(aX_1+X_2, Y) = a\,Cov(X_1,Y) + Cov(X_2,Y)$, X_1, X_2, Y v.a. con momento secondo assoluto finito, $a \in \mathbb{R}$
  - $Cov(X,Y) = Cov(Y,X)$

Cor: Se X, Y sono indipendenti, allora $Cov(X,Y)=0$

Il viceversa non vale:

Esempio: X uniforme su $\{-1,0,1\}$, $Y = X^2$

- X, Y non indip: $P\{Y=1 | X=1\} = 1 \neq P\{Y=1\}$
- $E[X]=0$, $XY = X$ quindi $E[XY]=0$, quindi $Cov(X,Y)=0$

Oss. importante: $Cov(X,Y)$ dipende solo dalla legge congiunta di (X,Y), ma non è individuata univocamente dalle leggi marginali.

Prop: Date $X_1, \dots X_n$ v.a. reali con $E[|X_i|^2]<\infty$ $\forall i$,

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

In particolare, se X_i sono indipendenti, o anche solo scorrelate a due a due,

$$Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i)$$

Dim:

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

Esempio:

$X \sim B(n,p)$, $X = X_1 + \dots + X_n$, $X_i \sim B(p)$ indipendenti (senza perdita di generalità)

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

Oss: La proprietà di additività della varianza per v.a. indipendenti cattura le cancellazioni nella somma dovute all'indipendenza e ci dice che, in media, "le cose vanno meglio rispetto al caso deterministico"

Esempio: passeggiata aleatoria simmetrica

$X_i \sim B(\frac{1}{2})$, $Y_i = 2X_i - 1 = \begin{cases} 1 & \text{con prob } 1/2 \\ -1 & \text{" " } 1/2 \end{cases}$ (Rademacher), Y_i indip.

$S_n = \sum_{i=1}^n Y_i$ "somma di 1 o -1" $E[S_n] = 0$, $Var(S_n) = \sum_{i=1}^n Var(Y_i) = n$

- stima deterministica (caso peggiore) $|S_n| \leq \sum_{i=1}^n |Y_i| = n$
- stima probabilistica $E[|S_n|] \leq E[|S_n|^2]^{1/2} = \sqrt{Var(S_n)} = \sqrt{n}$

Def: Siano X, Y v.a. reali con $E[|X|^2] < \infty$, $E[|Y|^2] < \infty$ e $Var(X) > 0$, $Var(Y) > 0$.

Coefficiente di correlazione tra X e Y: il numero

$$\rho(X,Y) := \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)} \qquad (\text{con } \sigma(X) = \sqrt{Var(X)},\ \sigma(Y) = \sqrt{Var(Y)})$$

Proprietà di ρ:

- $|\rho| \leq 1$:

  infatti per dis. di Schwarz: $Cov(X,Y) = E[(X - E[X])(Y - E[Y])] \leq Var(X)^{1/2} Var(Y)^{1/2}$

- $\rho(aX+b, cY+d) = \rho(X,Y)$ $\forall a,b,c,d \in \mathbb{R}$ con $a, c \neq 0$

  ("ρ non dipende da unità di misura")

  dim: esercizio

Prop: Siano X, Y v.a. reali con $E[|X|^2], E[|Y|^2] < \infty$, $Var(X), Var(Y) > 0$.

Allora la funzione

$$\mathbb{R}^2 \ni (a,b) \mapsto E[(Y - (aX+b))^2] \in \mathbb{R}$$

ammette un unico pt di minimo (a^*, b^*), che vale

$$a^* = \frac{Cov(X,Y)}{Var(X)} \qquad b^* = E[Y] - a^* E[X] \qquad (\Delta)$$

Inoltre il valore del minimo è

$$E[(Y - (a^* X + b^*))^2] = Var(Y)(1 - \rho(X,Y)^2) \qquad (\diamond)$$

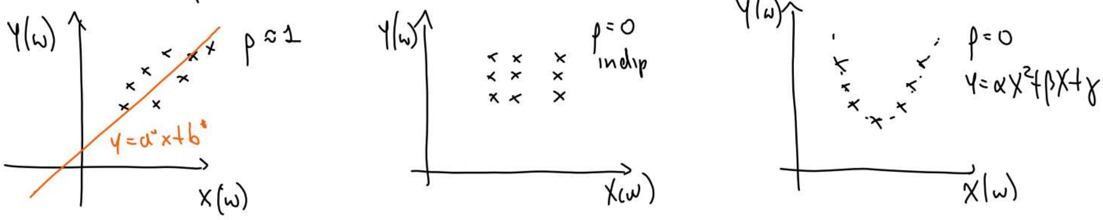
Def: La retta $y = a^* x + b^*$ è detta retta di regressione tra X e Y.

Significato:

- la retta di regressione è la "migliore" approssimazione lineare tra X e Y, nel senso che minimizza la media della distanza al quadrato tra Y e $aX+b$
- poiché il valore del minimo è proporzionale a $1-\rho^2$, $1-\rho^2$ indica quanto la relazione tra X e Y sia approssimabile da una retta, in particolare:
  - $|\rho|$ vicino a 0 indica che non c'è, o c'è debole relazione lineare
  - $|\rho|$ " " 1 " " c'è una forte relazione lineare
  - $|\rho| = 1 \Rightarrow Y = a^* X + b^*$ q.c.
- il segno di a^* coincide con il segno di ρ (e con il segno di Cov)

Oss: Il fatto che indip. $\Rightarrow$ non correlate e non corr. $\not\Rightarrow$ indipendenza corrisponde a:

- se due v.a. sono indip., allora non hanno alcuna relazione di tipo lineare;
- se due v.a. non hanno relazione lineare, potrebbero comunque avere qualche altra relazione non lineare ed essere quindi dipendenti.



Dim, sketch:

- la funzione $F(a,b) = E[(Y-(aX+b))^2]$ è C^1 (è un polinomio) e soddisfa
$$\lim_{|(a,b)| \to +\infty} F(a,b) = +\infty$$
quindi esiste almeno un pt di minimo, che verifica $\nabla F = 0$
- esiste un solo punto (a^*, b^*) t.c. $\nabla F(a^*, b^*) = 0$ ed è dato da (Δ)
in particolare, (a^*, b^*) è l'unico punto di minimo
- inserendo le espressioni in (Δ) per a^*, b^*, si trova $(\Diamond)$

Oss: Come abbiamo visto,

- se un problema coinvolge una sola v.a., tutte le quantità di interesse ($P(X \in A)$, $E[X]$, $Var(X)$ ..) dipendono solo della legge di X
- se un problema coinvolge più v.a. $X_1, \ldots X_n$, tutte le quantità di interesse dipendono solo della legge congiunta di $X_1, \ldots X_n$ (e, se le v.a. sono indip, solo delle leggi marginali)

Per questo, spesso si dà solo la legge di X, o la legge congiunta di $X_1, \ldots X_n$, senza specificare lo spazio Ω dove X o $X_1 \ldots X_n$ sono definite

L'esistenza di $(\Omega, \mathcal{P}(\Omega), P)$ e $(X_1, .. X_n) : \Omega \to S_1 \times .. \times S_n$ con la data legge congiunte è garantita dalla costruzione canonica

(nel caso $X_1, .. X_n$ indip di leggi $P_{X_1}, .. P_{X_n}$ date, la legge congiunta è la prob prodotto $P_{(X_1, .. X_n)} = P_{X_1} \otimes .. \otimes P_{X_n}$)

10
domenica 19 marzo 2023 11:49

Teoremi limite (LGN, TCL)

Esempio/motivazione:

1000 lanci di una moneta equilibrata, che cosa possiamo dire del n° di teste?

- legge dei grandi numeri (LGN): "frequenza relativa di "testa" su 1000 lanci è circa $\frac{1}{2}$"
  (dove freq relativa $= \frac{\# \text{teste}}{1000}$)
- teorema centrale del limite (TCL): "la distribuzione delle oscillazioni del n° di teste attorno al valor medio 500 è circa una gaussiana"

Oss: • queste affermazioni valgono per un grande numero di esperimenti ripetuti
$\rightsquigarrow$ la probabilità predice il comportamento statistico di un esperimento aleatorio
• come vedremo, il comportamento statistico della frequenza relativa e delle sue oscillazioni è lo stesso per un'ampia classe di distribuzioni
$\rightsquigarrow$ universalità

$(\Omega, \mathcal{P}(\Omega), P)$ sp. di probabilità discreto

Def: Data una sequenza (finita o infinita) di v.a. $X_1, \dots, X_n, \dots$, con $X_i : \Omega \to S$, X_i si dicono indipendenti e identicamente distribuite (i.i.d.) se sono indipendenti e hanno la stessa legge su S (cioè $P_{X_i} = P_{X_j}$ $\forall i, j$).

Esempio: n ripetizioni di un esperimento:

Consideriamo una v.a. X che rappresenta (una caratteristica di) un esperimento
(ad es. $X = \begin{cases} 1 & \text{se esce } 5 \\ 0 & \text{altrimenti} \end{cases}$ nel lancio di un dado equilibrato)

Ripetiamo l'esperimento n volte, sia X_i la v.a. che rappresenta l'i-sima ripetizione
(ad es., $X_i = \begin{cases} 1 & \text{se esce } 5 \text{ all' i-simo lancio} \\ 0 & \text{altrimenti} \end{cases}$)

Allora le X_i sono
- indipendenti, poiché prove distinte sono indipendenti
- con la stessa legge, che è la legge di X ($P_{X_i} = P_X$), poiché sono ripetizioni dello stesso esperimento

Precisamente, richiamandoci al modello della prob. prodotto:
- chiamiamo $(\Omega_0, \mathcal{P}(\Omega_0), P_0)$ lo sp. di prob. (discreto) del singolo esperimento,
  $X : \Omega_0 \to S$ una caratteristica dell'esperimento
- le n ripetizioni sono modellizzate da $(\Omega_{(n)} = \Omega_0^n, \mathcal{P}(\Omega_n), P_{(n)} = P_0^{\otimes n})$, le proiezioni canoniche $\pi_i : \Omega_0^n \to \Omega_0$ (l'esito dell' i-simo esperimento) sono indip. e con legge P_0

· detta $X_i = X \circ \pi_i : \Omega_0^n \to S$ la caratteristica dell'i-simo esperimento, $i=1,\dots,n$, le X_i:
- sono indip, poiché funzione di v.a. indip π_i
- hanno la stessa legge di X: $P_{(n)}\{X_i \in A\} = P_{(n)}\{\pi_i \in X^{-1}(A)\} = P_{(0)}(X^{-1}(A)) = P_{(0)}\{X \in A\}$

In statistica, $(X_1, \dots X_n)$ è detto campione (i.i.d.) di taglia n di X

Esempio: n estrazioni con reinserimento da una popolazione Ω_0 (n ripetizioni dell'esperimento di estrazione)
$X: \Omega_0 \to S$ caratteristica, X_i = caratteristica X per l'i-simo individuo estratto
$(\Omega_{(n)} = \Omega_0^n,\ P_{(n)} = P_0^{\otimes n}$ = uniforme, $X_i = X \circ \pi_i)$
$(X_1, \dots X_n)$ è un campione della popolazione (o meglio della caratteristica X della popolazione)

Oss: Abbiamo visto come costruire, tramite la prob prodotto, n v.a. indip e di legge data.
Si può estendere tale costruzione al caso di una successione (numerabile) di v.a. indip e di legge data (estensione non banale, Ω più che numerabile, non lo dimostriamo qui).

Def: Data una successione di v.a. reali $Y_n: \Omega \to \mathbb{R}$ e data $Y: \Omega \to \mathbb{R}$ v.a. reale, diciamo che $(Y_n)_n$ converge in probabilità a Y ($Y_n \xrightarrow{P} Y$) se
$$\lim_n P\{|Y_n - Y| > \varepsilon\} = 0 \quad \forall \varepsilon > 0$$

Oss: Se $Y = c$ costante q.c. allora la convergenza sopra dipende solo dalla legge di Y_n:
$$P\{|Y_n - c| > \varepsilon\} = P^{Y_n}([c-\varepsilon, c+\varepsilon]^c)$$
in particolare Y_n può essere definita su Ω_n dipendente da n.

Date $X_1, \dots X_n$ v.a., chiamiamo media campionaria $\bar{X}_n$ la loro media aritmetica:
$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$
(se $(X_1, \dots X_n)$ è un campione i.i.d., $\bar{X}_n$ è la media sul campione)

Teor (Legge (debole) dei grandi numeri, LGN):
Sia $X_1, \dots X_n, \dots$ una successione di v.a. i.i.d. dotate di momento secondo ($E[|X_1|^2] < \infty$), sia $m = E[X_1]$. Allora
$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} m \qquad \text{per } n \to \infty$$
cioè $\lim_n P\{|\bar{X}_n - m| > \varepsilon\} = 0 \quad \forall \varepsilon > 0$.

Dim:
$\bar{X}_n$ ha momento secondo finito (poiché combinazione lineare di v.a. con momento secondo finito)

$$E[\overline{X}_n] = \frac{1}{n}\sum_{i=1}^{n} \underset{\substack{\| \\ m \text{ per stessa legge}}}{E[X_i]} = m$$

$$Var(\overline{X}_n) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) \underset{\substack{\uparrow \\ \text{per indip.}}}{=} \frac{1}{n^2}\sum_{i=1}^{n} \underset{\substack{\| \\ Var(X_1) =: \sigma^2 \text{ per stessa legge}}}{Var(X_i)} = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n} \to 0$$

Per Chebyshev, $\forall \varepsilon > 0$

$$P\{|\overline{X}_n - m| > \varepsilon\} = P\{|\overline{X}_n - E[\overline{X}_n]| > \varepsilon\} \leq \frac{Var(\overline{X}_n)}{\varepsilon^2} = \frac{1}{\varepsilon^2}\frac{\sigma^2}{n} \to 0.$$

Oss: L'ipotesi di indipendenza si può sostituire con quella, più debole, che le X_i siano a due a due scorrelate.

Somme di v.a. a due a due scorrelate e a media nulla sono martingale.

Oss: Esiste una LGN forte, in cui le v.a. hanno solo momento primo e la nozione di convergenza è più forte.

Esempio:

Gioco del lotto: giocando 1€ su 2 numeri, ottengo 250€ con ambo, zero altrimenti

Se gioco tante volte, stima della vincita netta?

X = vincita netta in una giocata, $X = \begin{cases} 249 & \text{con prob } 1/400.5 =: p \\ -1 & \text{" " } 1-p \end{cases}$

$E[X] = 249 \cdot p - 1 \cdot (1-p) =: \mu \approx -0.376 < 0$

X_i = vincita i-sima giocata

X_i sono i.i.d con la stessa legge di X (ripetizioni del gioco del lotto)

$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ media delle n vincite nette

Per LGN, $\overline{X}_n \to \mu < 0$, quindi

vincita netta totale $= n\overline{X}_n \approx n\mu(1+o(1)) < 0$

non conviene giocare!

(anche con strategie più raffinate, si va di solito in perdita)

Esempio importante: LGN per frequenza empirica / binomiale

Consideriamo n ripetizioni di un esperimento, sia A un evento relativo a tale esperimento. Allora la frequenza relativa di A nelle n prove (freq. relativa campionaria) tende alla probabilità di A.

Infatti, detta $X_i = \begin{cases} 1 & \text{se } A \text{ all'i-sima ripetizione} \\ 0 & \text{altrimenti} \end{cases}$ $(\Omega_{(n)} = \Omega_0^n,\ P = P_0^{\otimes n},\ X_i = 1\!\!1_A \circ \pi_i)$

le X_i sono $B(p)$ indipendenti (sequenza di Bernoulli), con $p = P_0(A)$, quindi per LGN:

frequenza relativa di $A = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} E[X_1] = p = P_0(A)$

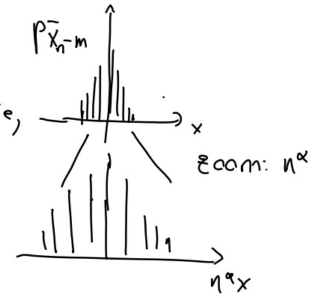
Notiamo che la freq assoluta $Y_n = X_1 + \dots + X_n$ è $\sim B(n,p)$, quindi la LGN dà

$$\frac{Y_n}{n} \xrightarrow{P} p \quad \text{per } Y_n \sim B(n,p)$$

Ad es, in n lanci di moneta equa, $\bar{X}_n$ = freq. relativa di "testa" $\xrightarrow{P} \frac{1}{2}$ per $n \to \infty$

in n lanci di dado equo, $\bar{X}_n$ = freq relativa di "5" $\to \frac{1}{6}$ per $n \to \infty$

"Zoom in" / "scaling"



Sappiamo che $\bar{X}_n - m \xrightarrow{P} 0$, cerchiamo una scala n^α, $\alpha > 0$, se esiste, t.c. $n^\alpha(\bar{X}_n - m)$ tenda a un limite non banale:

$$Var(n^\alpha(\bar{X}_n - m)) = n^{2\alpha} Var(\bar{X}_n) = n^{2\alpha - 1}\sigma^2 \quad \text{con } \sigma^2 = Var(X_1)$$

Per Chebyshev

$$P\{n^\alpha |\bar{X}_n - m| > \varepsilon\} \le \frac{1}{\varepsilon^2} n^{2\alpha-1}\sigma^2$$

Quindi per $\alpha < \frac{1}{2}$, $n^\alpha(\bar{X}_n - m) \xrightarrow{P} 0$

Per $\alpha = \frac{1}{2}$ ci si può attendere un limite non banale $\rightsquigarrow$ TCL

Teor. (teorema centrale del limite, TCL):

Sia $X_1, \dots X_n, \dots$ una successione di v.a. i.i.d dotate di momento secondo ($E[X_1^2] < \infty$), e non costanti q.c., chiamiamo $m = E[X_i]$, $\sigma^2 = Var(X_i)$ ($0 < \sigma^2 < \infty$ per ipotesi).

Allora, $\forall\ -\infty \le a < b \le +\infty$,

$$P\left\{a \le \frac{\sum_{i=1}^n X_i - nm}{\sqrt{n}\,\sigma} \le b\right\} \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \quad \text{per } n \to +\infty$$

$$\boxed{\sigma_n = \frac{\sigma}{\sqrt{n}}}$$

equivalentemente, $P\{a \le \frac{\sqrt{n}}{\sigma}(\bar{X}_n - m) \le b\} \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ per $n \to +\infty$.

Oss: Il limite non dipende dalla legge delle X_i! (purché valga $0 < \sigma^2 < \infty$) (universalità)

Oss: $\dfrac{\sum_{i=1}^n X_i - nm}{\sqrt{n}\sigma} = \dfrac{n\bar{X}_n - nm}{\sqrt{n}\sigma} = \dfrac{\sqrt{n}}{\sigma}(\bar{X}_n - m)$ (da cui l'equivalenza delle due formulazioni)

inoltre $E\left[\frac{\sqrt{n}}{\sigma}(\bar{X}_n - m)\right] = 0$

$$Var\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - m)\right) = \frac{n}{\sigma^2} Var(\bar{X}_n) = 1$$

$\sigma = \sqrt{Var}$ $\quad \sqrt{Var(\bar{X}_n)}$

come normalizzare $\quad E[X] = 0$, $Var(X) = 1$

Esempio importante: TCL per frequenza empirica / binomiale (teor. di De Moivre-Laplace)

Sia $Y_n \sim B(n,p)$, ad es Y_n = # successi in n ripetizioni di Bernoulli con prob di successo p,

Allora, $\forall\ -\infty \le a < b \le +\infty$,

$$P\{np + a\sqrt{np(1-p)} \le Y_n \le np + b\sqrt{np(1-p)}\} \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx \quad \text{per } n\to+\infty.$$

Infatti $Y_n \overset{(d)}{=} X_1 + \dots + X_n$ con $X_i \sim B(p)$ indipendenti e si applica il TCL.

L'integrale $\Phi(b) = \int_{-\infty}^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx$ non ha una forma chiusa in termini di funzioni elementari, ma si calcola in modo numerico, con ottima approssimazione, per molti valori di b, e quindi $\int_a^b = \Phi(b) - \Phi(a)$.

In particolare, si ha che $\left(\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx = 1 \text{ e}\right)$

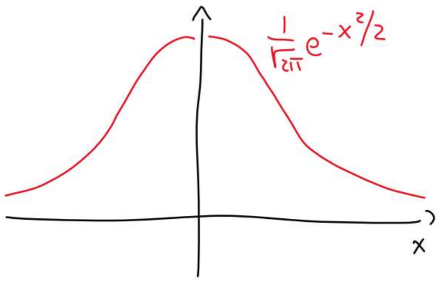
$$\int_{-3}^{3} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx \approx 0.997$$

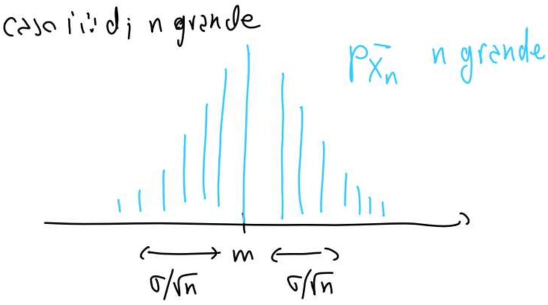
e dunque, per n grande, la prob che $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - m) \in [-3,3]$, cioè

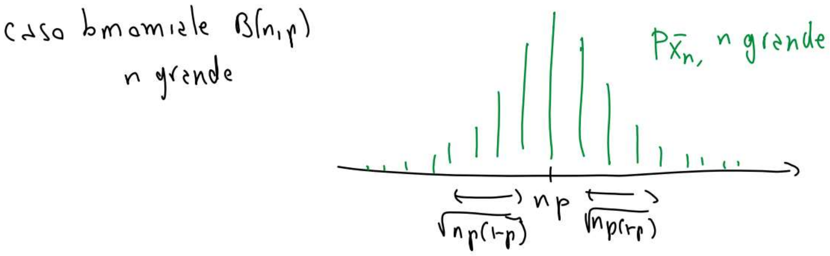
$\bar{X}_n \in \left[m - \frac{3\sigma}{\sqrt{n}},\ m + \frac{3\sigma}{\sqrt{n}}\right]$ (o equiv. $n\bar{X}_n \in [mn - 3\sigma\sqrt{n},\ mn + 3\sigma\sqrt{n}]$)

è ≈ 0.997.

Esempio: su 1000 lanci di moneta equa ($m = p = \frac{1}{2}$, $\sigma^2 = p(1-p) = \frac{1}{4}$, $n = 1000$), la freq. assoluta di "testa" è, con prob ≈ 0.997, compresa tra 452 e 548







Disuguaglianze di concentrazione: tasso di convergenza per $P\{|\bar{X}_n - m| > \varepsilon\}$, $X_1 \dots X_n$ i.i.d.

"quanto rapidamente $P\{|\bar{X}_n - m| \ge \varepsilon\}$ va a 0?"

Abbiamo visto, nella dim. della LGN: $(\sigma^2 = Var(X_1))$

$$P\{|\bar{X}_n - m| \ge \varepsilon\} \le \frac{1}{\varepsilon^2} Var(\bar{X}_n) = \frac{1}{\varepsilon^2}\frac{\sigma^2}{n}$$

dove abbiamo usato

a) monotonia di $\varphi(t) = t^2$ (per $t \ge 0$) e dis. di Markov

$\{|\bar{X}_n - m| \ge \varepsilon\} = \{|\bar{X}_n - m|^2 \ge \varepsilon^2\}$ e Markov

b) additività di $Var(\cdot)$ per v.a. indipendenti

Un'altra classe di "operatori" che soddisfa (a) e (b) è legata agli esponenziali:

a) $P\{\bar{X}_n - m \geq \varepsilon\} = P\{\sum_{i=1}^n (X_i - m) \geq n\varepsilon\} \leq e^{-\lambda n \varepsilon} E[e^{\lambda \sum_{i=1}^n (X_i - m)}]$

b) $E[\exp(\lambda \sum_{i=1}^n (X_i - m))] = \prod_{i=1}^n E[e^{\lambda(X_i - m)}] = E[e^{\lambda(X_1 - m)}]^n$

equivalentemente, $\log E[e^{\lambda \cdot}]$ è additivo per v.a. indip.

Quindi

$$P\{\bar{X}_n - m \geq \varepsilon\} \leq \exp[-n(\lambda\varepsilon - \log E[e^{\lambda(X_1 - m)}])]$$

Ottimizzando in λ,

$$P\{\bar{X}_n - m \geq \varepsilon\} \leq \exp[-n \sup_{\lambda > 0}(\lambda\varepsilon - \log E[e^{\lambda(X_1 - m)}])] = e^{-n I(\varepsilon)}$$

dove $I(t) = \sup_{\lambda > 0}(\lambda t - \log E[e^{\lambda(X_1 - m)}])$ (trasformata di Cramér)

(legata alla trasformata di Legendre)

Se $\exists \lambda > 0$ t.c. $E[e^{\lambda |X_1|}] < \infty$ (momento esponenziale finito), allora $I(t) > 0 \quad \forall t > 0$

e quindi $P\{\bar{X}_n - m > \varepsilon\}$ tende a 0 esponenzialmente in n

(e analogamente $P\{\bar{X}_n - m < -\varepsilon\}$)

Caso Bernoulli ($X_1, \dots X_n$ v.i.i.d $\sim B(p)$):

$$\log E[e^{\lambda(X_1 - p)}] = \log E[e^{\lambda X_1}] - \lambda p = \log(pe^\lambda + 1 - p) - \lambda p$$

$$g(\lambda) = \lambda t - \log(pe^\lambda + 1 - p) + \lambda p$$

- per $t > 1-p$, $\lim_{\lambda \to +\infty} g(\lambda) = +\infty$, in particolare $\sup_{\lambda > 0} g(\lambda) = +\infty$
- per $t = 1-p$, g è crescente, $\lim_{\lambda \to +\infty} g(\lambda) = \sup_{\lambda \to +\infty} g(\lambda) = \log(1/p)$
- per $0 < t < 1-p$, $g(0) = 0$, $\lim_{\lambda \to +\infty} g(\lambda) = -\infty$, $g'(\lambda) = 0 \iff \lambda = \lambda_t := \log \frac{(p+t)(1-p)}{p(1-p-t)}$

  in particolare λ_t è punto di massimo e $g(\lambda_t) = h_p(p+t)$

con $h_p(a) = (1-a)\log\frac{1-a}{1-p} + a\log\frac{a}{p}$ (= entropia relativa di $B(a)$ rispetto a $B(p)$)

quindi $P\{\bar{X}_n - p \geq \varepsilon\} \leq e^{-n h_p(p+\varepsilon)} \quad \forall\ 0 < \varepsilon < 1-p$

e analog. $P\{\bar{X}_n - p \leq -\varepsilon\} \leq e^{-n h_{1-p}(1-p+\varepsilon)} \quad \forall 0 < \varepsilon < p$

da cui $P\{|\bar{X}_n - p| \geq \varepsilon\} \leq 2\exp(-n \cdot \min\{h_p(p+\varepsilon), h_{1-p}(1-p+\varepsilon)\}) \quad \forall\ 0 < \varepsilon < \min\{p, 1-p\}$

11

mercoledì 29 marzo 2023 11:58

Leggi condizionali

Esempio/motivazione:

- n° di clienti in un negozio segue distrib. di Poisson di parametro $\lambda > 0$
- ogni cliente acquista almeno un prodotto con prob. $p \in (0,1)$, indipendentemente dagli altri
- qual è la distrib. del n° di clienti che acquista almeno un prodotto?

Che info ci dà il problema:

- $N = \#$ clienti $\sim Poisson(\lambda)$
- $X_i = \mathbb{1}_{i\text{-simo cliente compra qualcosa}} \sim B(p)$, $i \in \mathbb{N}^+$, $N, X_1, X_2, \dots$ indipendenti
- $M = \#$ clienti che acquistano qualcosa $= \sum_{i=1}^{N} X_i$ $\left(M(\omega) = \sum_{i=1}^{N(\omega)} X_i(\omega) \right)$

Se $N = n$, allora $M \sim B(n,p)$, cioè
la legge di M sotto la prob $P(\cdot \mid N=n)$ è $B(n,p)$.

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

Def: Date $X: \Omega \to S_1$ v.a., $Y: \Omega \to S_2$ v.a., $x \in R_{p_X}$ (cioè $x \in S_1$, $P\{X=x\} > 0$), legge condizionale di Y dato $X=x$,
$P_Y(\cdot \mid X=x)$ legge di Y sotto $P(\cdot \mid X=x)$, cioè
$P_Y(\cdot \mid X=x)$ probabilità su $(S_2, \mathcal{P}(S_2))$ (o su $(S_{2,Y}, \mathcal{P}(S_{2,Y}))$ con $S_{2,Y} = Y(\Omega)$) con
$P_Y(A \mid X=x) = P(Y \in A \mid X=x) \quad \forall A \in \mathcal{P}(S_2)$

$P_Y(\cdot \mid X=x)$ ha densità discreta

$$p_{Y|X}(y|x) = P\{Y=y \mid X=x\} = \frac{P\{X=x, Y=y\}}{P\{X=x\}} = \frac{p_{(X,Y)}(x,y)}{p_X(x)}$$

con p_X, $p_{(X,Y)}$ densità discrete rispettivamente di X e (X,Y)

In particolare, $P_Y(\cdot \mid X=x)$ dipende solo dalla legge congiunta di (X,Y).

(Si estende $p_{Y|X}(y|x)$ a $x \in S_1$ ponendo $p_{Y|X}(y|x) = 0 \quad \forall x \in S_1 \setminus R_{p_X}$)

Prop (della legge condizionata alla legge)

$$P_Y(A) = \sum_{x \in R_{p_X}} P_Y(A \mid X=x)\, P_X(x) \quad \forall A \in \mathcal{P}(S_2)$$

Dim: Dalla formula della partizione applicata alla partizione $\{X=x\}_{x \in R_{p_X}}$:

$$P_Y(A) = P\{Y \in A\} = \sum_{x \in R_{p_X}} P\{Y \in A \mid X=x\}\, P\{X=x\}$$

Oss: In realtà $\{X=x\}_{x \in R_{p_X}}$ non è una partizione di Ω, poiché $\{X \notin R_{p_X}\}$ può essere non vuoto, benché di misura P nulla. Però ci possiamo restringere a $\Omega_0 = \{X \in R_{p_X}\}$, che ha $P(\Omega_0) = 1$:
$P|_{\mathcal{P}(\Omega_0)}$ è una prob su $(\Omega_0, \mathcal{P}(\Omega_0))$ e qui applichiamo la formula della partizione

Quindi, nell'esempio precedente, si può ricavare la legge di M data $P_M(\cdot|N=n) = B(n,p)$

$$p_M(k) = \sum_{n=0}^{\infty} p_{M|N}(k|n)\, p_N(n) = \sum_{n=0}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} 1\!\!1_{k\le n} \frac{\lambda^n}{n!} e^{-\lambda} = \sum_{n=k}^{\infty} \dots$$

Def: Date X, Y come sopra, $\varphi: S_2 \to \mathbb{R}$, se $\varphi(Y)$ è ≥ 0 o integrabile,

valore atteso condizionale di $\varphi(Y)$ dato $X=x$, $x \in R_{p_X}$,

$$E[\varphi(Y)|X=x] = E^{P(\cdot|X=x)}[\varphi(Y)] \quad \text{(valore atteso rispetto alla prob } P(\cdot|X=x))$$

$$\left(= \sum_{\omega\in\Omega} \varphi(Y(\omega))\, P(\{\omega\}|X=x)\right)$$

$$= \sum_{y\in S_2} \varphi(y)\, p_{Y|X}(y|x)$$

Valore atteso condizionale di $\varphi(Y)$ dato X:

funzione di x, valutata in $x = X(\omega)$

$$E[\varphi(Y)|X] = \sum_{x\in R_{p_X}} E[\varphi(Y)|X=x]\, 1\!\!1_{\{X=x\}} = \text{"} E[\varphi(Y)|X=x]\Big|_{x=X} \text{"}$$

Oss: $E[\varphi(Y)|X=x] =: f(x)$ è una funzione di $x \in R_{p_X}$

(estesa a S_1 ponendo $E[\varphi(Y)|X=x] = 0 \quad \forall x \in S_1 \setminus R_{p_X}$)

$E[\varphi(Y)|X] = f(X)$ è una v.a. funzione di X.

Prop (del valore atteso condizionale al valore atteso):

Se $\varphi(Y)$ è ≥ 0 o integrabile, allora

$$E[\varphi(Y)] = E[E[\varphi(Y)|X]] = \sum_{x\in S_1} E[\varphi(Y)|X=x]\, P\{X=x\}$$

Dim:

$$E[\varphi(Y)] = \sum_{y\in S_2} \varphi(y)\, p_Y(y) = \sum_{y\in S_2,\, x\in S_1} \varphi(y)\, \underbrace{p_{(X,Y)}(x,y)}_{= p_{Y|X}(y|x) p_X(x)}$$

$$= \sum_{x\in S_1} \sum_{y\in S_2} \varphi(y)\, p_{Y|X}(y|x)\, p_X(x)$$

$$= \sum_{x\in S_1} E[\varphi(Y)|X=x]\, P\{X=x\}$$

$$= E[E[\varphi(Y)|X]]$$

Ad es,

$$E[M] = \sum_{n=0}^{\infty} E[M|N=n]\, P\{N=n\} \underset{\substack{\uparrow \\ M\sim B(n,p) \text{ sotto } N=n}}{=} \sum_{n=0}^{\infty} np \frac{\lambda^n}{n!} e^{-\lambda} = p\,E[N] = \lambda p$$