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

$A \subseteq B \leftrightarrow$ se accade A, allora necessariamente accade B — $A = \{1\} \subseteq B = \{$dispari$\}$

$A \cap B = \emptyset \leftrightarrow$ non possono accadere sia A sia B — $A = \{$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) $0 \leq P(A) \leq 1 \quad \forall A \in \mathcal{F}$
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
$$1 = P(\Omega) = P(\Omega \cup \emptyset \cup \emptyset \ldots) = P(\Omega) + \sum_{n=2}^{\infty} P(\emptyset), \text{ quindi deve essere } P(\emptyset) = 0$$
• Inoltre, P soddisfa:
iii') $\forall A_1, \ldots A_N \in \mathcal{F}$ 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 $A_{N+1} = A_{N+2} = \ldots = \emptyset$
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)
$P\{$al più due figli$\} = P\{$al più un figlio$\} + P\{$due figli$\}$

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

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

Esempio fondamentale: modello uniforme (o equiprobabile):

$\Omega \neq \phi$ 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)

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

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
($\Omega = \{1, 2 \dots 6\}$ $\mathcal{F} = \mathcal{P}(\Omega)$ 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, ...)
$\Omega = \{$oggetti nell'urna/popolazione/...$\}$ $\mathcal{F} = \mathcal{P}(\Omega)$ P uniforme

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

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 $\mathcal{F}$:

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

$$A \cup B = A \cup B \cup \emptyset \cup \emptyset \cup \ldots \in \mathcal{F}$$
$$A \cap B = A \cap B \cap \Omega \cap \Omega \cap \ldots \in \mathcal{F}$$
$$A \setminus B = A \cap B^c \in \mathcal{F}$$

Proprietà di una probabilità P:

a) $P(\emptyset) = 0$

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

c) $\forall A, B \in \mathcal{F}$, se $A \subseteq B$ allora $P(A) \le P(B)$ 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) $\forall A_1, \ldots A_n \in \mathcal{F}$,

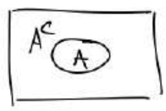
$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \le i < j \le n} P(A_i \cap A_j)$$
$$+ \sum_{1 \le i < j < k \le n} P(A_i \cap A_j \cap A_k) - \ldots \quad \text{(formula di inclusione - esclusione)}$$
$$+ (-1)^{n+1} P(A_1 \cap A_2 \cap \ldots \cap A_n)$$

Dimostrazione:

a) Già visto

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

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

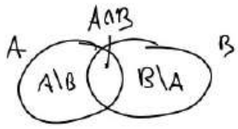


c) se $A \subseteq B$, allora $B = A \dot\cup (B \setminus A)$, quindi

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

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



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

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

Dim:

a) $B_1 = A_1$, $B_n := A_n \setminus A_{n-1}$, $\forall$ 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 $A_{N+1} = A_{N+2} = \dots = \emptyset$) $P(\bigcup_{n=1}^{N} A_n) \leq \sum_{n=1}^{N} P(A_n)$

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 (Ω, A, P), studiarne le proprietà
- " modellistico: dato un problema reale, scegliere il modello probabilistico più adeguato (cioè scrivere le informazioni date dal problema in termini di P)

Sequenze di esperimenti

sabato 14 marzo 2020 13:38

Modellizzare sequenza ordinate (finita o infinita) di esperimenti

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

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

  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 $\in \{5, \dots 10\}\}$

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

- 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)

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

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

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

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

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

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

(- 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

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

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

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: $B \mapsto P(A|B)$ non 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 $P(A_1 \cap \dots \cap A_n) > 0$,

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

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

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

$= \dots$ (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?

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

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

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

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

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

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

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)?

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

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

$$\forall J \subseteq I \text{ finito}(\neq\emptyset),\ P\Big(\bigcap_{i\in J} A_i\Big) = \prod_{i\in J} P(A_i)$$

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 teste se i primi due lanci hanno esito concorde, croce altrimenti

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

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

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

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

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

$A_i = \{$Teste al lancio i-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)

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

4

venerdì 30 dicembre 2022 16:51

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 discreta, 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)

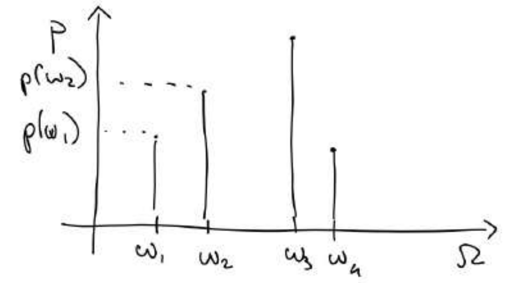
3) σ-additività: $A = \bigcup_{n \in \mathbb{N}} A_n$, A_n a due a due disgiunti

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

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

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

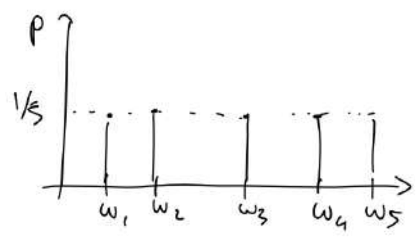
Rappresentazione della densità discreta tramite grafico a barre



(ω_i non necessariamente ordinati)

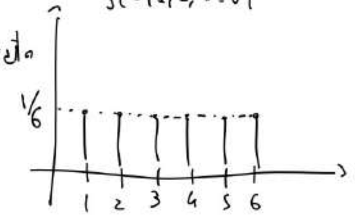
Il grafico a barre indica dove P assegna più "massa"

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



$\Omega=\{1,2,..6\}$

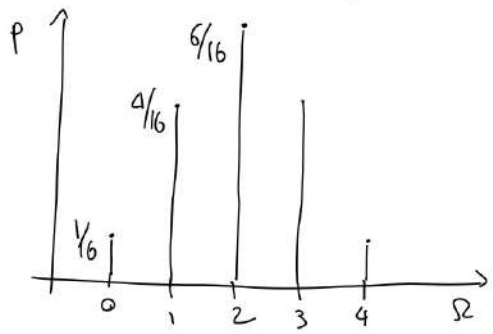
esito lancio dado equilibrato



Esempio: n° teste in 4 lanci di moneta:

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

densità di probabilità $p(\omega)=P\{\omega$ 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: Data $(\Omega, \mathcal{F})$ 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.

$P(\Omega_0) = 1$ $\qquad$ (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

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

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.

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

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

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

  Nel caso infinite prove: $\overline{\Omega}_\infty = \{0,1\}^{\mathbb{N}^+}$ $\overline{\mathcal{F}}_\infty = \sigma\{A_1, A_2, \dots\}$ "la più piccola σ-alg contenente $A_1, A_2, \dots$" (contiene tutti gli eventi rilevanti)

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

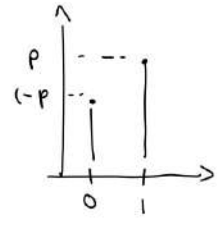
Esempi:

- 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 $p \in [0,1]$ $(B(p))$

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

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

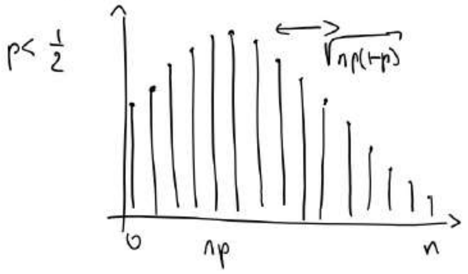


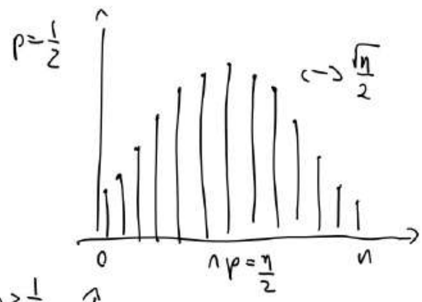
$B(p)$ rappresenta un esperimento con esito successo (1) o insuccesso (0), in cui il successo ha probabilità p (prova di Bernoulli) $(P\{1\} = \overline{P}_n\{\omega \mid \omega_i = 1\}\ \forall i)$

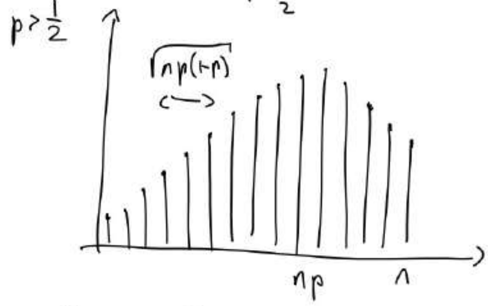
· Distribuzione binomiale di parametri $n \in \mathbb{N}^+$ e $p \in [0,1]$ ($B(n,p)$ o $Bin(n,p)$)

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

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







$B(n,p)$ 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 $\bar{P}_n(\bigcap_I A_i \cap \bigcap_{I^c} A_i^c) = \prod_{i \in I} \bar{P}_n(A_i) \cdot \prod_{i \notin I} \bar{P}_n(A_i^c) = p^k (1-p)^{n-k}$

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

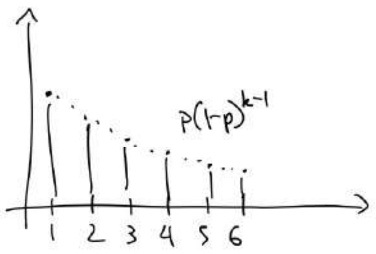
Quindi

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

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

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:

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

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

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

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

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

$BinNeg(h,p)$ 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 $1 \le N_1, n \le N$ $(H(N, N_1, n))$

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

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

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: $\tilde{\Omega} = \{S \subseteq \{1, \dots N\} \mid \# S = n\}$, $\tilde{\mathcal{F}} = \mathcal{P}(\tilde{\Omega})$, $\tilde{P}$ 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) $\cdot$

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

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

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

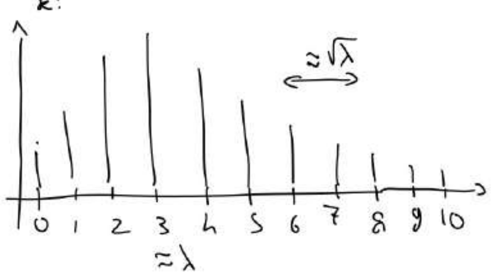
Più in generale, se gli N oggetti di un'urna sono divisi in gruppi $(a_1), \dots (a_m)$, con $N_1, \dots N_m$ elementi rispettivamente $(N_1 + \dots + N_m = N)$, ed estraiamo n oggetti, allora

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

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

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

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



$P(\lambda)$ 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 $p_n(k) = 0 \quad \forall k > n$).
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

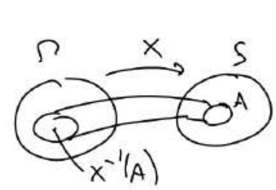
- $X_i := \begin{cases} 1 & \text{se esce T alla i-sima estrazione} \\ 0 & \text{altrimenti} \end{cases}$

  $X_i((\omega_1, \dots \omega_n)) = \begin{cases} 1 & \text{se } \omega_i = T \\ 0 & \text{se } \omega_i = C \end{cases}$

- $S_n =$ 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,

$X^{-1}(\bigcup_{i\in I} A_i) = \bigcup_{i \in I} X^{-1}(A_i)$

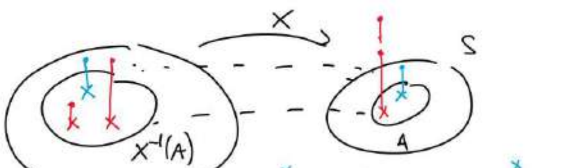
$X^{-1}(A^c) = X^{-1}(A)^c$ (facile verifica per esercizio)

$X^{-1}(\bigcap_{i\in I} A_i) = \bigcap_{i\in I} X^{-1}(A_i)$

Def. Dati Ω spazio discreto, P probabilità su $(\Omega, \mathcal{P}(\Omega))$, $X: \Omega \to S$ v.a., $S_X := X(\Omega) = \{X(\omega) \mid \omega \in S\}$, 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

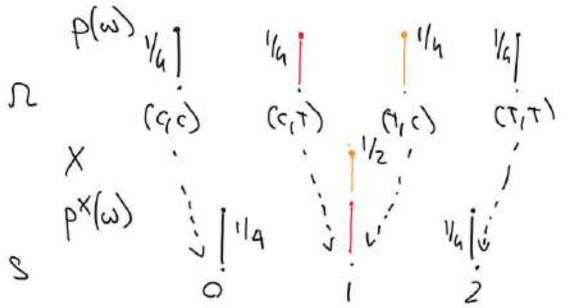
$P^X(A) = P\{X \in A\} = P(X^{-1}(A))$, $A \subseteq S_X$



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



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: dato 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. $1/6$)

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

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

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

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

$\Omega = \{$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

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

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

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

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:

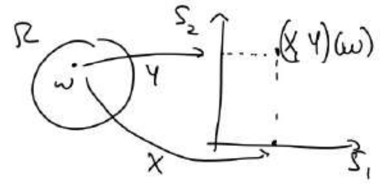
$\Omega = S$, $P = Q$, $X: \Omega \to S$, X = identità $(X(\omega) = \omega \ \forall \omega \in \Omega)$

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

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

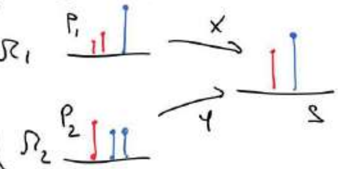


- Se $X:\Omega\to\mathbb{R}$, $Y:\Omega\to\mathbb{R}$ sono v.d. reali, allora
  $X+Y$, XY, ... sono v.d.
  Infatti $X+Y=f(X,Y)$ con $f:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$, $(x,y)\mapsto x+y$
  $XY=f(X,Y)$ $\quad f:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$, $(x,y)\mapsto xy$

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

Def: Date $X:\Omega\to S$, $Y:\Omega\to S$ v.d. 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.d., 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 $R_{p_1^X}=\{x\in S\,|\,p_1^X(x)>0\}$

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

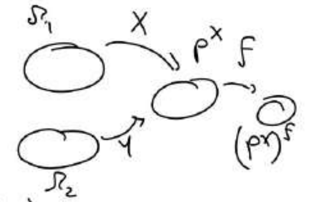
Dim: · Se $X=Y$ q.c., allora, $\forall A\in\mathcal{P}(S)$,

$$
\begin{aligned}
P\{X\in A\}&=P\{X\in A, X=Y\}+\underbrace{P\{X\in A, X\neq Y\}}_{\leq P\{X\neq Y\}=0}\\
&=P\{Y\in A, X=Y\}+\overbrace{P\{Y\in A, X\neq Y\}}\\
&=P\{Y\in A\}
\end{aligned}
$$

Esempio di X,Y v.d. 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:

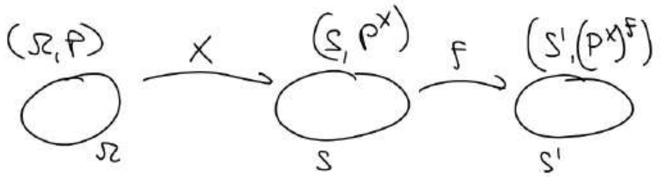
- Se $X:\Omega\to S$ e $Y:\Omega\to S$ sono uguali q.c e $f:S\to S'$, allora $f(X)=f(Y)$ q.c.
- Se $X:\Omega_1\to S$ e $Y:\Omega_2\to S$ hanno la stessa legge e $f:S\to S'$, allora $f(X)$ e $f(Y)$ hanno la stessa legge:



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)

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

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

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

Oss: La legge congiunta contiene informazioni più ricche su $(X_1, \dots X_n)$ 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 $X_1: \Omega \to S_1, \dots X_n: \Omega \to S_n$ v.a., $(X_1, \dots X_n)$ si dice famiglia di v.a. indipendenti se

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

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

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

Dim: $\Leftarrow$: da def.

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

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

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

Invece, $(X_1, \dots X_n)$ indipendenti e $X_i \overset{(d)}{=} Y_i \ \forall i \not\Rightarrow (Y_1, \dots Y_n)$ 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.d. 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: $P\{X_i = x\} = \frac{1}{N} \quad \forall x \in U$
- $X_1, \dots X_n$ sono indipendenti

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

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

- ciascuna Y_i è v.d. uniforme su U $\qquad P\{Y_i = x\} = \frac{1}{N} \quad \forall x \in U$
- $Y_1, \dots Y_n$ 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

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

$\beta \in \mathbb{R}$ 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., $\quad p_{X_J}(x_J) = \prod_{i \in J} p_{X_i}(x_i) \quad \forall J \subseteq I$ 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)

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

(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 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

$\Omega = \{0,1\}^n$ $\mathcal{F} = \mathcal{P}(\Omega)$ $A_i := \{a \in \Omega \mid a_i = 1\}$ (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, $\mathcal{A} = \mathcal{P}(\Omega)$, P prob su $(\Omega, \mathcal{P}(\Omega))$ con densità discreta p.

Def: Sia $X: \Omega \to \mathbb{R}$ v.d. 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.d., $\mathbb{R}_X = X(\Omega)$. 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 $\varphi(X) \geq 0$,

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

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

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 $X(\omega)p(\omega) \leq Y(\omega)p(\omega) \quad \forall \omega \in \Omega$, 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 $X_1, \dots X_n \sim B(p)$ indipendenti, $X := X_1 + \dots + X_n \sim B(n,p)$
  $E[X] = \sum_{i=1}^{n} E[X_i] = np$
  Oss: Non è restrittivo supporre X somma di Bernoulli indip: se $Y \sim B(n,p)$, allora
  $E[Y] = E[X] = np$

- $X \sim \mathcal{G}(p)$

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

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

- $X \sim$ Poisson (λ), $\lambda > 0$

  $X \geq 0$ 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 $X_1, \dots X_n \sim \mathcal{B}(p)$ anche non indipendenti, allora $E[X_1 + \dots + X_n] = np$

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

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

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

$X = \sum_{i=1}^{N} \mathbb{1}_{C_i}$, 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

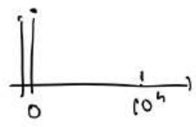
- valore atteso $E[X]$
- mediana: ogni valore $m_X \in \mathbb{R}$ t.c. $P\{X \leq m_X\} \geq \frac{1}{2}$, $P\{X \geq m_X\} \geq \frac{1}{2}$ (esercizio: esiste almeno una mediana)
- moda: ogni valore $mo_X \in \mathbb{R}$ di massimo per p_X (esercizio: esiste almeno una moda)

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:

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

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

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

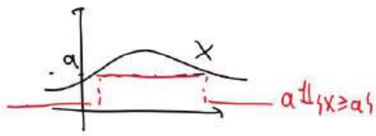
Prop (disuguaglianza di Markov):

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

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

Dim:

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



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

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

Dim:

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

($\{|X| \ge a\}$, 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

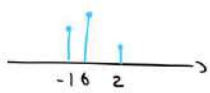
$$Var(X) = E[(X - E[X])^2] \in [0, +\infty)$$

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

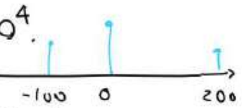


1) supponiamo di vincere 2 se esce 6, -1 se esce 1 o 2, 0 altrimenti

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 (Ω = {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) $Var(X) \geq 0$, $Var(X) = 0 \iff X$ = 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 $= 0 \iff (X - E[X])^2 \geq 0 \iff X = E[X]$ = cost. q.c.

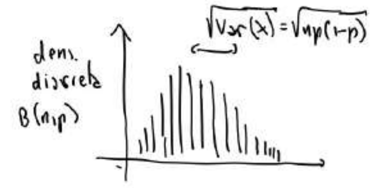
Esempi notevoli:

- $X \sim B(p)$

  $E[X^2] = 1^2 \cdot P\{X=1\} + 0^2 \cdot P\{X=0\} = p$, $Var(X) = p - p^2 = p(1-p)$

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

  $Var(X) = np(1-p)$ (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 $I_1, I_2, \ldots$ 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

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

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 $E[\varphi(X,Y)]=\sum_{i,j}\varphi(x_i,y_j)\,p_{(X,Y)}(x_i,y_j)$

$$E[XY]=\sum_{i,j}x_iy_j\,p_{(X,Y)}(x_i,y_j)\underset{\uparrow}{=}\sum_{i,j}x_iy_j\,p_X(x_i)\,p_Y(y_j)=\sum_i x_i\,p_X(x_i)\cdot\sum_j y_j\,p_Y(y_j)=E[X]E[Y]$$

indip: $p_{(X,Y)}(x,y)=p_X(x)p_Y(y)$

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, $\forall a\in\mathbb{R}$
  - $Cov(X,Y) = Cov(Y,X)$

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

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

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

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

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:

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

Esempio:

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

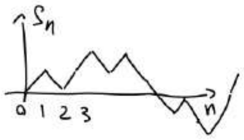
$Var(X) = \sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)$

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

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

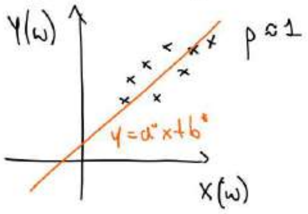
Def: La retta $y = a^* x + b^*$ è detta retta di regressione tra X e Y.

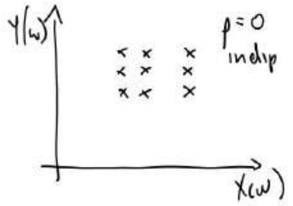
Significato:

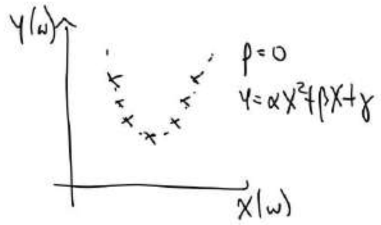
- la retta di regressione è la "migliore" approssimazione lineare tra X e Y, nel senso che minimizza la media della distanza al quadrato tra Y e $aX+b$
- poiché il valore del minimo è proporzionale a $1-\rho^2$, $1-\rho^2$ indica quanto la relazione tra X e Y sia approssimabile da una retta, in particolare:
  - $|\rho|$ 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. $\not\Rightarrow$ indipendente 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 dalla legge di X
- se un problema coinvolge più v.a. $X_1 \ldots X_n$, tutte le quantità di interesse dipendono solo dalla legge congiunta di $X_1 \ldots X_n$ (e, se le v.a. sono indip, solo dalle 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 congiunta 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 ripetati
  $\rightsquigarrow$ 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

· dette $X_i = X \circ \pi_i : \Omega_0^n \to S$ 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

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} m \quad \text{per } n \to \infty$$

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)

$$E[\bar{X}_n] = \frac{1}{n}\sum_{i=1}^{n} \underbrace{E[X_i]}_{\substack{\shortparallel \\ m \text{ per stessa legge}}} = m$$

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

Per Chebyshev, $\forall \varepsilon > 0$

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

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)

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

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

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

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

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

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

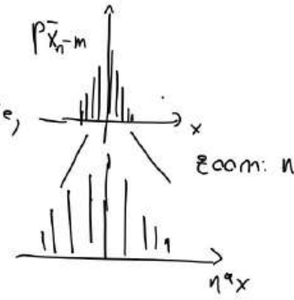
Notiamo che la freq assoluta $Y_n = X_1 + \dots + X_n$ è $\sim B(n,p)$, quindi la LGN dà

$\frac{Y_n}{n} \xrightarrow{P} p$ 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:

$Var(n^\alpha(\bar{X}_n - m)) = n^{2\alpha} Var(\bar{X}_n) = n^{2\alpha - 1}\sigma^2$ 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 $\alpha < \frac{1}{2}$, $n^\alpha(X_n - m) \xrightarrow{P} 0$

Per $\alpha = \frac{1}{2}$ ci si può attendere un limite non banale $\sim$ 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, $\forall\ -\infty \le a < b \le +\infty$,

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

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

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

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

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

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

Esempio importante: TCL per frequenze empiriche / 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 $\left(\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx = 1 \text{ e}\right)$

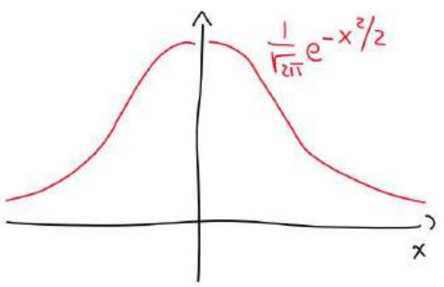
$$\int_{-3}^{3} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx \simeq 0.997$$

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

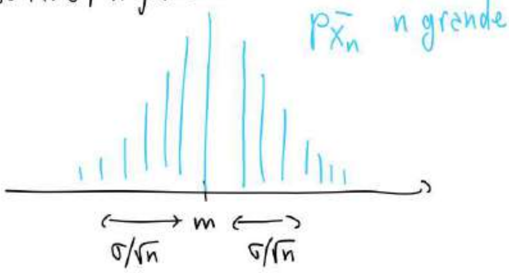
$\overline{X}_n \in \left[m - \frac{3\sigma}{\sqrt{n}}, m + \frac{3\sigma}{\sqrt{n}}\right]$ (o equiv. $n\overline{X}_n \in [mn - 3\sigma\sqrt{n}, mn + 3\sigma\sqrt{n}]$)

è ≈ 0.997.

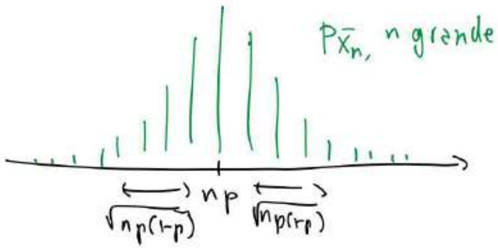
Esempio: su 1000 lanci di moneta equa ($m = p = \frac{1}{2}$, $\sigma^2 = p(1-p) = \frac{1}{4}$, $n = 1000$), la freq. assoluta di "testa" è, con prob ≈ 0.997, compresa tra 452 e 548



caso i.i.d. n grande



caso binomiale $B(n,p)$
n grande



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

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

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

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

dove abbiamo usato

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

$\{|\overline{X}_n - m| \ge \varepsilon\} = \{|\overline{X}_n - m|^2 > \varepsilon^2\}$ 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,

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

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 $I(t)>0$ $\forall t>0$

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

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

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 $= \sum_{i=1}^{N} X_i \quad \left(M(\omega) = \sum_{i=1}^{N(\omega)} X_i(\omega)\right)$

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 (cioè $x \in S_1$, $P\{X=x\} > 0$)

Def: Date $X: \Omega \to S_1$ v.a., $Y: \Omega \to S_2$ v.a., $x \in R_{p_X}$, 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
$P_Y(A \mid X=x) = P(Y \in A \mid X=x) \qquad \forall A \in \mathcal{P}(S_2)$

$P_Y(\cdot \mid X=x)$ 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)
$$P_Y(A) = \sum_{x \in R_{p_X}} P_Y(A \mid X=x)\, P_X(x) \qquad \forall A \in \mathcal{P}(S_2)$$

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 $P(\Omega_0) = 1$:
$\llcorner$ $P|_{\mathcal{P}(\Omega_0)}$ 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 $X=x$, $x \in R_{p_X}$,

$$E[\varphi(Y)|X=x] = E^{P(\cdot|X=x)}[\varphi(Y)] \quad \text{(valore atteso rispetto alla prob } P(\cdot|X=x)\text{)}$$

$$\left(= \sum_{\omega\in\Omega} \varphi(Y(\omega))\, P(\{\omega\}|X=x) \right)$$

$$= \sum_{y\in S_2} \varphi(y)\, p_{Y|X}(y|x)$$

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:

$$E[\varphi(Y)] = \sum_{y\in S_2} \varphi(y)\, p_Y(y) = \sum_{y\in S_2,\, x\in S_1} \varphi(y)\, p_{(X,Y)}(x,y) \overset{p_{Y|X}(y|x)p_X(x)}{=}$$

$$= \sum_{x\in S_1} \sum_{y\in S_2} \varphi(y)\, p_{Y|X}(y|x)\, p_X(x)$$

$$= \sum_{x\in S_1} E[\varphi(Y)|X=x]\, P\{X=x\}$$

$$= E[E[\varphi(Y)|X]]$$

Ad es,

$$E[M] = \sum_{n=0}^{\infty} E[M|N=n]\, P\{N=n\} \underset{\uparrow}{=} \sum_{n=0}^{\infty} np\, \frac{\lambda^n}{n!} e^{-\lambda} = p\,E[N] = \lambda p$$

$M \sim B(n,p)$ sotto $N=n$

Probabilità sulla retta reale

Def: Dati $\Omega \neq \phi$, $\mathcal{C} \subseteq \mathcal{P}(\Omega)$, σ-algebra generata da $\mathcal{C}$

$\sigma(\mathcal{C})$ = la più piccola σ-algebra contenente $\mathcal{C} = \bigcap_{\substack{\mathcal{F}\ \sigma\text{-algebra su } \Omega \\ \mathcal{F} \supseteq \mathcal{C}}} \mathcal{F}$

Oss: Intersezione di una famiglia di σ-algebre è σ-algebra (esercizio)

in particolare $\sigma(\mathcal{C}) = \bigcap_{\mathcal{F}\ \sigma\text{-alg} \supseteq \mathcal{C}} \mathcal{F}$ è una σ-algebra (la più piccola contenente $\mathcal{C}$)

Def. Dato $\mathcal{X} \neq \phi$ spazio metrico separabile (spesso $\mathcal{X} \subseteq \mathbb{R}^d$)

σ-algebra dei boreliani su $\mathcal{X}$: σ-algebra generata dagli aperti di $\mathcal{X}$

$\mathcal{B}(\mathcal{X}) = \sigma\{A \subseteq \mathcal{X} \mid A \text{ aperto}\}$

Proprietà dei boreliani:

- $\mathcal{B}(\mathcal{X})$ contiene tutti gli insiemi aperti, chiusi ed è generata dai chiusi
  tutti gli insiemi finiti o numerabili (unione numerabile di $\{x\}$, chiusi)
- $\mathcal{B}(\mathbb{R})$ contiene tutti gli intervalli (aperti, chiusi, semiaperti, semichiusi) e tutte le semirette
- $\mathcal{B}(\mathbb{R}) = \sigma\{(a,b] \mid a,b \in \mathbb{R}, a<b\}$
  $= \sigma\{(-\infty, x] \mid x \in \mathbb{R}\}$
- $\mathcal{B}(\mathbb{R}^d) = \sigma\{A_1 \times \dots \times A_d \mid A_1, \dots A_d \in \mathcal{B}(\mathbb{R})\}$
  $= \sigma\{(-\infty, x_1] \times \dots \times (-\infty, x_d] \mid x_1, \dots x_d \in \mathbb{R}\}$
- $\mathcal{B}(\mathbb{R}^d) \neq \mathcal{P}(\mathbb{R}^d)$
- Se $\mathcal{X}$ è metrico separabile e $\mathcal{Y} \subseteq \mathcal{X}$, $\mathcal{Y} \neq \phi$, ha la metrica indotta da $\mathcal{X}$,
  $\mathcal{B}(\mathcal{Y}) = \{A \cap \mathcal{Y} \mid A \in \mathcal{B}(\mathcal{X})\}$

(in pratica, $\mathcal{B}(\mathbb{R}^d)$ contiene tutti gli insiemi interessanti, ma non tutti i sottoinsiemi di $\mathbb{R}^d$)

Def: Dato $(\Omega, \mathcal{F})$ spazio misurabile, una misura μ su $(\Omega, \mathcal{F})$ è una funzione

$\mu: \mathcal{F} \to [0, +\infty]$ con $\mu(\phi) = 0$ e μ σ-additiva

Oss: P è prob su $(\Omega, \mathcal{F}) \iff P$ è misura su $(\Omega, \mathcal{F})$ e $P(\Omega) = 1$

Def: $N \in \mathcal{F}$ si dice μ-trascurabile: se $\mu(N) = 0$

Una proprietà q vale μ-q.o.: se $\exists N \in \mathcal{F}$ μ-trascurabile t.c. q vale su N^c.

Prop (misura di Lebesgue): Esiste un'unica misura m su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ che soddisfi $m([a,b]) = b-a$ $\forall a,b \in \mathbb{R}$, $a<b$. Essa è detta misura di Lebesgue (lunghezza)

Oss: $m|_{[0,1]} : \mathcal{B}([0,1]) \to [0,1]$ è misura di probabilità (probabilità uniforme su $[0,1]$)

Si può dimostrare, assumendo l'assioma della scelta, che $m|_{[0,1]}$ non si estende a $\mathcal{P}([0,1])$

Sia $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ e sia P probabilità su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

Def: Funzione di ripartizione (FdR) di P

$F: \mathbb{R} \to [0,1]$ definita da $F(x) = P((-\infty, x])$, $x \in \mathbb{R}$

Prop (proprietà di F):

a) F è non decrescente

b) F è continua a destra (cioè $\forall x \in \mathbb{R}$, $\lim_{y \downarrow x} F(y) = F(x)$)

c) $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to +\infty} F(x) = 1$

Dim:

a) $\forall x \le y$, $F(x) = P((-\infty, x]) \overset{\text{monotonia}}{\le} P((-\infty, y]) = F(y)$

b) Sia $x_n \downarrow x$. $(-\infty, x_n] \downarrow (-\infty, x]$ (cioè $(-\infty, x_n] \supseteq (-\infty, x_{n+1}]$, $\bigcap_n (-\infty, x_n] = (-\infty, x]$)

quindi per continuità di P per succ. decrescenti, $F(x_n) = P((-\infty, x_n]) \downarrow P((-\infty, x]) = F(x)$

c) Sia $x_n \downarrow -\infty$. $(-\infty, x_n] \downarrow \emptyset$, quindi per cont. per succ decrescenti $F(x_n) \downarrow P(\emptyset) = 0$

Sia $x_n \uparrow +\infty$. $(-\infty, x_n] \uparrow \mathbb{R}$, quindi per cont. per succ. crescenti, $F(x_n) \uparrow P(\mathbb{R}) = 1$

Prop (prob $\overset{1-1}{\longleftrightarrow}$ FdR):

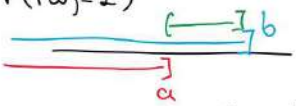
Data una funzione $F: \mathbb{R} \to \mathbb{R}$ che soddisfa (a), (b), (c) della prop. precedente, esiste un'unica probabilità P su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ avente F per funzione di ripartizione.

In particolare, F determina univocamente P.

La dimostrazione, che non vediamo, si basa sul fatto che le semirette $(-\infty, x]$, $x \in \mathbb{R}$, generano $\mathcal{B}(\mathbb{R})$ e sono chiuse per intersezione finita (π-sistema).

Calcolo di prob di intervalli con FdR: $(F(-\infty) = 0,\ F(+\infty) = 1)$

- $P((a,b]) = F(b) - F(a)$ $\quad \forall -\infty \le a < b \le +\infty$

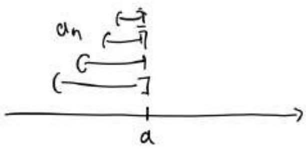


infatti $(a,b] = (-\infty, b] \setminus (-\infty, a]$ e quindi $P((a,b]) = P((-\infty, b]) - P((-\infty, a])$

- $P\{a\} = F(a) - F(a-)$ $\quad \forall a \in \mathbb{R}$, dove $F(a-) = \lim_{x \uparrow a} F(x)$

infatti, se $a_n \uparrow a$, allora $(a_n, a] \downarrow \{a\}$ e quindi

$P\{a\} = \lim_n P((a_n, a]) = \lim_n F(a) - F(a_n) = F(a) - F(a-)$



- $P([a,b]) = F(b) - F(a-)$ $\quad \forall -\infty \le a < b \le +\infty$

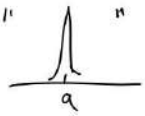
poiché $P([a,b]) = P((a,b]) + P\{a\}$

$$\left.\begin{aligned} P((a,b)) &= F(b-) - F(a) \\ P([a,b)) &= F(b-) - F(a-) \end{aligned}\right\} \text{(esercizio)}$$

Def: P si dice continua se F è continua, cioè se
$P\{a\}=0 \quad \forall a \in \mathbb{R}$

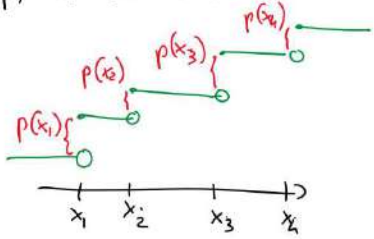
Esistono due grandi classi (non esaustive) di prob su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$:

1) Prob. discrete: $\exists\, \Omega_0 \subseteq \mathbb{R}$ al più numerabile t.c. $P(\Omega_0)=1$ $\quad (\Omega_0 = \{x_1, x_2, \dots\})$
senza perdita di generalità, possiamo prendere $\Omega_0 = R_P$ il range di P

Esempi: uniforme su $\{x_1, \dots x_N\}$, $B(p)$, $B(n,p)$, $G(p)$, $H(N, N_1, n)$, $P(\lambda)$, ...
Come abbiamo visto, assegnare P discreta equivale ad assegnare Ω_0 e $p: \Omega_0 \to \mathbb{R}$ densità discreta
Oss: Dato $a \in \mathbb{R}$, definiamo la misura di prob. δ_a (Delta di Dirac in a) su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$:
$$\delta_a(A) = \begin{cases} 1 & \text{se } a \in A \\ 0 & \text{se } a \notin A \end{cases}$$
"  " massa concentrata in a

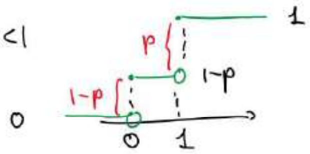
Allora P discreta si scrive come $P = \sum_i p(x_i)\,\delta_{x_i}$

Data P discreta, con range $R_P = \{x_1, x_2, \dots\}$ e densità discreta p, la FdR F di P è
$$F(x) = \sum_{i,\, x_i \le x} p(x_i)$$
infatti $F(x) = P((-\infty, x]) = \sum_{x_i \in (-\infty, x]} p(x_i)$



Se R_P non ha punti di accumulazione, F è costante a tratti, con salti negli x_i e ampiezza dei salti $p(x_i)$

Esempio: $B(p)$, $0<p<1$



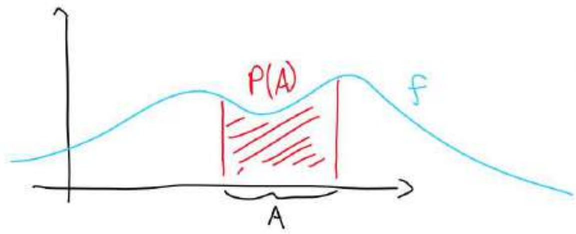
Esempio: dati $(p_r)_{r \in \mathbb{Q}}$ $\quad p_r > 0 \ \forall r$, $\sum_r p_r = 1$
$$P = \sum_{r \in \mathbb{Q}} p_r \delta_r$$ è prob discreta con FdR "non costante a tratti"

2) Probabilità assolutamente continue (rispetto alla misura di Lebesgue):

Def: P prob. su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ è assolutamente continua: se
$\exists\, f: \mathbb{R} \to \mathbb{R}$ boreliana (cioè t.c. $f^{-1}(A) \in \mathcal{B}(\mathbb{R})$ $\forall A \in \mathcal{B}(\mathbb{R})$) t.c.
$$P(A) = \int_A f(x)\,dx \quad \forall A \in \mathcal{B}(\mathbb{R})$$

Una tale funzione f si dice densità di P (rispetto alla misura di Lebesgue).
[Si assume implicitamente che f soddisfi $\int_{\mathbb{R}} |f(x)|\,dx < \infty$]

Oss: L'integrale $\int_A f(x)\,dx$ è inteso nel senso di Lebesgue, ma in molti esempi si riduce all'integrale di Riemann (eventualmente improprio)



Oss: La prob $P(A)$ è l'area sottesa da f su A.

Prop (caratterizzazione e unicità della densità)

a) Se P prob su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ha densità f, allora g è densità per $P \iff f = g$ Lebesgue q.o.

b) Se P prob su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ha densità f, allora f soddisfa

(i) $f \geq 0$ Lebesgue q.o.

(ii) $\int_{\mathbb{R}} f(x)\, dx = 1$

c) Viceversa, data $f: \mathbb{R} \to \mathbb{R}$ boreliana che soddisfa (i) e (ii), allora esiste un'unica prob P su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ avente f per densità, ed è data da

$$P(A) = \int_A f(x)\,dx, \quad A \in \mathcal{B}(\mathbb{R})$$

Dim: [q.o. = Lebesgue q.o.]

Ricordiamo: • Se $h \geq 0$ q.o. allora • $\int h\, dx \geq 0$

• $\int h\, dx = 0 \iff h = 0$ q.o.

• teor conv. monotona: $h_n \geq 0$ q.o., $h_n \uparrow h \Rightarrow \int h_n dx \uparrow \int h\, dx$

a) Se f è densità e $g = f$ q.o., allora

$$\left|P(A) - \int_A g(x)dx\right| = \left|\int_A f\,dx - \int_A g\,dx\right| \leq \int_{\mathbb{R}} |f-g|\,dx = 0$$

Viceversa, se f e g sono densità per P

$$\int \mathbb{1}_{\{f>g\}}(f-g)\,dx = \int_{\{f>g\}} f\,dx - \int_{\{f>g\}} g\,dx = P\{f>g\} - P\{f>g\} = 0,$$

Poiché $\mathbb{1}_{\{f>g\}}(f-g) \geq 0$, deve essere $\mathbb{1}_{\{f>g\}}(f-g) = 0$ q.o, cioè $f \leq g$ q.o.

Per simmetria, deve essere anche $f \geq g$ q.o. e quindi $f = g$ q.o.

b) Se f è densità, allora

$$0 \geq \int_{\{f<0\}} f(x)dx = P\{f<0\} \geq 0, \text{ quindi } \int \mathbb{1}_{\{f<0\}} f\,dx = 0 \text{ quindi } \mathbb{1}_{\{f<0\}} f = 0 \text{ q.o.} \checkmark$$

cioè $f \geq 0$ q.o

$$\int_{\mathbb{R}} f(x)dx = 1$$

c) Unicità: Se P ha densità f, deve essere $P(A) = \int_A f(x)dx \quad \forall A \in \mathcal{B}(\mathbb{R})$

Esistenza: mostriamo che $P(A) = \int_A f(x)dx$ verifica la def. di prob.

• $P(A) = \int_A f(x)dx \underset{f \geq 0 \text{ q.o.}}{\geq} 0 \qquad \forall A \in \mathcal{B}(\mathbb{R})$

- $P(\mathbb{R}) = \int_{\mathbb{R}} f(x)\,dx = 1$
- dati A_n a due a due disgiunti, $n \in \mathbb{N}^+$, abbiamo

$$\mathbb{1}_{\bigcup_n A_n} = \sum_n \mathbb{1}_{A_n} \quad \left(\text{infatti } \mathbb{1}_{\bigcup_n A_n}(\omega) = 1 \Leftrightarrow \omega \in \bigcup_n A_n \Leftrightarrow \omega \in A_{n_0} \text{ per uno e un solo } n_0 \Leftrightarrow \sum_n \mathbb{1}_{A_n}(\omega) = 1 \right)$$

quindi

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \int \mathbb{1}_{\bigcup_n A_n} f(x)\,dx = \int \sum_{n=1}^{\infty} \mathbb{1}_{A_n} f(x)\,dx$$

$$\overset{\text{teor. conv. monotona}}{=} \lim_{N\to\infty} \int \sum_{n=1}^{N} \mathbb{1}_{A_n} f(x)\,dx$$

$$\overset{\text{additività}}{=} \lim_{N\to\infty} \sum_{n=1}^{N} \int \mathbb{1}_{A_n} f(x)\,dx = \sum_{n=1}^{\infty} P(A_n)$$

FdR di prob. assolutamente continua: P prob su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, con FdR F

- Se P è assolutamente continua, con densità f, allora

$$F(x) = P((-\infty, x]) = \int_{-\infty}^{x} f(y)\,dy \quad \forall x \in \mathbb{R}$$

[Si può dimostrare anche il viceversa]

- Se P è assolutamente continua, allora è continua (poiché F è continua)

  Il viceversa è falso

  esempio: P prob associata a F scala di Cantor

- Se F è continua su $\mathbb{R}$ e C^1 a tratti (cioè F è C^1 eccetto che in un insieme di pt isolati), allora P è assolutamente continua, con densità

$$f(x) = F'(x) \quad (\text{dove esiste } F')$$

Dim:

Prendiamo $f = F'$ dove esiste, estendendo $f = 0$ dove F' non esiste. Notiamo che $f \geq 0$ poiché F è non-decrescente e, per il teor fondamentale del calcolo,

$$F(x) = \int_{-\infty}^{x} f(y)\,dy \quad \forall x \in \mathbb{R}$$

In particolare, $\int_{-\infty}^{+\infty} f(x)\,dx = \lim_{x\to+\infty} F(x) = 1$ e quindi f è una densità. Prendiamo Q prob su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ avente f per densità. Allora P e Q hanno la stessa FdR e quindi $P = Q$ e P ha densità $f = F'$.

Oss.: Se P è assol. continua con densità f, allora $P\{f=0\} = \int_{\{f=0\}} f\,dx = 0$

Def: Data P prob su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, supporto (topologico) di P: $\operatorname{supp} P \subseteq \mathbb{R}$ chiuso t.c.

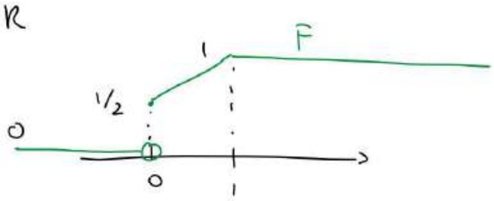
$$(\operatorname{supp} P)^c = \bigcup_{A \text{ aperto} \subseteq \mathbb{R},\ P(A)=0} A$$

Lemma: $P((\operatorname{supp} P)^c) = 0$

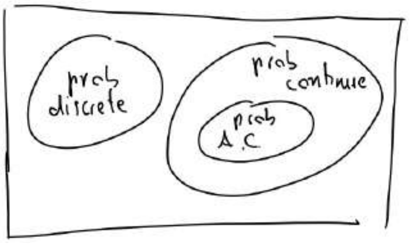
Esistono prob P su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ che non sono né discrete né continue

Es: P prob associata alla FdR

$$F(x) = \begin{cases} 0 & x<0 \\ \frac{1}{2}(1+x) & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$



Prob reali

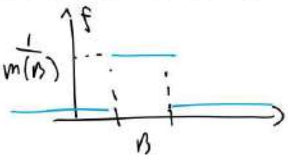


Esempi notevoli di probabilità assolutamente continue

P prob su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ con densità f, m misura di Lebesgue

- Dato $B \in \mathcal{B}(\mathbb{R})$ con $0 < m(B) < \infty$, distribuzione uniforme su B $(\mathcal{U}(B))$

$$f(x) = \frac{1}{m(B)} \mathbb{1}_B(x) = \begin{cases} \frac{1}{m(B)} & x \in B \\ 0 & x \notin B \end{cases}$$

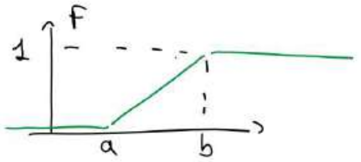


$$\forall A \in \mathcal{B}(\mathbb{R}),\ P(A) = \int_A \frac{1}{m(B)} \mathbb{1}_B(x)\,dx = \frac{m(A \cap B)}{m(B)}$$

In particolare, se $B = (a,b)$ intervallo, $a<b$,

$$f(x) = \frac{1}{b-a} \mathbb{1}_{(a,b)}(x)$$

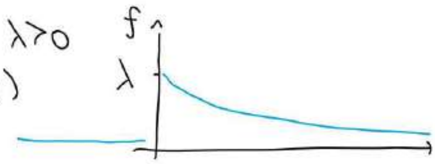
FdR: $$F(x) = \begin{cases} 0 & x<a \\ \frac{x-a}{b-a} & a \le x < b \\ 1 & x \ge b \end{cases}$$



Significato: "scegliamo un pt a caso su (a,b), senza preferenze"

- Distribuzione esponenziale di parametro $\lambda > 0$ $(\mathcal{E}xp(\lambda))$

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{(0,+\infty)}(x)$$



FdR $F(x) = (1 - e^{-\lambda x}) \mathbb{1}_{x \ge 0}$

Significato: "tempi di attesa senza memoria"

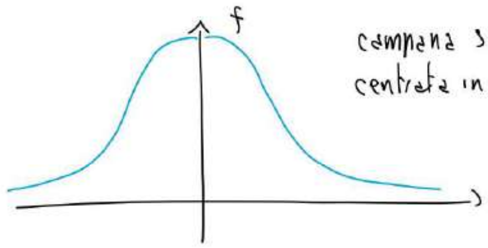
- Distribuzione Gamma di parametri $r>0$ e $\lambda > 0$ $(\Gamma(r,\lambda))$

$$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} \mathbb{1}_{(0,+\infty)}(x)$$

dove $\Gamma(r) = \int_0^\infty x^{r-1} e^{-x}\,dx$ $\quad (\Gamma(r) = (r-1)!\ \forall r \in \mathbb{N}^+)$

- Distribuzione gaussiana, o normale, standard: $\mathcal{N}(0,1)$:

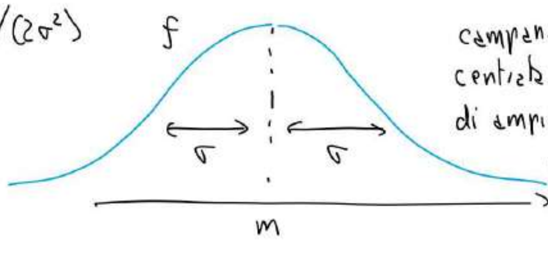
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$



campana simmetrica centrata in $x=0$

- Distribuzione gaussiana, o normale, di media $m \in \mathbb{R}$ e varianza $\sigma^2 > 0$: $\mathcal{N}(m, \sigma^2)$:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/(2\sigma^2)}$$



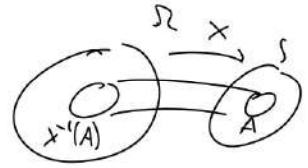
campana simmetrica centrata in $x=m$
di ampiezza proporzionale a σ
pt di flesso in $m \pm \sigma$

14
lunedì 3 aprile 2023 21:58

Variabili aleatorie generali

Def: Dati $(\Omega, \mathcal{F})$, $(S, \mathcal{S})$ spazi misurabili, una variabile aleatoria (v.a) da Ω in S è una funzione $X: \Omega \to S$ misurabile da $\mathcal{F}$ a $\mathcal{S}$, cioè t.c.

$X^{-1}(A) \in \mathcal{F} \quad \forall A \in \mathcal{S}$



Tipicamente: $S = \mathbb{R}$ e in questo caso si prende $\mathcal{S} = \mathcal{B}(\mathbb{R})$: v.a. reale
$S = \mathbb{R}^d$ " " " " " " $\mathcal{S} = \mathcal{B}(\mathbb{R}^d)$: vettore aleatorio

Ricordiamo: dato $A \subseteq S$, $X^{-1}(A) := \{X \in A\} := \{\omega \in \Omega \mid X(\omega) \in A\}$
X^{-1} commuta con le operazioni insiemistiche $\cup, \cap, ^c$

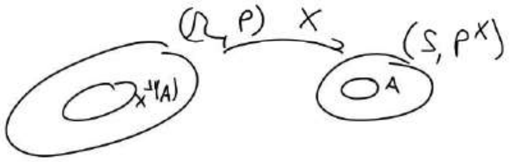
Oss: Se $\mathcal{F} = \mathcal{P}(\Omega)$, ogni funzione $X: \Omega \to S$ è misurabile

Fatto: Se $\mathcal{C} \subseteq \mathcal{P}(S)$ genera $\mathcal{S}$ (cioè $\mathcal{S} = \sigma(\mathcal{C})$) e $X^{-1}(C) \in \mathcal{F}$ $\forall C \in \mathcal{C}$, allora X è misurabile

In particolare, se $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, X è misurabile se, ad es, $(-\infty, x] \in \mathcal{F}$ $\forall x \in \mathbb{R}$

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

$P^X(A) = P\{X \in A\} = P(X^{-1}(A))$, $A \in \mathcal{S}$



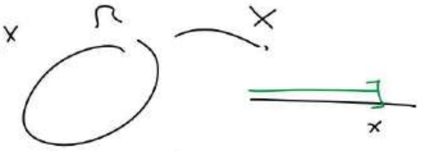
Lemma: P^X è una probabilità su $(S, \mathcal{S})$

Dim come nel caso discreto.

Per v.a. X reale,

- chiamiamo funzione di ripartizione F^X di X la FdR di P^X

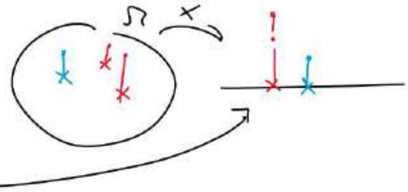
$F^X(x) = P^X((-\infty, x]) = P\{X \le x\} \quad \forall x \in \mathbb{R}$



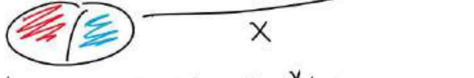
- diciamo che X è discreta, con densità discreta p^X, se P^X è discreta con dens. discr. p^X:

$p^X(x) = P^X\{x\} = P\{X = x\}$,

$P\{X \in A\} = P^X(A) = \sum_{x \in A} p^X(x) \left(= \sum_{x \in A \cap R_{p^X}} p^X(x) \right) \quad \forall A \in \mathcal{B}(\mathbb{R})$



Oss: non è detto che Ω sia discreto



- diciamo che X è continua se P^X è continua, cioè $P\{X = x\} = P^X\{x\} = 0$ $\forall x \in \mathbb{R}$

· diciamo che X è assolutamente continua con densità f, se P^X è assol. cont. con dens. f

$$P\{X\in A\} = P^X(A) = \int_A f(x)\,dx \qquad \forall A \in \mathcal{B}(\mathbb{R})$$

Analogamente per X v.a. a valori in $\mathbb{R}^d$

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

$A \in \mathcal{F} \iff \mathbb{1}_A$ è misurabile (da $\mathcal{F}$ a $\mathcal{B}(\mathbb{R})$) (esercizio)

Costruzione canonica: dato $(S, \mathcal{S})$ sp misurabile e Q probabilità su $(S, \mathcal{S})$, prendiamo $(\Omega, \mathcal{F}) = (S, \mathcal{S})$, $P = Q$ prob su $(\Omega, \mathcal{F})$ e $X : \Omega \to S$, $X = id$

allora X è v.a. di legge $P^X = Q$

Composizione di v.a.: dati $(\Omega, \mathcal{F}), (S, \mathcal{S}), (S', \mathcal{S}')$ sp. misurabili,

$X : \Omega \to S$ v.a., $\varphi : S \to S'$ misurabile, allora $\varphi \circ X : \Omega \to S'$ è v.a. (cioè misurabile):

infatti $\forall A \in \mathcal{S}'$, $\{\varphi(X) \in A\} = \{X \in \underbrace{\varphi^{-1}(A)}_{\in \mathcal{S} \text{ per mis. di } \varphi}\} \in \mathcal{F}$ per mis. di X

Oss: Se X è discreta, con range R_{p_X}, allora $\varphi(X)$ è discreta, con range $\varphi(R_{p_X})$

infatti · $P\{\varphi(X) \in \varphi(R_{p_X})\} \geq P\{X \in \varphi(R_{p_X})\} = 1$

· $\# \varphi(R_{p_X}) \leq \# R_{p_X} \leq \# \mathbb{N}$

· $\forall y = \varphi(x)$, $x \in R_{p_X}$, $P\{\varphi(X) = y\} \geq P\{X = x\} > 0$

Invece se $X : \Omega \to \mathbb{R}$ è continua, non è detto che $\varphi(X)$ sia continua.

es: $\varphi \equiv 0$.

I concetti e le proprietà di:

· uguaglianza q.c. per v.a. ($X = Y$ q.c. se $P\{X = Y\} = 1$)

· uguaglianza in legge per v.a. ($X \overset{(d)}{=} Y$ se $P^X = P^Y$)

si estendono senza difficoltà al caso generale.

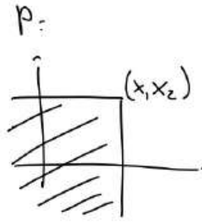
Distribuzioni congiunte e indipendenza per v.a. generali

Lemma (Dynkin): Sia $(\Omega, \mathcal{F})$ sp. misurabile, sia $\mathcal{C} \subseteq \mathcal{F}$ con $\mathcal{C}$ chiusa per intersezione finita (cioè $A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C}$) e t.c. $\sigma(\mathcal{C}) = \mathcal{F}$. Siano P, Q due probabilità su $(\Omega, \mathcal{F})$. Se P e Q coincidono su $\mathcal{C}$, allora $P = Q$.

Probabilità su $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$:

- Prop: Esiste un'unica misura m su $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ t.c.
  $m([a_1,b_1] \times \dots \times [a_d, b_d]) = (b_1 - a_1) \cdot \dots \cdot (b_d - a_d) \quad \forall a_1, b_1, \dots a_d, b_d$, con $a_i < b_i \ \forall i$
  Essa è detta misura di Lebesgue d-dimensionale (volume)

- Def: Data P prob su $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, funzione di ripartizione di P:
  $F: \mathbb{R}^d \to \mathbb{R}$ definita da $F(x_1, \dots x_d) = P((-\infty, x_1] \times \dots \times (-\infty, x_d])$



Prop:
- F è non decrescente rispetto a ogni variabile
  $F(x_1, \dots x_{i-1}, x, x_{i+1}, \dots x_d) \le F(x_1, \dots x_{i-1}, y, x_{i+1}, \dots x_d) \quad \forall x \le y, \ \forall i$
- F è continua a destra rispetto a ogni variabile
  $\lim_{y \downarrow x} F(x_1, \dots x_{i-1}, y, x_{i+1}, \dots x_d) = F(x_1, \dots x_{i-1}, x, x_{i+1}, \dots x_d) \quad \forall x, \ \forall i$
- $\forall i, \ \lim_{x \to -\infty} F(x_1, \dots x_{i-1}, x, x_{i+1}, \dots x_d) = 0$
  $\lim_{x_1 \to +\infty} \dots \lim_{x_d \to +\infty} F(x_1, \dots x_d) = 1$

Prop: F determina univocamente P (cioè se P e Q hanno la stessa FdR, allora $P = Q$)

Dim:

Se P e Q hanno la stessa FdR, allora P e Q coincidono su
$\mathcal{C} = \{(-\infty, x_1] \times \dots \times (-\infty, x_d] \mid (x_1, \dots x_d) \in \mathbb{R}^d\}$. Poiché $\mathcal{C}$ è chiusa per inters. finita e $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R}^d)$, per il lemma di Dynkin, abbiamo $P = Q$ su $\mathcal{B}(\mathbb{R}^d)$.

- P prob su $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ è detta discreta: se $\exists \ \Omega_0 \subseteq \mathbb{R}^d$ al più numerabile t.c. $P(\Omega_0) = 1$
- P prob su $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ è detta assolutamente continua: se $\exists f: \mathbb{R}^d \to \mathbb{R}$ boreliana t.c.
  $P(A) = \int_A f(x)\,dx \quad \forall A \in \mathcal{B}(\mathbb{R}^d)$ (integrale di Lebesgue)
  f si dice densità di P.

Prop: a) Se P ha densità f, allora P ha densità $g \iff f = g$ Lebesgue q-o.

b) Se P ha densità f, allora

i) $f \geq 0$ Lebesgue-q.o

ii) $\int_{\mathbb{R}^d} f(x)\,dx = 1$

c) Viceversa, se $f: \mathbb{R}^d \to \mathbb{R}$ boreliana soddisfa (i) e (ii), allora esiste un'unica prob P su $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ che ha densità f

FdR: Se P è assolutamente continua con densità f, allora P ha FdR

$$F(x_1 \dots x_d) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} f(y_1, \dots y_d)\, dy_1 \dots dy_d \quad \forall (x_1 \dots x_d) \in \mathbb{R}^d \quad (\Delta)$$

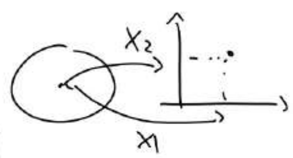
Viceversa, se P ha F che soddisfa (Δ) per qualche densità f (cioè $f: \mathbb{R}^d \to \mathbb{R}$ boreliana che soddisfa (i) e (ii)), allora P ha densità f.

infatti, detta Q la prob con densità f, Q e P hanno le stesse FdR e quindi $P = Q$.

"Se F è sufficientemente regolare, allora $f(x_1 \dots x_d) = \partial_{x_1} \dots \partial_{x_d} F(x_1 \dots x_d)$"

In generale, si può usare la formula $f = \partial_{x_1} \dots \partial_{x_d} F$ per indovinare f e poi verificare che f è densità e che $F(x_1 \dots x_d) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} f(y_1 \dots y_d)\, dy_1 \dots dy_d$ (e quindi P ha densità f)

Date $X_1: \Omega \to S_1, \dots X_n: \Omega \to S_n$, $A_1 \subseteq S_1, \dots A_n \subseteq S_n$,



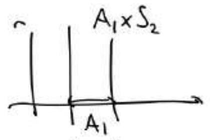
chiamiamo $(X_1, \dots X_n): \Omega \to S_1 \times \dots \times S_n$, $(X_1, \dots X_n)(\omega) = (X_1(\omega), \dots X_n(\omega))$

vettore aleatorio

Ricordiamo

$$\{X_1 \in A_1, \dots X_n \in A_n\} = \{X_1 \in A_1, \dots X_n \in A_n\} = \bigcap_{i=1}^{n} \{X_i \in A_i\}$$

$$\{X_i \in A_i\} = \{(x_1, \dots x_n) \in S_1 \times \dots \times S_{i-1} \times A \times S_{i+1} \times \dots \times S_n\}$$



Def. Dati $(S_1, \mathcal{S}_1), \dots (S_n, \mathcal{S}_n)$ spazi misurabili, si definisce σ-algebra prodotto $\mathcal{S}_1 \otimes \dots \otimes \mathcal{S}_n$ su $S_1 \times \dots \times S_n$:

$$\mathcal{S}_1 \otimes \dots \otimes \mathcal{S}_n = \sigma\{\underbrace{A_1 \times \dots \times A_n}_{\text{insiemi cilindrici}} \mid A_1 \in \mathcal{S}_1, \dots A_n \in \mathcal{S}_n\}$$

Oss: Dato $(\Omega, \mathcal{F})$ sp. mis., $X_i: \Omega \to S_i$, $i = 1, \dots n$,

X_i è misurabile da $\mathcal{F}$ a $\mathcal{S}_i$, $\forall i = 1, \dots n \iff (X_1, \dots X_n): \Omega \to S_1 \times \dots \times S_n$ è misurabile da $\mathcal{F}$ a $\mathcal{S}_1 \otimes \dots \otimes \mathcal{S}_n$

In particolare, $(X_1, \dots X_n): \Omega \to \mathbb{R}^n$ è misurabile (in $\mathcal{B}(\mathbb{R}^n)$) $\iff X_i$ è mis. (in $\mathcal{B}(\mathbb{R})$) $\forall i$.

Dim: $\Rightarrow)$ $\forall A_1 \in \mathcal{S}_1, \dots A_n \in \mathcal{S}_n$,

$$\underbrace{\{(X_1, \dots X_n) \in A_1 \times \dots \times A_n\}}_{} = \underbrace{\{X_1 \in A_1\}}_{\in \mathcal{F}} \cap \dots \cap \underbrace{\{X_n \in A_n\}}_{\in \mathcal{F}} \in \mathcal{F}$$ poiché X_i sono mis. da $\mathcal{F}$ a $\mathcal{S}_i$

poiché gli insiemi cilindrici generano $\mathcal{S}_1 \otimes \dots \otimes \mathcal{S}_n$, $(X_1 \dots X_n)$ è mis da $\mathcal{F}$ a $\mathcal{S}_1 \otimes \dots \otimes \mathcal{S}_n$

$\Leftarrow)$: $\forall i = 1, \dots n$, $\forall A_i \in \mathcal{S}_i$, $\{X_i \in A_i\} = \{(X_1 \dots X_n) \in S_1 \times \dots \times S_{i-1} \times A_i \times S_{i+1} \times \dots \times S_n\} \in \mathcal{F}$ poiché $(X_1 \dots X_n)$ è mis da $\mathcal{F}$ a $\mathcal{S}_1 \otimes \dots \otimes \mathcal{S}_n$

Def: Data $X = (X_1, \dots X_n) : \Omega \to S_1 \times \dots \times S_n$ v.a (misurabile da $\mathcal{F}$ a $\mathcal{S}_1 \otimes \dots \otimes \mathcal{S}_n$)

legge congiunta di $(X_1, \dots X_n)$: la legge P^X di X su $(S_1 \times \dots \times S_n, \mathcal{S}_1 \otimes \dots \otimes \mathcal{S}_n)$

leggi marginali: le leggi $P^{X_1}, \dots P^{X_n}$ rispettivamente di $X_1, \dots X_n$ su $(S_1, \mathcal{S}_1), \dots (S_n, \mathcal{S}_n)$

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

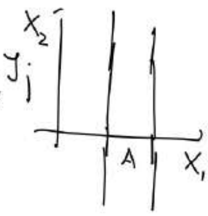
Oss: $\forall A_i \in \mathcal{S}_i$, $P\{X_1 \in A_1, \dots X_n \in A_n\} = P\{(X_1, \dots X_n) \in A_1 \times \dots \times A_n\} = P_{(X_1, \dots X_n)}(A_1 \times \dots \times A_n)$

Prop: La legge congiunta determina le leggi marginali: precisamente, $\forall i = 1, \dots n$

$$P_{X_i}(A) = P\{X_i \in A\} = P\{(X_1, \dots X_n) \in S_1 \times \dots \times S_{i-1} \times A_i \times S_{i+1} \times \dots \times S_n\} \quad \forall A \in \mathcal{S}_i$$

e più in generale, $\forall J \subsetneq I$,

$$P_{(X_j)_{j \in J}}(A) = P\{(X_j)_{j \in J} \in A, (X_i)_{i \in I \setminus J} \in \times_{i \in I \setminus J} S_i\} \quad \forall A \in \otimes_{j \in J} \mathcal{S}_j$$



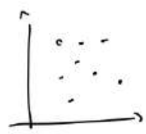
Dim: (come nel caso discreto)

La prima formula segue dell' osservazione

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

e analogamente la seconda da $\{(X_j)_{j \in J} \in A\} = \{(X_j)_{j \in J} \in A, (X_i)_{i \in I \setminus J} \in \times_{i \in I \setminus J} S_i\}$

Caso discreto:



$X = (X_1, \dots X_n) : \Omega \to \mathbb{R}^n$ è discreto $\Leftrightarrow$ $X_i : \Omega \to S_i$ è discreta $\forall i = 1, \dots n$

Infatti: · se X è discreto con range R_{P_X}, allora detta π_i la proiezione canonica, $X_i = \pi_i(X)$ è discreto con range $\pi_i(R_{P_X})$

· se X_i sono discreti con range $R_{P_{X_i}}$, allora $X \in R_{P_{X_1}} \times \dots \times R_{P_{X_n}}$ q.c. e $\# R_{P_{X_1}} \times \dots \times R_{P_{X_n}} \leq \# \mathbb{N}$, quindi $R_{P_{(X_1 \dots X_n)}} \subseteq R_{P_{X_1}} \times \dots \times R_{P_{X_n}}$

Dal caso discreto abbiamo:

Prop: Dette $p_{(X_1 \dots X_n)}$ la densità discreta congiunta di $(X_1 \dots X_n)$ e p_{X_i} le densità discrete marginali, vale

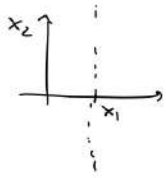
$$p_{X_i}(x) = \sum_{x_j \in \mathbb{R}, j \neq i} p_{(X_1 \dots X_n)}(x_1, \dots x_{i-1}, x, x_{i+1}, \dots x_n) \qquad x \in \mathbb{R}$$

dove $\sum_{x_j \in \mathbb{R},\, j \neq i} = \sum_{(x_1, \dots x_{i-1}, x, x_{i+1}, \dots x_n) \in R_{p_{(X_1 \dots X_n)}}}$

Caso assolutamente continuo:

Prop: Se $X = (X_1 \dots X_n) : \Omega \to \mathbb{R}^n$ è assolutamente continuo con densità congiunta $f : \mathbb{R}^n \to \mathbb{R}$, allora, $\forall i = 1, \dots n$, X_i è assolutamente continua con densità marginale $f_i : \mathbb{R} \to \mathbb{R}$ data da

$$f_i(x) = \int_{\mathbb{R}^{n-1}} f(x_1, \dots x_{i-1}, x, x_{i+1}, \dots x_n)\, dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n, \quad x \in \mathbb{R}$$



Dim: $i = 1$ per semplicità di notazione

$$P\{X_1 \in A\} = P\{X_1 \in A, X_2 \in \mathbb{R}, \dots X_n \in \mathbb{R}\}$$
$$= \int_{A \times \mathbb{R}^{n-1}} f(x_1, x_2, \dots x_n)\, dx_1, \dots dx_n$$

Fubini-Tonelli
$$= \int_A \left(\int_{\mathbb{R}^{n-1}} f(x_1, x_2, \dots x_n)\, dx_2, \dots dx_n \right) dx_1$$
$$= \int_A f_1(x_1)\, dx_1 \qquad \forall A \in \mathcal{B}(\mathbb{R})$$

quindi f_1 è densità per X_1

Fubini-Tonelli:

Se $f : \mathbb{R}^2 \to \mathbb{R}$ è boreliana ≥ 0 o integrabile, allora

- $f(\cdot, x_2)$, $x_2 \mapsto \int f(x_1, x_2)\, dx_1$ sono boreliane, e analogamente scambiando x_1 e x_2
- $\iint f(x_1, x_2)\, dx_1 dx_2 = \int \left(\int f(x_1, x_2)\, dx_1 \right) dx_2 = \int \left(\int f(x_1, x_2)\, dx_2 \right) dx_1$

Analogamente per $f : \mathbb{R}^n \to \mathbb{R}$.

Oss: Se X_1, X_2 v.a. reali sono assolutamente continue, non è detto che $(X_1, X_2) : \Omega \to \mathbb{R}^2$ sia assolutamente continuo

Esempio: X_1 assolutamente continua, $X_2 = X_1$, (X_1, X_1) non è assolutamente continuo

idea: (X_1, X_1) è concentrato su $\{(x, x) \mid x \in \mathbb{R}\}$, che ha misura di Lebesgue 2D nulla

per assurdo: (X_1, X_1) ha densità f, allora

$$0 = \iint_{\underbrace{\{x_1 = x_2\}}_{= \{(x_1, x_1) \mid x_1 \in \mathbb{R}\}}} f(x_1, x_2)\, dx_1 dx_2 = P\{X_1 = X_2\} = 1 \quad \text{↯}$$

Indipendenza di v.a.: $((\Omega, \mathcal{F}, P)$ sp di prob., $(S_i, \mathcal{I}_i)$ sp misurabili, $i = 1, \dots n)$

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

$P\{X_1\in A_1, \dots X_n \in A_n\} = P\{X_1 \in A_1\} \cdot \dots \cdot P\{X_n \in A_n\} \quad \forall A_1 \in \mathcal{J}_1, \dots A_n \in \mathcal{J}_n$

Oss: Come nel caso discreto, $(X_i)_{i=1}^n$ sono indip $\Leftrightarrow \{X_i \in A_i\}_{i=1}^n$ sono indip. $\forall A_i \in \mathcal{J}_i, i=1,\dots n$

Oss: Come nel caso discreto, l'indipendenza è una proprietà della legge congiunta di $(X_1 \dots X_n)$

Def: Sia $(S_i, \mathcal{J}_i)$, $i\in I$, una famiglia di spazi misurabili (con I possibilmente infinito) sia $X_i : \Omega \to S_i$, $i \in I$, una famiglia di v.a. $(X_i)_{i\in I}$ si dice famiglia di v.a. indipendenti: se, $\forall J \subseteq I$ finito, $(X_j)_{j\in J}$ è famiglia di v.a. indipendenti.

Prop: Siano $X_i : \Omega \to \mathbb{R}$, $i=1,\dots n$, v.a. reali, siano F la FdR di $(X_1 \dots X_n)$ e, per $i=1,\dots n$, F_i la FdR di X_i. Allora

$(X_1 \dots X_n)$ è famiglia di v.a. indipendenti se e solo se

$F(x_1 \dots x_n) = F_1(x_1) \cdot \dots \cdot F_n(x_n) \quad \forall (x_1 \dots x_n) \in \mathbb{R}^n$

Dim:

$\Rightarrow$): Se $X_1 \dots X_n$ sono indipendenti, allora

$F(x_1 \dots x_n) = P\{(X_1, \dots X_n) \in (-\infty, x_1] \times \dots \times (-\infty, x_n]\} = P\{X_1 \le x_1, \dots X_n \le x_n\}$

indip $\to = P\{X_1 \le x_1\} \cdot \dots \cdot P\{X_n \le x_n\} = F_1(x_1) \cdot \dots \cdot F(x_n)$

$\Leftarrow$): Siano $Y_1 \dots Y_n$ v.a. reali con $Y_i \overset{(d)}{=} X_i$ $\forall i=1,\dots n$ e Y_i indipendenti

[Si può dimostrare che esistono tali Y_i su un opportuno spazio, si veda teor successivo]

Poiché $X_i \overset{(d)}{=} Y_i$, Y_i ha F_i come FdR, e per indipendenza la FdR di $(Y_1 \dots Y_n)$ è

$F_{(Y_1 \dots Y_n)}(x_1 \dots x_n) = F_1(x_1) \cdot \dots \cdot F(x_n) = F(x_1, \dots x_n) \forall (x_1 \dots x_n) \in \mathbb{R}^n$

Quindi $(X_1 \dots X_n)$ e $(Y_1 \dots Y_n)$ hanno la stessa FdR, quindi hanno la stessa legge. In particolare, $X_1 \dots X_n$ sono indipendenti

Prop: Sia $(X_1, \dots X_n) : \Omega \to \mathbb{R}^n$ un vettore aleatorio discreto ($\Leftrightarrow X_i$ sono v.a discrete).

$(X_1, \dots X_n)$ è famiglia di v.a. indipendenti se e solo se

$p_{(X_1 \dots X_n)}(x_1 \dots x_n) = p_{X_1}(x_1) \cdot \dots \cdot p_{X_n}(x_n) \quad \forall (x_1 \dots x_n) \in \mathbb{R}^n$

(equivalentemente $\forall (x_1 \dots x_n) \in R_{p_{(X_1 \dots X_n)}}$)

(come visto nel caso discreto)

Prop: Sia $(X_1, \dots X_n) : \Omega \to \mathbb{R}^n$ un vettore aleatorio.

a) Se $(X_1 \dots X_n)$ è assolutamente continuo e la sua densità congiunta f soddisfa

$f(x_1, \dots x_n) = f_1(x_1) \cdot \dots \cdot f_n(x_n) \quad \forall (x_1, \dots x_n) \in \mathbb{R}^n \quad (*)$

con $f_1, \dots f_n$ densità marginali, allora $(X_1, \dots X_n)$ è famiglia di v.a. indipendenti.

b) Viceversa, se $X_1, \dots X_n$ sono v.a. reali assolutamente continue e indipendenti, allora $(X_1, \dots X_n)$ è assolutamente continuo e la sua densità congiunta f soddisfa (*).

Dim:

a) $\forall A_1, \dots A_n \in \mathcal{B}(\mathbb{R}^n)$

$$P\{X_1 \in A_1, \dots X_n \in A_n\} = \int_{A_1 \times \dots \times A_n} f(x_1 \dots x_n)\, dx_1 \dots dx_n$$

$$\underset{\text{ipotesi}}{=} \int_{A_1 \times \dots \times A_n} f(x_1) \dots f(x_n)\, dx_1 \dots dx_n$$

$$\underset{\text{Fubini Tonelli}}{=} \int_{A_1} \int_{A_2} \dots \int_{A_n} f(x_1) f(x_2) \dots f(x_n)\, dx_n \dots dx_2\, dx_1$$

$$= \int_{A_1} f(x_1)\, dx_1 \cdot \dots \cdot \int_{A_n} f(x_n)\, dx_n = P\{X_1 \in A_1\} \cdot \dots \cdot P\{X_n \in A_n\}$$

quindi $X_1 \dots X_n$ sono indipendenti

b) Usiamo il criterio per FdR

$$F(x_1, \dots x_n) \underset{\text{indip.}}{=} F(x_1) \cdot \dots \cdot F(x_n) =$$

$$= \int_{-\infty}^{x_1} f(y_1)\, dy_1 \cdot \dots \cdot \int_{-\infty}^{x_n} f(y_n)\, dy_n$$

$$\underset{\text{Fubini Tonelli}}{=} \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(y_1) \cdot \dots \cdot f(y_n)\, dy_1 \dots dy_n \qquad \forall (x_1 \dots x_n) \in \mathbb{R}^n$$

Esercizio: $f(x_1) \dots f(x_n)$ è una densità su $\mathbb{R}^n$

Quindi $P_{(X_1 \dots X_n)}$ ha densità $f(x_1) \dots f(x_n)$.

Dim alternativa con lemma di Dynkin:

Dobbiamo verificare che

$$P_{(X_1 \dots X_n)}(A) = \int_A f_1(x_1) \dots f_n(x_n)\, dx_1 \dots dx_n \qquad \forall A \in \mathcal{B}(\mathbb{R}^n)$$
$$= Q(A)$$

dove Q è la prob. di densità $f_1(x_1) \dots f_n(x_n)$ (esercizio: $f_1(x_1) \dots f_n(x_n)$ è una densità)

Poiché $\mathcal{C} = \{A_1 \times \dots \times A_n \mid A_1, \dots A_n \in \mathcal{B}(\mathbb{R}^n)\}$ è chiusa per intersezione finita e genera $\mathcal{B}(\mathbb{R}^n)$, basta verificare $P_{(X_1 \dots X_n)}(A_1 \times \dots \times A_n) = Q(A_1 \times \dots \times A_n)$ $\forall A_i \in \mathcal{B}(\mathbb{R})$

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

$$\underset{\text{indip}}{=} P\{X_1 \in A_1\} \cdot \dots \cdot P\{X_n \in A_n\}$$

$$= \int_{A_1} f_1(x_1)\, dx_1 \cdot \dots \cdot \int_{A_n} f_n(x_n)\, dx_n$$

$$\underset{\text{Fubini Tonelli}}{=} \int_{A_1 \times \dots \times A_n} f(x_1) \cdot \dots \cdot f(x_n)\, dx_1 \dots dx_n$$

Lemma (stabilità per composizione): Siano $(\Omega, \mathcal{F}, P)$ sp di prob,
$(S_i, \mathcal{S}_i), (S_i', \mathcal{S}_i')$ spazi misurabili, Se $X_i : \Omega \to S_i$, $i \in I$, sono v.a. indipendenti e
$\varphi_i : S_i \to S_i'$ sono misurabili, allora $\varphi_i(X_i) : \Omega \to S_i$, $i \in I$, sono v.a. indipendenti

Dim come nel caso discreto

Lemma (gruppi disgiunti di v.a. indip sono indip.)
Se $X_i : \Omega \to S_i$, $i \in I$, sono v.a. indip. e $J_1, \dots J_m \subseteq I$ sono disgiunti e finiti,
allora $(X_j)_{j \in J_1} : \Omega \to \times_{j \in J_1} S_j, \dots (X_j)_{j \in J_2} : \Omega \to \times_{j \in J_m} S_j$ sono v.a. indipendenti

La dimostrazione, che non vediamo, usa il lemma di Dynkin.
Esercizio: dimostrare il lemma per I finito, X_i assolutamente continue

Oss. (costruzione canonica): Data Q legge su $(S_1 \times \dots \times S_n, \mathcal{S}_1 \otimes \dots \otimes \mathcal{S}_n)$, la costruzione canonica fornisce $(\Omega, \mathcal{F}, P)$, $X = (X_1 \dots X_n)$ v.a. a valori in $S_1 \times \dots \times S_n$ di legge Q

Teor: Dati $(S_1, \mathcal{S}_1, Q_1), \dots (S_n, \mathcal{S}_n, Q_n)$ spazi di prob, esiste un'unica probabilità (detta prob prodotto) $Q (= Q_1 \otimes \dots \otimes Q_n)$ su $(S_1 \times \dots \times S_n, \mathcal{S}_1 \otimes \dots \otimes \mathcal{S}_n)$ che sia la legge di $(X_1 \dots X_n)$, con $X_1, \dots X_n$ indipendenti e di leggi marginali $Q_1 \dots Q_n$ rispettivamente.

16
martedì 11 aprile 2023 17:43

Valore atteso, momenti e varianza

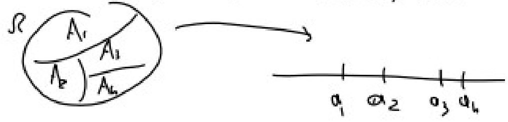
$(\Omega, \mathcal{F}, P)$ spazio di prob..

Data $X: \Omega \to \mathbb{R}$ v.a reale, vogliamo definire il valore atteso $E[X]$ di X in modo t.c.

- per X discreta, $E[X]$ coincida con la definizione/caratterizzazione vista nel caso discreto
  $E[X] = \sum_i x_i p_X(x_i) = \sum_i x_i P\{X = x_i\}$
- E abbia "buone proprietà" di passaggio al limite: "$X_n \to X$ in un senso opportuno $\Rightarrow E[X_n] \to E[X]$"

Def: $X: \Omega \to \mathbb{R}$ v.a. si dice semplice se assume un numero finito di valori, cioè

$$X = \sum_{i=1}^{m} a_i \mathbb{1}_{A_i}$$



con $a_1 \dots a_m \in \mathbb{R}$, $A_1, \dots A_m \in \mathcal{F}$, $m \in \mathbb{N}^+$.

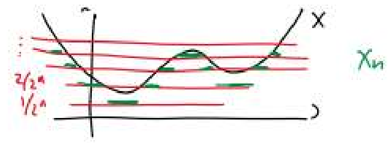
Data X semplice, $X = \sum_{i=1}^{m} a_i \mathbb{1}_{A_i}$, definiamo

$$E[X] = \int_\Omega X \, dP := \sum_{i=1}^{m} a_i P(A_i)$$

Oss (esercizio):
- $E[X]$ non dipende dalla scelta degli A_i e degli a_i
- In particolare, possiamo prendere gli a_i come tutti distinti, in questo caso $A_i = \{X = a_i\}$, e troviamo $E[X] = \sum_{i=1}^{m} a_i P\{X = a_i\}$, come visto nel caso discreto

Lemma: Data $X: \Omega \to \mathbb{R}$ v.a ≥ 0, esiste $(X_n)_n$ successione di v.a. semplici ≥ 0, nondecrescente (cioè $X_n(\omega) \leq X_{n+1}(\omega)$ $\forall n$, $\forall \omega \in \Omega$) e t.c. $X_n \nearrow X$ puntualmente (cioè $X_n(\omega) \nearrow X(\omega)$ $\forall \omega$).

Dim:
Basta prendere $X_n = \sum_{k=0}^{n2^n - 1} \frac{k}{2^n} \mathbb{1}_{[\frac{k}{2^n}, \frac{k+1}{2^n})}(X) + n \mathbb{1}_{[n, +\infty)}(X)$



Prop/def: Data $X: \Omega \to \mathbb{R}$ v.a ≥ 0, sia $(X_n)_n$ successione non decrescente di v.a. semplici, ≥ 0, con $X_n \nearrow X$ puntualmente. Definiamo valore atteso di X il numero

$$E[X] := \int_\Omega X \, dP := \lim_n E[X_n] \in [0, +\infty] \quad \left(\text{si scrive anche } \int_\Omega X(\omega) P(d\omega) = \int_\Omega X(\omega) dP(\omega)\right)$$

Il limite esiste e non dipende dalla successione $(X_n)_n$ scelta.

Def. Data $X: \Omega \to \mathbb{R}$ v.a., $X^+ = \max\{X, 0\}$, $X^- = \max\{-X, 0\}$, se $E[X^+] < \infty$ oppure $E[X^-] < \infty$ definiamo

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

Diciamo che X è integrabile se $E[X^+] < \infty$ e $E[X^-] < \infty$, equivalentemente se $E[|X|] < \infty$

Oss: $|X| = X^+ + X^-$, $X = X^+ - X^-$

$E[|X|] = E[X^+] + E[X^-]$

se X è integrabile, $|E[X]| \leq E[|X|] < \infty$

Def: Data $X:\Omega\to\mathbb{R}$ v.a. ≥ 0 o integrabile, $A\in\mathcal{F}$,

$$\int_A X\,dP := \int_\Omega \mathbb{1}_A X\,dP = E[\mathbb{1}_A\cdot X]$$

Oss: $E[\mathbb{1}_A]=P(A)$

Lemma (proprietà di $E[X]$):

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

b) $X\overset{(d)}{=}Y$, $X\geq 0$ o integrabile $\Rightarrow Y\geq 0$ q.c. o integrabile, $E[Y]=E[X]$

in particolare, $X=Y$ q.c. $\Rightarrow E[Y]=E[X]$

c) $X\geq 0$ q.c $\Rightarrow E[X]\geq 0$

d) $X\geq 0$ q.c., $E[X]=0 \Rightarrow X=0$ q.c.

e) $X\geq Y$ q.c., $E[Y^-]<\infty$ oppure $E[X^+]<\infty \Rightarrow E[X]\geq E[Y]$

f) linearità: X,Y entrambe integrabili, $a\in\mathbb{R} \Rightarrow aX+Y$ integrabile, $E[aX+Y]=aE[X]+E[Y]$

lo stesso se X,Y sono entrambe ≥ 0 q.c. e $a\geq 0$.

Lemma (Fatou): Siano X_n, $n\in\mathbb{N}^+$, v.a reali ≥ 0 q.c. Allora

$$E[\liminf_n X_n]\leq \liminf_n E[X_n]$$

Teor (di convergenza monotona): Siano X_n, $n\in\mathbb{N}$, X v.a. reali con $0\leq X_n\nearrow X$ q.c. (cioè, per q.o. ω, $X_n(\omega)\geq 0$, $(X_n(\omega))_n$ non decrescente, $X_n(\omega)\nearrow X(\omega)$). Allora

$$E[X_n]\nearrow E[X]$$

Teor (di convergenza dominata): Siano X_n, $n\in\mathbb{N}$, X v.a reali, con $X_n\to X$ q.c (cioè $P\{X_n\to X\}=1$)

Supponiamo che esista Y integrabile ≥ 0 con $|X_n|\leq Y$ q.c. $\forall n$. Allora X_n e X sono integrabili e $E[X_n]\to E[X]$.

Oss: Le stesse def. e proprietà si estendono all'integrale secondo Lebesgue

$$\int_\Omega \varphi\,d\mu$$

con $\varphi:\Omega\to\mathbb{R}$ boreliana, μ misura su $(\Omega,\mathcal{F})$. L'unica differenza è che nella def. di funzione semplice $\varphi=\sum_{i=1}^m a_i\mathbb{1}_{A_i}$, si chiede anche che $\mu(A_i)<\infty$ $\forall i$.

In particolare, per $\mu=m$ misura di Lebesgue su $\mathbb{R}$ o su $\mathbb{R}^d$, si trova la def di

$\int_{\mathbb{R}}\varphi\,dx$, $\int_{\mathbb{R}^d}\varphi\,dx$ (integrale rispetto alla misura di Lebesgue)

Prop (valore atteso in funzione della legge): $(\Omega,\mathcal{F},P)$ sp. di prob.

a) Sia $X:\Omega\to\mathbb{R}$ v.a.

- X è integrabile $\Leftrightarrow \int_{\mathbb{R}}|x|\,P^X(dx)<\infty$

- Se X è integrabile o ≥ 0 q.c., allora
$$E[X] = \int_{\mathbb{R}} x \, P^X(dx)$$

b) Più in generale, date $(S, \mathcal{S})$ sp. mis, $X: \Omega \to S$ v.a., $\varphi: S \to \mathbb{R}$ boreliana,

- $\varphi(X)$ è integrabile $\iff \int_S |\varphi(x)| \, P^X(dx) < \infty$
- Se $\varphi(X)$ è integrabile o ≥ 0 q.c., allora
$$E[\varphi(X)] = \int_S \varphi(x) P^X(dx)$$

Dim:

(a) Segue da (b) con $\varphi(x) = x$.

(b) Dimostriamo (b) per 1. $\varphi = \mathbb{1}_A$, 2. φ semplice, 3. $\varphi \geq 0$, 4. φ generica

1. Se $\varphi = \mathbb{1}_A$, allora $\varphi(X) = \mathbb{1}_A(X) = \mathbb{1}_{\{X \in A\}}$ e quindi
$$E[\mathbb{1}_A(X)] = E[\mathbb{1}_{X \in A}] = P\{X \in A\} = P^X(A) = \int_S \mathbb{1}_A \, P^X(dx)$$

2. Per φ semplice (combinazione lin. di indicatrici) si usa (1.) e linearità di E

3. Per $\varphi \geq 0$, siano φ_n semplici, $0 \leq \varphi_n \uparrow \varphi$, quindi $0 \leq \varphi(X_n) \uparrow \varphi(X)$

   Per (2), $E[\varphi_n(X)] = \int_S \varphi_n(x) \, P^X(dx)$ $\forall n$.

   Per conv monotona, $E[\varphi(X_n)] \uparrow E[\varphi(X)]$
   $$\int_S \varphi_n(x) P^X(dx) \uparrow \int_S \varphi(x) P^X(dx)$$
   quindi $E[\varphi(X)] = \int_S \varphi(x) P^X(dx)$

4. Per φ generica (boreliana), $\varphi(X)$ è integrabile $\iff E[|\varphi(X)|] \overset{(3)}{=} \int |\varphi(x)| P^X(dx) < \infty$

   Scomponendo $\varphi = \varphi^+ - \varphi^-$, e applicando (3) a φ^+, φ^-, si ottiene $E[\varphi(X)] = \int \varphi(x) P^X(dx)$

Oss importante: $E[\varphi(X)]$ dipende solo dalla legge di X

(cioè se $X \overset{(d)}{=} Y$, allora $E[\varphi(X)] = E[\varphi(Y)]$)

Calcolo del valore atteso:

- Caso $X: \Omega \to S$ discreta, con densità discreta p_X: come abbiamo visto, per $\varphi: S \to \mathbb{R}$ boreliana,
  - $\varphi(X)$ integrabile $\iff E[|\varphi(X)|] = \sum_{x \in R_{p_X}} |\varphi(x)| p_X(x) < \infty$
  - se $\varphi(X)$ è integrabile o ≥ 0 q.c., $E[\varphi(X)] = \sum_{x \in R_{p_X}} \varphi(x) p_X(x)$
- Caso $X: \Omega \to \mathbb{R}$ assolutamente continua, con densità f: per $\varphi: \mathbb{R} \to \mathbb{R}$ boreliana
  - $\varphi(X)$ integrabile $\iff E[|\varphi(X)|] = \int_{\mathbb{R}} |\varphi(x)| f(x) \, dx < \infty$
  - se $\varphi(X)$ è integrabile o ≥ 0 q.c., $E[\varphi(X)] = \int_{\mathbb{R}} \varphi(x) f(x) \, dx$

Dim:

Seguendo lo schema indicatrici, semplici, ≥ 0, generiche

1. $\varphi = \mathbb{1}_A$, $A \in \mathcal{B}(\mathbb{R})$: $E[\mathbb{1}_A(X)] = P\{X \in A\} = \int_A f(x)\,dx$
2. φ semplice: da (1.) usando linearità di E
3. $\varphi \geq 0$: prendiamo φ_n semplici, $0 \leq \varphi_n \uparrow \varphi$, allora per conv dominata e (2)

$$E[\varphi(X)] \underset{\text{conv. monotona}}{=} \lim_n E[\varphi_n(X)] \underset{(2)}{=} \lim_n \int \varphi_n(x) f(x)\,dx \underset{\text{conv monotona per } \varphi_n f}{=} \int \varphi(x) f(x)\,dx$$

4. φ generica: la condizione di integrabilità segue da (3) per $|\varphi|$
   l'uguaglianza $E[\varphi(X)] = \int \varphi(x) f(x)\,dx$ segue da (3) per φ_+, φ_-

Esempi notevoli:

- $X \sim \mathcal{U}((a,b))$, $a<b$ $\left(f(x) = \frac{1}{b-a} \mathbb{1}_{(a,b)}(x)\right)$

  X è limitata q.c., quindi $E[|X|] = \int |x| f(x)\,dx < \infty$

  $$E[X] = \int x f(x)\,dx = \frac{1}{b-a}\int_a^b x\,dx = \frac{1}{b-a}\,\frac{1}{2}(b^2-a^2) = \frac{a+b}{2}$$

- $X \sim \mathcal{E}xp(\lambda)$, $\lambda > 0$ $\left(f(x) = \lambda e^{-\lambda x} \mathbb{1}_{(0,+\infty)}(x)\right)$

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

  $$E[X] = \int_0^{+\infty} x \lambda e^{-\lambda x}\,dx \underset{\text{per parti}}{=} -x e^{-\lambda x}\Big|_{x=0}^{+\infty} + \int_0^{\infty} e^{-\lambda x}\,dx = 0 - \frac{1}{\lambda} e^{-\lambda x}\Big|_{x=0}^{+\infty} = \frac{1}{\lambda}$$

- $X \sim \Gamma(r,\lambda)$, $r>0$, $\lambda>0$

  $E[X] = \frac{r}{\lambda}$

- $X \sim \mathcal{N}(0,1)$ $\left(f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\right)$

  $$E[|X|] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |x| e^{-x^2/2}\,dx < \infty$$

  $$E[X] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x e^{-x^2/2}\,dx = \frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^0 x e^{-x^2/2}\,dx + \int_0^{+\infty} x e^{-x^2/2}\,dx\right) = 0$$

  $$\int_{-\infty}^0 x e^{-x^2/2}\,dx \underset{y=-x}{=} \int_0^{+\infty} -y\, e^{-y^2/2}\,dy$$

- $X \sim \mathcal{N}(m,\sigma^2)$

  $E[X] = m$ (con cambio di variabile oppure standardizzazione, vedi dopo)

Def. Data $X: \Omega \to \mathbb{R}$ v.a. reale, mediana di X: ogni valore $m \in \mathbb{R}$ t.c. $P\{X \leq m\} \geq \frac{1}{2}$ e $P\{X \geq m\} \geq \frac{1}{2}$.

Esercizio: $m = \inf\{t \in \mathbb{R} \mid F_X(t) \geq \frac{1}{2}\}$ (con F_X FdR di X) è una mediana

Le definizioni e proprietà dei momenti si estendono senza difficoltà dal caso discreto al caso generale in particolare ricordiamo disuguaglianza di Markov per $X \geq 0$: $\forall a > 0$

$P\{X \geq a\} \leq \frac{1}{a} E[X]$

e il suo corollario $P\{|X| > a\} \leq \frac{1}{a^p} E[|X|^p]$

Esempio notevole: momenti di una gaussiana standard $\mathcal{N}(0,1)$: per $X \sim \mathcal{N}(0,1)$, $p>0$

$$E[|X|^p] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |x|^p e^{-x^2/2} dx < \infty$$

quindi X ammette momenti di ogni ordine $p>0$; per $p \in \mathbb{N}^+$:

$$E[X^p] = \int_{\mathbb{R}} x^p \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \begin{cases} 0 & \text{per } p \text{ dispari, per simmetria} \\ (p-1)!! & \text{per } p \text{ pari, usando induzione e integrazione per parti (esercizio)} \end{cases}$$

dove $(p-1)!! = (p-1)\cdot(p-3)\cdot \ldots \cdot 3 \cdot 1$

Le definizioni e proprietà di varianza e deviazione standard, per X con $E[|X|^2] < \infty$,

$$Var(X) = E[(X - E[X])^2], \quad \sigma(X) = \sqrt{Var(X)}$$

si estendono senza difficoltà dal caso discreto al caso generale, in particolare ricordiamo la dis. di Chebyshev

$$P\{|X - E[X]| > a\} \leq \frac{1}{a^2} Var(X)$$

Esempi notevoli:

- $X \sim \mathcal{U}((a,b))$ $\quad \left(f(x) = \frac{1}{b-a} 1_{(a,b)}(x)\right)$

  $E[X^2] = \int x^2 f(x) dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{1}{b-a} \frac{1}{3}(b^3 - a^3) = \frac{a^2+ab+b^2}{3}$

  $Var(X) = E[X^2] - E[X]^2 = \frac{(b-a)^2}{12}$

- $X \sim \mathcal{E}xp(\lambda)$, $\lambda > 0$ $\quad \left(f(x) = \lambda e^{-\lambda x} 1_{(0,+\infty)}(x)\right)$

  $E[X^2] = \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx = -x^2 e^{-\lambda x}\Big|_{x=0}^{+\infty} + \int_0^{+\infty} 2x e^{-\lambda x} dx = \frac{2}{\lambda^2}$

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

- $X \sim \Gamma(r, \lambda)$

  $Var(X) = \frac{r}{\lambda^2}$

- $X \sim \mathcal{N}(0,1)$

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

- $X \sim \mathcal{N}(m, \sigma^2)$

  $Var(X^2) = \sigma^2$

17
sabato 15 aprile 2023 16:18

Indipendenza, covarianza e correlazione

$(\Omega, \mathcal{F}, P)$ sp. di prob.

Oss: Date $X: \Omega \to S_1$, $Y: \Omega \to S_2$ v.a. (con $(S_1, \mathcal{F}_1)$, $(S_2, \mathcal{F}_2)$ sp. misurabili), con legge congiunta $P_{(X,Y)}$ su $(S_1 \times S_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$, data $\varphi: S_1 \times S_2 \to \mathbb{R}$ boreliana con $\varphi(X,Y) \geq 0$ q.c. o $\varphi(X,Y)$ integrabile, vale

$$E[\varphi(X,Y)] = \int_{S_1 \times S_2} \varphi(x,y)\, P_{(X,Y)}(d(x,y))$$

Calcolo nel caso discreto: se (X,Y) è discreta ($\Leftrightarrow X, Y$ discrete), $E[\varphi(X,Y)] = \sum_{(x,y)} \varphi(x,y)\, p_{(X,Y)}(x,y)$ con $p_{(X,Y)}$ densità congiunta

Calcolo nel caso assolutamente continuo:

Se $(X,Y): \Omega \to \mathbb{R}^2$ è assolutamente continua con densità $f: \mathbb{R}^2 \to \mathbb{R}$, allora

- $\varphi(X,Y)$ è integrabile $\Leftrightarrow \int_{\mathbb{R}^2} |\varphi(x,y)| f(x,y)\, dx dy < \infty$
- se $\varphi(X,Y)$ è ≥ 0 q.c. o integrabile, allora
$$E[\varphi(X,Y)] = \int_{\mathbb{R}^2} \varphi(x,y) f(x,y)\, dx dy$$

Dim: come nel caso di una v.a. reale assolutamente continua (indicatrici, semplici, ≥ 0, generali)

Queste formule si generalizzano in modo naturale al caso $(X_1, \dots X_n): \Omega \to S_1 \times \dots \times S_n$

- $E[\varphi(X_1, \dots X_n)] = \int_{S_1 \times \dots \times S_n} \varphi(x_1, \dots x_n)\, P_{(X_1, \dots X_n)}(d(x_1, \dots x_n))$
- caso discreto: $E[\varphi(X_1 \dots X_n)] = \sum_{(x_1, \dots x_n)} \varphi(x_1, \dots x_n)\, p_{(X_1, \dots X_n)}(x_1, \dots x_n)$
- caso $(X_1 \dots X_n): \Omega \to \mathbb{R}^n$
$$E[\varphi(X_1 \dots X_n)] = \int_{\mathbb{R}^n} \varphi(x_1, \dots x_n) f(x_1, \dots x_n)\, dx_1 \dots dx_n$$

La dis di Schwarz si estende senza difficoltà dal caso discreto al caso generale

$$E[|XY|] \leq E[X^2]^{1/2} E[Y^2]^{1/2} \quad \text{per } X, Y \text{ v.a. reali}$$

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

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

Dim: Secondo lo schema indicatrici, semplici, ≥ 0, generali

1. $X = 1\!\!1_A \quad Y = 1\!\!1_B$, $A, B \in \mathcal{F}$

   $1\!\!1_A, 1\!\!1_B$ sono indip $\Leftrightarrow$ A e B sono indip.

   dim di $\Rightarrow$: $A = \{1\!\!1_A = 1\}$, $B = \{1\!\!1_B = 1\}$, quindi $P(A \cap B) = P\{1\!\!1_A = 1, 1\!\!1_B = 1\}$
   $= P\{1\!\!1_A = 1\} \cdot P\{1\!\!1_B = 1\} = P(A)P(B)$

   $$E[\underbrace{1\!\!1_A 1\!\!1_B}_{= 1\!\!1_{A \cap B}}] = P(A \cap B) \underset{\text{indip. di A e B}}{=} P(A)P(B) = E[1\!\!1_A]E[1\!\!1_B]$$

2. X, Y semplici: da (1) e linearità: $X = \sum_{i=1}^{n} a_i 1\!\!1_{A_i}$, $Y = \sum_{j=1}^{m} b_j 1\!\!1_{B_j}$

   $$E[XY] = E\left[\sum_{i,j} a_i b_j 1\!\!1_{A_i} 1\!\!1_{B_j}\right] = \sum_{i,j} a_i b_j E[1\!\!1_{A_i} 1\!\!1_{B_j}]$$
   $$\underset{(1)}{=} \sum_{i,j} a_i b_j E[1\!\!1_{A_i}]E[1\!\!1_{B_j}] = \sum_i a_i E[1\!\!1_{A_i}] \cdot \sum_j b_j E[1\!\!1_{B_j}]$$
   $$= E[X]E[Y]$$

3. $X, Y \geq 0$: da (2) e conv. monotona

   siano $X_n = h_n(X)$, $Y_n = h_n(Y)$, con $h_n: \mathbb{R} \to \mathbb{R}$, $h_n(x) = \sum_{k=0}^{2^n - 1} \frac{k}{2^n} 1\!\!1_{[\frac{k}{2^n}, \frac{k+1}{2^n})}(x) + n 1\!\!1_{[n, +\infty)}(x)$

   X_n, Y_n sono semplici, $0 \leq X_n \uparrow X$, $0 \leq Y_n \uparrow Y$ e quindi $X_n Y_n$ sono semplici, $0 \leq X_n Y_n \uparrow XY$

   $\forall n$, X_n e Y_n sono indipendenti, perché funzioni di v.a. indipendenti

   $$E[XY] \underset{\text{conv. mon.}}{=} \lim_n E[X_n Y_n] \underset{(2)}{=} \lim_n E[X_n]E[Y_n] \underset{\text{conv mon.}}{=} E[X]E[Y]$$

4. X, Y integrabili: da (3)

   $X^+ = \max\{X, 0\}$, $Y^+ = \max\{Y, 0\}$ sono indip. poiché funz. di indip.,

   analogamente X^-, Y^+, sono indip, X^+, Y^- sono indip, X^-, Y^- sono indip.

   quindi si decompone $XY = X^+Y^+ - X^-Y^+ - X^+Y^- + X^-Y^-$ e si usa (3).

Dim per (X, Y) v.a. assolutamente continua:

Se (X, Y) è assol cont con densità f, $\quad$ X e Y sono indip. $\Leftrightarrow f(x,y) = f_X(x) f_Y(y)$
con f_X e f_Y densità di X e Y risp.

$$E[|XY|] = \iint_{\mathbb{R}^2} |xy| f(x,y)\, dx\, dy \underset{\text{indip}}{=} \iint_{\mathbb{R}^2} |x| f_X(x) |y| f_Y(y)\, dx\, dy$$

$$\underset{\text{Fubini-Tonelli}}{=} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |x| f_X(x)\, dx\right) |y| f_Y(y)\, dy = \int_{\mathbb{R}} |x| f_X(x)\, dx \int_{\mathbb{R}} |y| f_Y(y)\, dy$$

$$= E[|X|]\, E[|Y|] < \infty \text{ se } X, Y \text{ sono integrabili}$$

Per $E[XY]$, si ripetono i passaggi sopra senza i moduli e si ottiene $E[XY] = E[X]E[Y]$

Come nel caso discreto, si dimostra

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

X, Y sono indip. $\iff E[g(X)h(Y)] = E[g(X)]E[h(Y)]$ $\forall g: S_1 \to \mathbb{R}$, $h: S_2 \to \mathbb{R}$ boreliane

con $g(X), h(Y)$ entrambe ≥ 0 q.c.

o entrambe integrabili

Le def. e le proprietà di covarianza, coefficiente di correlazione, retta di regressione si estendono senza difficoltà dal caso discreto al caso generale, in particolare ricordiamo:

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

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

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

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

18
sabato 15 aprile 2023 19:53

Trasformazioni di v.a.

Oss: Abbiamo visto che, se $X:\Omega \to \mathbb{R}$ è assolutamente continua e $\varphi:\mathbb{R}\to\mathbb{R}$ è boreliana, anche regolare, non è detto che $\varphi(X)$ sia assolutamente continua (es: $\varphi \equiv 0$)

Prop (Formula di cambio variabili): Sia $X:\Omega\to\mathbb{R}^d$ v.a. assolutamente continua, con densità f_X, con $f_X=0$ fuori da un aperto $O\subseteq\mathbb{R}^d$. Sia $\varphi: O\to O'$ un diffeomorfismo C^1 con $O'\subseteq\mathbb{R}^d$ aperto (cioè $\varphi:O\to O'$ è C^1, invertibile e φ^{-1} è C^1). Allora la v.a. $Y=\varphi(X):\Omega\to\mathbb{R}^d$ è assolutamente continua con densità

$$f_Y(y)=f_X(\varphi^{-1}(y))\,|\det D\varphi^{-1}(y)|\,\mathbb{1}_{O'}(y) \qquad y\in\mathbb{R}^d$$

[Idea: " $y=\varphi(x)$ $\quad x=\varphi^{-1}(y)$ $\quad dx=|\det D\varphi^{-1}(y)|\,dy$ "]

Dim:

Per $A\in\mathcal{B}(\mathbb{R}^d)$, dobbiamo dim $P\{Y\in A\}=\int_A f_Y(y)\,dy$, f_Y come sopra

$$P\{\underset{\varphi(X)}{Y}\in A\}=P\{X\in\varphi^{-1}(A)\}=\int \mathbb{1}_{\varphi^{-1}(A)}(x)\,f_X(x)\,\mathbb{1}_O(x)\,dx$$

cambio variabile: $x=\varphi^{-1}(y)$, $dx=\det D\varphi^{-1}(y)dy$

$$=\int \underbrace{\mathbb{1}_{\varphi^{-1}(A)}(\varphi^{-1}(y))}_{=\mathbb{1}_A(y)}\,f_X(\varphi^{-1}(y))\,|\det D\varphi^{-1}(y)|\,\overbrace{\mathbb{1}_O(\varphi^{-1}(y))}^{=\mathbb{1}_{O'}(y)}\,dy$$

$$=\int \mathbb{1}_A(y)\,f_Y(y)\,dy$$

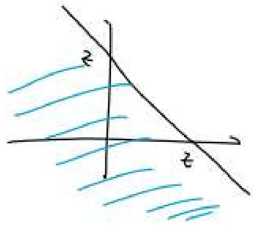
Nel caso generale, per $X:\Omega\to\mathbb{R}^d$ v.a., $\varphi:\mathbb{R}^d\to\mathbb{R}^m$ boreliana, si può calcolare la FdR di $Y=\varphi(X)$ per studiare le proprietà della legge di Y

$$F_Y(y_1,\dots,y_m)=P\{\varphi_1(X)\le y_1,\dots,\varphi_m(X)\le y_m\}$$

Prop (densità della somma):

Siano $X:\Omega\to\mathbb{R}$, $Y:\Omega\to\mathbb{R}$ v.a. reali con (X,Y) assolutamente continua di densità $f_{(X,Y)}$. Allora $X+Y$ è assolutamente continua, con densità

$$f_{X+Y}(z)=\int_{\mathbb{R}} f_{(X,Y)}(x,z-x)\,dx=\int_{\mathbb{R}} f_{(X,Y)}(z-y,y)\,dy$$



Dim:

Sia F_{X+Y} la FdR di $X+Y$, vogliamo mostrare

$$F_{X+Y}(z)=\int_{-\infty}^{z}\dots \qquad \forall z\in\mathbb{R}$$

$$F_{X+Y}(z) = P\{X+Y \le z\} = P\{(X,Y) \in \{(x,y) | x+y \le z\}\}$$
$$= \iint 1\!\!1_{x+y\le z}\, f_{(X,Y)}(x,y)\,dx\,dy \quad \left[\text{dove } 1\!\!1_{x+y\le z} = 1\!\!1_{\{(x',y')|x'+y'\le z\}}(x,y)\right]$$
$$= \int \Big(\underbrace{\int 1\!\!1_{x+y\le z}\, f_{(X,Y)}(x,y)\,dy}_{} \Big)\,dx$$

cambio var $y = z'-x$, $dy = dz'$

$$= \int 1\!\!1_{z'\le z}\, f_{(X,Y)}(x, z'-x)\,dz'$$
$$= \iint \Big(\int 1\!\!1_{z'\le z}\, f_{(X,Y)}(x, z'-x)\,dx\Big)dy = \int_{-\infty}^{z} f_{X+Y}(z')\,dz'$$

L'altra formula si dimostra scambiando x e y.

Cor: (formula della convoluzione):

Se $X:\Omega\to\mathbb{R}$, $Y:\Omega\to\mathbb{R}$ sono v.a. reali, assolutamente continue con densità risp. f_X, f_Y, e indipendenti, allora $X+Y$ è assolutamente continua con densità

$$f_{X+Y}(z) = \int_{\mathbb{R}} f_X(x) f_Y(z-x)\,dx = \int_{\mathbb{R}} f_X(z-y) f_Y(y)\,dy =: f_X * f_Y(z)$$

(convoluzione di f_X e f_Y)

Dim: dalla prop precedente con $f_{(X,Y)}(x,y) = f_X(x) f_Y(y)$.

Prop Siano X, Y v.a. (discrete) a valori interi, con densità discreta congiunta $p_{(X,Y)}$. Allora $X+Y$ ha densità discreta $p_{X+Y}(k) = \sum_{j\in\mathbb{Z}} p_{(X,Y)}(j, k-j) = \sum_{j\in\mathbb{Z}} p_{(X,Y)}(j,k-j)$, $k\in\mathbb{Z}$.

In particolare, se X e Y sono indipendenti, di densità discrete risp. p_X, p_Y,

$$p_{X+Y}(k) = \sum_{j\in\mathbb{Z}} p_X(j)\, p_Y(k-j) = \sum_{j\in\mathbb{Z}} p_X(k-j)\, p_Y(j), \quad k\in\mathbb{Z}$$

Applicazioni:

- $X\sim B(n,p)$, $Y\sim B(m,p)$ indipendenti $\Rightarrow X+Y \sim B(n+m,p)$ (riproducibilità delle binomiali)
- $X\sim \mathcal{P}(\lambda)$, $Y\sim\mathcal{P}(\mu)$ indipendenti $\Rightarrow X+Y\sim\mathcal{P}(\lambda+\mu)$ (" " Poisson)
- $X\sim\Gamma(r,\lambda)$, $Y\sim\Gamma(s,\lambda)$ indipendenti $\Rightarrow X+Y\sim\Gamma(r+s,\lambda)$ (" " Gamma)

in particolare $X_1,\dots X_n \sim \mathcal{E}(\lambda) = \Gamma(1,\lambda)$ indip. $\Rightarrow X_1+\dots+X_n \sim \Gamma(n,\lambda)$

Dim: $X\sim\Gamma(r,\lambda)$, $Y\sim\Gamma(s,\lambda)$ indip $\quad f_X(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} 1\!\!1_{(0,+\infty)}(x)$, $z>0$

$$f_{X+Y}(z) = \int_{\mathbb{R}} f_X(x) f_Y(z-x)\,dx = \int \frac{\lambda^r}{\Gamma(r)}\frac{\lambda^s}{\Gamma(s)} x^{r-1}(z-x)^{s-1} e^{-\lambda x} e^{-\lambda(z-x)} 1\!\!1_{x>0} 1\!\!1_{z-x>0}\,dx$$
$$= \frac{\lambda^{r+s}}{\Gamma(r)\Gamma(s)} \int_0^z x^{r-1}(z-x)^{s-1}\,dx \; e^{-\lambda z}\; 1\!\!1_{z>0}$$

$$\begin{aligned} &x = zu \\ &dx = z\,du \end{aligned} \quad = \frac{\lambda^{r+s}}{\Gamma(r)\Gamma(s)} \int_0^1 z^{r-1} u^{r-1} z^{s-1} (1-u)^{s-1} z\,du \;\; e^{-\lambda z}\, \mathbb{1}_{z>0}$$

$$= \frac{\lambda^{r+s}}{\Gamma(r)\Gamma(s)} \underbrace{\int_0^1 u^{r-1}(1-u)^{s-1}\,du}_{B(r,s)=\frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}} \; z^{r+s-1} e^{-\lambda z} \mathbb{1}_{z>0} = \frac{\lambda^{r+s}}{\Gamma(r+s)} z^{r+s-1} e^{-\lambda z} \mathbb{1}_{z>0}$$

V. a. gaussiana: standardizzazione

Prop: $X \sim \mathcal{N}(m, \sigma^2)$, $a, b \in \mathbb{R}$ con $a \neq 0 \Rightarrow aX+b \sim \mathcal{N}(am+b, a^2\sigma^2)$

Dim: $\left[f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-(x-m)^2/2\sigma^2\right) \right]$

Con formula di cambio variabile per $\varphi(x) = ax+b$, $|D\varphi^{-1}(y)| = \frac{1}{|a|}$: $Y = aX+b$ ha densità

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\left(\frac{y-b}{a} - m\right)^2 / 2\sigma^2\right) \frac{1}{|a|}$$

$$= \frac{1}{\sqrt{2\pi(a\sigma)^2}} \exp\left(-(y-(am+b))^2 / 2(a\sigma)^2\right)$$

Cor (standardizzazione): $X \sim \mathcal{N}(m, \sigma^2) \Leftrightarrow \frac{X-m}{\sigma} \sim \mathcal{N}(0,1)$

equivalentemente, $Z \sim \mathcal{N}(0,1) \Leftrightarrow \sigma Z + m \sim \mathcal{N}(m, \sigma^2)$

Calcoli con v.a. gaussiane: data $X \sim \mathcal{N}(m, \sigma^2)$, dati $-\infty \leq a < b \leq +\infty$, $P\{a \leq X \leq b\} = ?$

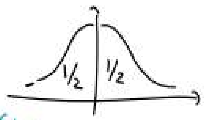
$P\{a \leq X \leq b\} = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx$ ma questo integrale non ha una forma esplicita (salvo per particolari valori di a e b)

- Se $X \sim \mathcal{N}(0,1)$, allora

$$P\{a \leq X \leq b\} = \Phi(b) - \Phi(a) \quad \text{con } \Phi \text{ FdR di } \mathcal{N}(0,1)$$

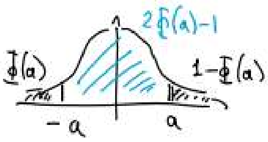
e, per molti $x \in \mathbb{R}$, $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$ è noto con ottima approssimazione (ed è presente nei software più comuni)

Proprietà di Φ:



- $\Phi(0) = \frac{1}{2}$, cioè $P\{X \leq 0\} = P\{X \geq 0\} = \frac{1}{2}$
- $\Phi(-x) = 1 - \Phi(x)$ $\forall x \in \mathbb{R}$

in particolare $P\{-a \leq X \leq a\} = 2\Phi(a) - 1$



- $\Phi(x) \approx 1$ per $x \geq 5$

$\Phi(x) \approx 0$ per $x \leq -5$

cioè $P\{-5 \leq X \leq 5\}$ è quasi 1.

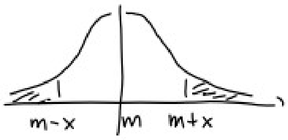
- $P\{-3 \leq X \leq 3\} \approx 0.997$

- Se $X \sim \mathcal{N}(m, \sigma^2)$, ci si riconduce a $\mathcal{N}(0,1)$ tramite standardizzazione

$$P\{a \leq X \leq b\} = P\Big\{\frac{a-m}{\sigma} \leq \underbrace{\frac{X-m}{\sigma}}_{\sim \mathcal{N}(0,1)} \leq \frac{b-m}{\sigma}\Big\} = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right)$$

In particolare

- $P\{X \leq m\} = P\{X \geq m\} = \frac{1}{2}$
- $P\{X \leq m-x\} = P\{X \geq m+x\} \quad \forall x \in \mathbb{R}$
- $P\{m-5\sigma \leq X \leq m+5\sigma\} \simeq 1$
- $P\{m-3\sigma \leq X \leq m+3\sigma\} \simeq 0.997$



Valore atteso, varianza e momenti di $X \sim \mathcal{N}(m, \sigma^2)$

$X = \sigma Z + m$, $Z \sim \mathcal{N}(0,1)$, quindi

- $E[X] = \sigma E[Z] + m = m$
- $Var(X) = \sigma^2 Var(Z) = \sigma^2$
- $E[|X-m|^p] = \sigma^p E[|Z|^p] = c_p \, Var(X)^{p/2} \quad \forall p > 0 \quad$ con $c_p = E[|Z|^p]$

(per una v.a. gaussiana, i momenti assoluti p-simi sono controllati dalla varianza)

Riproducibilità delle gaussiane:

Prop: $X \sim \mathcal{N}(m_1, \sigma_1^2)$, $Y \sim \mathcal{N}(m_2, \sigma_2^2)$ indipendenti $\Rightarrow X+Y \sim \mathcal{N}(m_1+m_2, \sigma_1^2+\sigma_2^2)$

Dim:

1. $X \sim \mathcal{N}(0,1)$, $Y \sim \mathcal{N}(0, \sigma^2)$ indip $\Rightarrow X+Y \sim \mathcal{N}(0, \sigma^2+1)$

$$f_{X+Y}(z) = \int f_X(x) f_Y(z-x)\,dx = \int \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(z-x)^2/2\sigma^2} dx$$

$$= \int \frac{1}{2\pi\sigma} \exp\left[-\frac{1}{2\sigma^2}\left((\sigma^2+1)x^2 - 2xz + z^2\right)\right] dx$$

$$= \int \frac{1}{2\pi\sigma} \exp\left[-\frac{1}{2\sigma^2}\left(\sqrt{\sigma^2+1}\,x - \frac{1}{\sqrt{\sigma^2+1}} z\right)^2 - \frac{1}{2\sigma^2} \cdot \frac{\sigma^2}{\sigma^2+1} z^2\right] dx$$

$$= \exp\left(-\frac{1}{2(\sigma^2+1)} z^2\right) \cdot \frac{1}{2\pi\sigma} \int \exp\left(-\frac{1}{2\sigma^2}\left(\sqrt{\sigma^2+1}\,x - \frac{1}{\sqrt{\sigma^2+1}} z\right)^2\right) dx$$

$y = \sqrt{\sigma^2+1}\,x - \frac{1}{\sqrt{\sigma^2+1}} z$

$dy = \sqrt{\sigma^2+1}\,dx$

$$= \exp\left(-\frac{1}{2(\sigma^2+1)} z^2\right) \cdot \frac{1}{2\pi\sigma} \cdot \frac{1}{\sqrt{\sigma^2+1}} \underbrace{\int \exp\left(-\frac{y^2}{2\sigma^2}\right) dy}_{= \sqrt{2\pi\sigma^2}}$$

$$= \exp\left(-\frac{1}{2(\sigma^2+1)} z^2\right) \cdot \frac{1}{\sqrt{2\pi(\sigma^2+1)}} \quad \text{densità } \mathcal{N}(0, \sigma^2+1)$$

2. Caso generale

$Z = \frac{X-m_1}{\sigma_1} \sim \mathcal{N}(0,1)$, $V = \frac{Y-m_2}{\sigma_1} \sim \mathcal{N}\left(0, \left(\frac{\sigma_2}{\sigma_1}\right)^2\right)$ indip., quindi

$$X+Y = m_1+m_2 + \sigma_1 \underbrace{(Z+V)}_{\sim \mathcal{N}\left(0, \left(\frac{\sigma_2}{\sigma_1}\right)^2+1\right)} \sim \mathcal{N}\left(m_1+m_2, \sigma_1^2\left(\frac{\sigma_2^2}{\sigma_1^2}+1\right) = \sigma_1^2+\sigma_2^2\right)$$

19
martedì 25 aprile 2023 18:12

Teoremi limite (LGN, TCL)

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

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

La def di conv. in probabilità si estende senza difficoltà dal caso discreto al caso gen.

Teor (Legge (debole) dei grandi numeri, LGN):

Sia $X_1, \dots X_n$, una successione di v.a. reali 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 $P\{|\bar{X}_n - m| > \varepsilon\} \to 0$ per $n \to \infty$, $\forall \varepsilon > 0$.

La dim. è la stessa del caso discreto

Oss: Come nel caso discreto, l'ipotesi di indip può essere sostituita con quella che le X_i siano a due a due scorrelate

Cor: Sia $X_1, \dots X_n, \dots$ una successione di v.a. i.i.d. a valori in S, con $(S, \mathcal{J})$ sp. mis.
Sia $\varphi : S \to \mathbb{R}$ boreliana con $E[\varphi(X_1)^2] < \infty$. Allora

$$\frac{1}{n}\sum_{i=1}^{n} \varphi(X_i) \xrightarrow{P} E[\varphi(X_1)] \qquad \text{per } n \to \infty$$

Dim: Basta applicare la LGN alle v.a. reali i.i.d. $\varphi(X_i)$.

Oss: In particolare, per $\varphi = \mathbb{1}_A$, $A \in \mathcal{J}$, ($\mathbb{1}_A(X) = \mathbb{1}_{X \in A} \sim B(p)$ con $p = P\{X \in A\}$), troviamo freq relativa di $\{X_i \in A\} = \frac{1}{n} \#\{i \mid X_i \in A\} \xrightarrow{P} P\{X \in A\}$

Applicazione: metodo Monte-Carlo per il calcolo approssimato degli integrali
Vogliamo calcolare $\int_a^b \varphi(x)dx$ per qualche $\varphi : (a,b) \to \mathbb{R}$ boreliana integrabile
assumiamo $\int_a^b |\varphi(x)|^2 dx < \infty$ (si potrebbe rimuovere queste ipotesi)

- notiamo che $\frac{1}{b-a}\int_a^b \varphi(x)dx = E[\varphi(U)]$ con $U \sim \mathcal{U}((a,b))$ (uniforme su (a,b))
- per la LGN, prese U_i i.i.d $\sim \mathcal{U}((a,b))$, $\frac{1}{n}\sum_{i=1}^{n} \varphi(U_i) \xrightarrow{P} E[\varphi(U)] = \frac{1}{b-a}\int_a^b \varphi(x)dx$

- quindi basta generare n v.a. U_i uniformi e calcolare
  $\frac{1}{n}\sum_{i=1}^{n} \varphi(U_i)$

Analogamente, $\int_{(0,1)^d} \varphi(x)dx$, con $\varphi: (0,1)^d \to \mathbb{R}$ boreliana t.c. $\int_{(0,1)^d} |\varphi(x)|^2 dx < \infty$,
può essere approssimato con $\frac{1}{n}\sum_{i=1}^{n} \varphi(U_i)$, con U_i i.i.d uniformi su $(0,1)^d$ (di densità $\mathbb{1}_{(0,1)^d}$)

Notare che (direttamente)
$$P\left\{ \left| \frac{1}{n}\sum_{i=1}^{n} \varphi(U_i) - \int_{(0,1)^d} \varphi(x)dx \right| > \varepsilon \right\} \leq \frac{1}{\varepsilon^2 n} Var(\varphi(U_1)) = O_{\varepsilon,\varphi}\left(\frac{1}{n}\right)$$ non dipende direttamente dalla dim. d

Teor (Teorema centrale del limite, TCL/TLC)

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

Allora
$$\frac{\sum_{i=1}^{n} X_i - nm}{\sqrt{n}\,\sigma} = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - m)$$ converge in legge a $Z \sim \mathcal{N}(0,1)$, cioè
detta F_n la FdR di $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - m)$, Φ la FdR di $\mathcal{N}(0,1)$,
$$\lim_n F_n(x) = \Phi(x) \quad \forall x \in \mathbb{R}$$

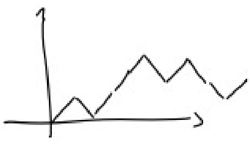
Oss: Da $P\{a \leq Z \leq b\} = F(b) - F(a)$, si ricava
$$P\left\{a \leq \frac{\sqrt{n}}{\sigma}(\bar{X}_n - m) \leq b\right\} \to P\{a \leq Z \leq b\} \quad \forall a,b \in \mathbb{R},\ a<b$$
e analogamente per le disuguaglianze strette

Oss Si può dimostrare che vale anche $E\left[\varphi\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - m)\right)\right] \to E[\varphi(Z)] \quad \forall \varphi \in C_b(\mathbb{R})$

Oss: Per standardizzazione, $\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n} X_i - nm\right) = \sqrt{n}(\bar{X}_n - m) \xrightarrow{\text{in legge}} Z_\sigma$, con $Z_\sigma \sim \mathcal{N}(0,\sigma^2)$
cioè $P\{a \leq \sqrt{n}(\bar{X}_n - m) \leq b\} \to P\{a \leq Z_\sigma \leq b\} \quad \forall a,b \in \mathbb{R},\ a<b$

Applicazione: convergenza delle leggi marginali di una passeggiata aleatoria

$(X_n)_{n \in \mathbb{N}^+}$ i.i.d. Rademacher, cioè $P\{X_i = 1\} = P\{X_i = -1\} = \frac{1}{2}$
$S_n = X_1 + \dots + X_n$ passeggiata aleatoria simmetrica



consideriamo, per $t>0$, $S_{[nt]}$
(comportamento a tempi grandi / passeggiata a step temporale $\frac{1}{n}$)

Per il TCL, per $n \to \infty$
$$\frac{1}{\sqrt{n}} S_{[nt]} = \sqrt{\frac{[nt]}{n}} \cdot \frac{1}{\sqrt{[nt]}} S_{[nt]} \xrightarrow{\text{in legge}} \sqrt{t}\, \underset{\sim\mathcal{N}(0,1)}{Z} \overset{(d)}{=} W_t \quad \text{con } W_t \sim \mathcal{N}(0,t)$$
questo è la convergenza delle leggi marginali: a t fissato, $\frac{1}{\sqrt{n}} S_{[nt]} \xrightarrow{\text{in legge}} W_t$

In realtà vale anche la convergenza come processo, cioè la convergenza delle leggi congiunte

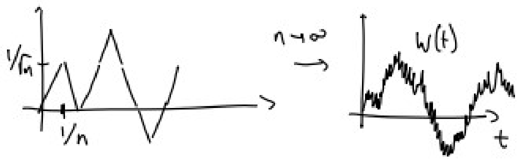
$$\left(\frac{1}{\sqrt{n}} S_{[nt_1]}, \dots \frac{1}{\sqrt{n}} S_{[nt_n]}\right) \overset{\text{"in legge"}}{\longrightarrow} (W_{t_1}, \dots W_{t_n}) \qquad \forall\, t_1 < t_2 < \dots < t_n$$

per un opportuno processo stocastico (= famiglia di v.a. indicizzate da t) W, noto come moto browniano o processo di Wiener

Notiamo

1. se lo step temporale scala come $\frac{1}{n}$, lo step spaziale scala come $\frac{1}{\sqrt{n}}$

   cioè $\delta W \approx \sqrt{\delta t}$

   ci aspettiamo allora che W sia $\approx \frac{1}{2}$-Hölder in t

   in effetti, è $\left(\frac{1}{2} - \varepsilon\right)$-Hölder $\forall \varepsilon > 0$, ma non $\frac{1}{2}$-Hölder

2. la densità di W_t è $f_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/(2t)}$ $\forall t > 0$, e soddisfa

   $$\partial_t f(t,x) = \frac{1}{2} \partial_x^2 f(t,x) \qquad \forall (t,x) \in (0,+\infty) \times \mathbb{R}$$

   equazione della diffusione o equazione del calore

20

domenica 30 aprile 2023 15:18

Statistica descrittiva

Statistica descrittiva: analisi dei dati $x_1, \dots x_n$ (generalmente ottenuti da ripetizioni di un esperimento) e della loro distribuzione, senza l'interpretazione di un modello probabilistico (cioè non ci interessa la distribuzione di prob. associata all'esperimento)

Ad es: lanciamo 10 volte una moneta, studiamo la distrib del n° di teste tra i 10 lanci
considerando 20 individui, studiamo la distribuzione dell'altezza tra questi 20 individui

Supponiamo qui che i dati siano quantitativi, cioè $x_i \in \mathbb{R}$ (o $\mathbb{R}^d$)

Alcuni indici statistici (funzioni di $x_1, \dots x_n$) rilevanti

- media campionaria o empirica
  $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$
  indice di centralità (indica il centro della distrib)
- mediana campionaria o empirica:
  ordinati i dati in senso crescente, il dato centrale se n è dispari, la media dei due dati centrali se n è pari.
  indice di centralità
- varianza campionaria
  $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$
  indice di dispersione (indica la dispersione dei dati)

Oss: · la media campionaria è il valore atteso di una v.a. con distrib. uniforme sugli x_i
· $\frac{n-1}{n} s^2$ è la varianza di una v.a con distrib uniforme sugli x_i

Per coppie di dati $(x_1, y_1), \dots (x_n, y_n)$, ci interessa la distribuzione congiunta degli (x_i, y_i)

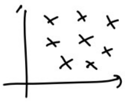
Ad es: dati 20 individui, x_i = altezza, y_i = peso dell' i-simo individuo

Alcuni indici statistici rilevanti:

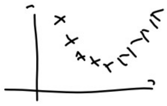
- coefficiente di correlazione campionario
  $$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\left(\sum_{i=1}^{n} (x_i - \bar{x})^2\right)^{1/2} \left(\sum_{i=1}^{n} (y_i - \bar{y})^2\right)^{1/2}} \in [-1, 1]$$

indice quanto i dati sono allineati

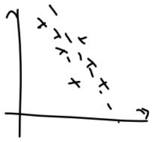
- $|r| \approx 0$: i dati sono poco allineati



o



- $|r| \approx 1$: i dati sono molto vicini ad una retta



- retta di regressione campionaria

$y = \alpha^* x + \beta^*$ con

$$\alpha^* = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad \beta^* = \bar{y} - \alpha \bar{x}$$

è la retta che meglio approssima i dati

21

domenica 30 aprile 2023 17:05

Teoria degli stimatori

Statistica inferenziale: a partire dai dati $x_1, \dots x_n$ ottenuti da un esperimento ripetuto, ricavare informazioni sulla distribuzione di probabilità associata a quell'esperimento

Esempio: lanciamo una moneta n volte, ottenendo risultati $x_1, \dots x_n$, ($x_i = \begin{cases} 1 & \text{se esce testa all'i-simo lancio} \\ 0 & \text{altrimenti} \end{cases}$)
vogliamo avere info sulla prob p di testa (non nota):

1) stima per p? $\to$ stimatore ($\bar{x}$ = freq relativa di "teste")
2) intervallo "ragionevole" per p? $\rightsquigarrow$ intervalli di fiducia
3) dato $p_0 \in (0,1)$, l'ipotesi $p = p_0$ è "ragionevole" (sulla base dei dati)? $\to$ test di ipotesi

Altri esempi:

- date le altezze di n persone (adulte di sesso maschile in Italia), avere info sulla distribuzione delle altezze su tutta la popolazione (persone adulte di sesso maschile in Italia)
- dati i risultati di un sondaggio sul gradimento del governo tra n persone, avere info sul gradimento del governo tra tutta la popol.
- dati gli esiti di un certo farmaco su n persone, avere info sulla distrib. degli esiti del farmaco su un individuo generico

Qui ci occupiamo di statistica (inferenziale) parametrica: supponiamo che la distrib. dell'esperimento sia nota a meno di uno o più parametri (o comunque siamo interessati a determinare uno o più parametri incogniti).

Def: Chiamiamo modello statistico (parametrico) una terna $(S, \mathcal{S}, (Q_\theta)_{\theta \in \Theta})$ con

- $(S, \mathcal{S})$ spazio misurabile
- Θ insieme dei parametri ($\neq \emptyset$)
- Q_θ prob su $(S, \mathcal{S})$, $\forall \theta \in \Theta$

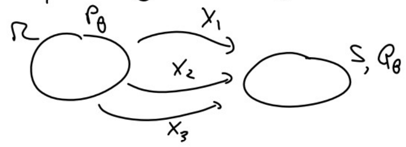
Q_θ rappresenta la distribuzione di un (carattere di) un esperimento aleatorio.

Esempi:

- lancio di moneta: $\theta = p$ = prob. di testa $\in \Theta = [0,1]$
  $(S, \mathcal{S}) = (\{0,1\}, \mathcal{P}(\{0,1\}))$ (con 1 = testa)
  $Q_\theta = B(\theta)$ Bernoulli di parametro θ
- carica elettrica di un corpo, supponiamo di distrib. gaussiana di media e varianza non note
  $\theta = (m, \sigma^2) \in \mathbb{R} \times (0, +\infty) = \Theta$
  $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$
  $Q_{(m,\sigma^2)} = \mathcal{N}(m, \sigma^2)$

Def: Un campione (i.i.d.) di taglia n e legge Q_θ è una famiglia $X_1, \dots X_n$ di v.a. i.i.d., a valori in S, di legge Q_θ, "al variare di $\theta \in \Theta$".

Più precisamente, un campione è una famiglia $X_1, \dots X_n$ con $X_i : \Omega \to S$ v.a., $(\Omega, \mathcal{F})$ spazio mis., tali che esista una famiglia di prob. $(P_\theta)_{\theta \in \Theta}$ su $(\Omega, \mathcal{F})$ tali che, $\forall \theta \in \Theta$: sotto P_θ $X_1, \dots X_n$ siano i.i.d. di legge Q_θ



Tipicamente, se Q_θ è la distribuzione di un esperimento aleatorio, $X_1, \dots X_n$ rappresentano gli esiti di n ripetizioni dell'esperimento.

Esempi:

- n lanci di moneta

$$X_i = \begin{cases} 1 & \text{se esce testa all'i-simo lancio} \\ 0 & \text{altrimenti} \end{cases}$$

$\Omega = \{0,1\}^n$, $\mathcal{F} = \mathcal{P}(\Omega)$, $P_\theta = Q_\theta^{\otimes n}$, cioè $p_\theta(x_1, \dots x_n) = q_\theta(x_1) \cdot \ldots \cdot q_\theta(x_n)$ (con p_θ, q_θ densità discrete di P_θ, Q_θ risp.)

- n misurazioni della carica elettrica di un corpo

X_i = esito della i-sima misurazione $X_i(\omega) = \omega_i$

$\Omega = \mathbb{R}^n$, $\mathcal{F} = \mathcal{B}(\mathbb{R}^n)$, P_θ misura con densità $f_\theta(x_1, \dots x_n) = g_\theta(x_1) \cdot \ldots \cdot g_\theta(x_n)$

(con $g_\theta = g_{(m, \sigma^2)}$ densità $\mathcal{N}(m, \sigma^2)$)

- dato un carattere di una popolazione, con distrib. Q_θ (ad es., n° di figli di un italiano)

X_i = carattere per l'i-simo individuo estratto dalla popolazione (n° di figli dell'i-simo individuo)

[spesso si identificano una popolazione e il suo carattere, quando ci interessa solo tale carattere]

Oss: Data una famiglia di prob $(Q_\theta)_{\theta \in \Theta}$ su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, tutte discrete o tutte assolutamente continue, esiste un campione $(X_1, \dots X_n)$ di legge Q_θ:

- $\Omega = \mathbb{R}^n$, $\mathcal{F} = \mathcal{B}(\mathbb{R}^n)$
- $X_i(\omega_1, \dots \omega_n) = \omega_i \quad i = 1, \dots n$
- P_θ : nel caso discreto: avente densità discreta $p_\theta(x_1, \dots x_n) = q_\theta(x_1) \cdot \ldots \cdot q_\theta(x_n)$ con q_θ densità discreta di Q_θ

  nel caso assol. cont: avente densità $f_\theta(x_1, \dots x_n) = g_\theta(x_1) \cdot \ldots \cdot g_\theta(x_n)$ con g_θ densità di Q_θ

Nel caso generale, $Q_\theta = P_\theta^{\otimes n}$ è la prob prodotto

Nella statistica inferenziale, osserviamo i risultati $x_1, \dots x_n$ di un campione $X_1, \dots X_n$ di legge Q_θ e vogliamo stimare il parametro incognito θ a partire da questi risultati.

Def: Chiamiamo statistica una v.a. funzione del campione $(X_1, .. X_n)$: $g(X_1, .. X_n)$

Def: Chiamiamo stimatore una statistica che non dipende direttamente dal parametro θ.
Scopo di uno stimatore è stimare $h(\theta)$, con $h: \Theta \to \mathbb{R}$ funzione data.

Esempi notevoli:

- media campionaria ($X_1, .. X_n$ campione i.i.d. di v.a. reali)
$$\bar{X} = \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$
come vedremo, è un "buon" stimatore del valore atteso $E[X_1]$

- varianza campionaria ($X_1, .. X_n$ campione i.i.d. di v.a. reali)
$$S^2 = S^2_{n-1} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \right)$$
($\approx$ media campionaria degli scarti quadratici da $\bar{X}$)
come vedremo, è un "buon" stimatore della varianza $Var(X_1)$

- dato un campione bivariato $(X_1, Y_1), .. (X_n, Y_n)$, ($X_i, Y_i$ v.a. reali, (X_i, Y_i) i.i.d. sotto P_θ)
coefficiente di correlazione campionario
$$r^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\left(\sum_{i=1}^{n} (X_i - \bar{X})^2 \cdot \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \right)^{1/2}}$$
è un "buon" stimatore del coeff. di correlazione tra X_i e Y_i $\rho(X_1, Y_1)$

Oss: In una sequenza di Bernoulli di n ripetizioni di esperimento successo/insuccesso, con p prob (incognita) di successo ($X_i = \begin{cases} 1 & \text{se successo all'i-sima prova} \\ 0 & \text{altrimenti} \end{cases}$)
$\bar{X}$ = freq. relativa del successo sul campione
$E[X_1] = p$ = prob del successo

Criteri per valutare la bontà di uno o più stimatori:

Def: Uno stimatore U è uno stimatore corretto, o non distorto (unbiased), di $h(\theta)$: se, $\forall \theta \in \Theta$, U è P^θ-integrabile e vale
$$E^\theta[U] = h(\theta) \quad \forall \theta \in \Theta$$
"la media di U su tutti i campioni possibili è il valore $h(\theta)$"

Esempi notevoli:

- Se $X_1, .. X_n$ è campione i.i.d. di v.a. reali con $E^\theta[|X_1|] < \infty$ $\forall \theta \in \Theta$, $(h(\theta) = E^\theta[X_1])$
$\bar{X}$ è uno stimatore corretto di $E^\theta[X_1]$: infatti
$$E^\theta[\bar{X}] = \frac{1}{n} \sum_{i=1}^{n} E^\theta[X_i] = E^\theta[X_1]$$

- Se $X_1, .. X_n$ è campione i.i.d. di v.a. reali con $E^\theta[X_1^2] < \infty$ $\forall \theta \in \Theta$, $(h(\theta) = Var^\theta(X_1))$,
S^2 è uno stimatore corretto di $Var^\theta(X_1)$: infatti

$$E^\theta[\bar{X}^2] = \frac{1}{n^2}\sum_{i,j=1}^{n} E^\theta[X_iX_j] = \frac{1}{n^2}\sum_{i=1}^{n} E^\theta[X_i^2] + \frac{1}{n^2}\sum_{i\neq j} E^\theta[X_i]E^\theta[X_j]$$
$$= \frac{1}{n}E^\theta[X_1^2] + \frac{n-1}{n}E^\theta[X_1]^2$$
$$E^\theta[S^2] = \frac{1}{n-1}\left(\sum_{i=1}^{n} E^\theta[X_i^2] - nE^\theta[\bar{X}^2]\right) = \frac{1}{n-1}\left(nE^\theta[X_1^2] - E^\theta[X_1^2] - (n-1)E^\theta[X_1]^2\right)$$
$$= E^\theta[X_1^2] - E^\theta[X_1]^2 = Var^\theta(X_1)$$

Sia $(S, \mathcal{S}, (Q_\theta)_{\theta\in\Theta})$ un modello e, $\forall n\in\mathbb{N}$, sia $X_1, \dots X_n$ un campione i.i.d. di taglia n e di legge Q_θ (è possibile def. tutte le X_i sullo stesso modello $(\Omega, \mathcal{F}, (P_\theta)_{\theta\in\Theta})$)

Def: Una successione di stimatori $U_n = g_n(X_1, \dots X_n)$, $n\in\mathbb{N}^+$, di $h(\theta)$ è asintoticamente non distorta se, $\forall n\in\mathbb{N}^+$, $\forall\theta\in\Theta$, U_n è P_θ-integrabile e
$$\lim_n E^\theta[U_n] = h(\theta) \quad \forall\theta\in\Theta$$

Def: Una successione di stimatori $U_n = g_n(X_1, \dots X_n)$, $n\in\mathbb{N}^+$, di $h(\theta)$ è consistente: se
$$\lim_n P^\theta\{|U_n - h(\theta)| > \varepsilon\} = 0 \quad \forall\varepsilon>0 \quad \forall\theta\in\Theta$$
cioè $(U_n)_n$ converge in P^θ-prob. a $h(\theta)$, $\forall\theta\in\Theta$

"per n grande, U_n è vicino con alta prob. a $h(\theta)$"

Esempi notevoli:

- Se $X_1, \dots X_n, \dots$ è un campione i.i.d. di v.a. reali con $E^\theta[X_1^2]<\infty$ $\forall\theta\in\Theta$, $(\bar{X}_n)_{n\in\mathbb{N}}$ è una successione di stimatori consistenti di $E^\theta[X_1]$:
  infatti per LGN, $\bar{X}_n \xrightarrow{P_\theta} E^\theta[X_1]$ $\forall\theta\in\Theta$
- Se $X_1, \dots X_n, \dots$ è un campione i.i.d. di v.a. reali con $E^\theta[X_1^4]<\infty$ $\forall\theta\in\Theta$, $(S^2_{n-1})_{n\in\mathbb{N}}$ è una successione di stimatori consistenti di $Var^\theta(X_1)$
  infatti per LGN,
  $$S^2_{n-1} = \underbrace{\frac{n}{n-1}}_{\to 1}\left(\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i^2}_{\xrightarrow{P_\theta} E[X_1^2]} - \underbrace{\bar{X}_n^2}_{\xrightarrow{P_\theta} E[X_1]}\right) \xrightarrow{P_\theta} E^\theta[X_1^2] - E^\theta[X_1]^2 = Var(X_1)$$
  (dove abbiamo usato che la convergenza in prob. è stabile per somma e prodotto)

Def: Dato U stimatore di $h(\theta)$, rischio quadratico di U
$$R_\theta(U) = E^\theta[(U-h(\theta))^2]$$
Dati U e V stimatori di $h(\theta)$, diciamo che U è preferibile a V: se $R_\theta(U)\le R_\theta(V)$ $\forall\theta\in\Theta$

"U è meno incerto di V"

Oss: Se U è corretto, $R_\theta(U) = Var^\theta(U)$

Oss: $R_\theta(\bar{X}_n) = Var^\theta(\bar{X}_n) = \frac{1}{n}Var^\theta(X_1) \downarrow 0$ per $n\to\infty$.

22
mercoledì 3 maggio 2023 19:12

Stimatori di massima verosimiglianza

Sia $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (Q_\theta)_{\theta\in\Theta})$ un modello statistico
- discreto: Q_θ discreta con densità discreta $m_\theta = p_\theta$, $\forall\theta\in\Theta$
- oppure assolutamente continuo: Q_θ assol continua con densità $m_\theta = f_\theta$, $\forall\theta\in\Theta$

Def. Funzione di verosimiglianza: $L: \Theta\times\mathbb{R}^n \to \mathbb{R}$ data da
$$L(\theta, x_1, \dots x_n) = L_\theta(x_1,\dots x_n) = m_\theta(x_1)\cdot\ldots\cdot m_\theta(x_n)$$

Oss: • Nel caso discreto, L_θ è la densità discreta congiunta di un campione $\overset{\text{i.i.d.}}{\vee}$ $X_1,\dots X_n$ di legge Q_θ
$$L_\theta(x_1,\dots x_n) = P_\theta\{X_1=x_1\}\cdot\ldots\cdot P_\theta\{X_n=x_n\} = P_\theta\{X_1=x_1,\dots X_n=x_n\}$$
• Nel caso assol. cont., L_θ è la densità congiunta di un campione $\overset{\text{i.i.d}}{\vee}$ $X_1\dots X_n$ di legge Q_θ

Sia $(X_1 \dots X_n)$ un campione i.i.d. di legge Q_θ

maximum likelihood estimator

Def: Uno stimatore U è detto stimatore di massima verosimiglianza di θ (MLE in inglese): se
$$L_U(X_1,\dots X_n) = \sup_{\theta\in\Theta} L_\theta(X_1,\dots X_n)$$
(cioè $L_{U(\omega)}(X_1(\omega),\dots X_n(\omega)) = \sup_{\theta\in\Theta} L_\theta(X_1(\omega),\dots X_n(\omega))$ $\forall\omega\in\Omega$, si assume implicitamente $U(\omega)\in\Theta$)

Poiché $U = g(X_1,\dots X_n)$, trovare U di max. ver. equivale a trovare g t.c.
$$L_{g(x_1,\dots x_n)}(x_1,\dots x_n) = \sup_{\theta\in\Theta} L_\theta(x_1,\dots x_n) \quad \forall(x_1,\dots x_n)\in\mathbb{R}^n$$

Significato: • Caso discreto: se i dati del campione sono $(x_1,\dots x_n)$, scegliamo la stima $g(x_1,\dots x_n)$ che massimizza $L_\theta(x_1,\dots x_n)$ = prob. P_θ di ottenere $X_1=x_1,\dots X_n=x_n$
• Caso assol cont: scegliamo la stima $g(x_1,\dots x_n)$ che massimizza $L_\theta(x_1,\dots x_n)$ "$\approx$ prob di ottenere dati $X_1\approx x_1,\dots X_n\approx x_n$ $\cdot\, \delta x_1\cdot\ldots\cdot\delta x_n$"

Esempio 1.: $Q_\theta = B(\theta)$ (Bernoulli di par. θ), $\theta\in[0,1]$
$$m_\theta(x) = \begin{cases}1-\theta & \text{se } x=0\\ \theta & \text{se } x=1\end{cases} = \theta^x(1-\theta)^{1-x} \qquad x\in\{0,1\}$$
$$L_\theta(x_1,\dots x_n) = \prod_{i=1}^n m_\theta(x_i) = \prod_{i=1}^n \theta^{x_i}(1-\theta)^{x_i}$$
$$= \theta^{\sum_{i=1}^n x_i}(1-\theta)^{\sum_{i=1}^n(1-x_i)} = \theta^{n\bar{x}}(1-\theta)^{n(1-\bar{x})}$$
con $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$
$$\log L_\theta(x_1\dots x_n) = n\bar{x}\log\theta + n(1-\bar{x})\log(1-\theta)$$
$$\frac{d}{d\theta}\log L_\theta(x_1\dots x_n) = \frac{n\bar{x}}{\theta} + \frac{n(1-\bar{x})}{1-\theta}$$
$\frac{d}{d\theta}\log L_\theta(x_1\dots x_n) = 0 \iff \theta = \bar{x}$ e $\log L_\theta(x_1\dots x_n) = -\infty$ per $\theta = 0,1$

quindi $L_\theta(x_1\dots x_n)$ ha un unico max in $\theta=\bar{x}$

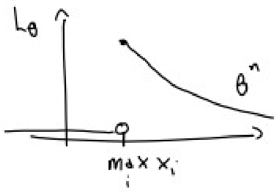
quindi $\bar{X}$ è lo stimatore di max. verosimiglianza di θ

Esempio 2: $Q_\theta = \mathcal{U}([0,\theta])$, $\theta>0$

$m_\theta(x) = \frac{1}{\theta}\mathbb{1}_{[0,\theta]}(x)$

$L_\theta(x_1,\dots x_n) = \prod_{i=1}^{n} m_\theta(x_i) = \frac{1}{\theta^n}\prod_{i=1}^{n}\mathbb{1}_{[0,\theta]}(x_i) = \frac{1}{\theta^n}\mathbb{1}_{0\le \min_i x_i \le \max_i x_i \le \theta}$

$= \mathbb{1}_{[0,+\infty)}(\min_i x_i)\ \frac{1}{\theta^n}\ \mathbb{1}_{[\max_i x_i,\, +\infty)}(\theta)$



L_θ ha max in $\theta = \max_i x_i$

Quindi $\max\{X_1,\dots X_n\}$ è lo stimatore di max ver. di θ

Oss: Usando $\theta = 2E[X_i]$, potremmo stimare θ anche con $2\overline{X}$.

Quale tra $\max_i X_i$ e $2\overline{X}$ è "meglio" come stimatore di θ? si dimostra

- $\max_i X_i$ è distorto, ma asintoticamente non distorto
- $\max_i X_i$ è consistente
- $R(\max_i X_i) \le R(2\overline{X})$ quindi $\max_i X_i$ è preferibile a $\overline{X}$

Def: Modello esponenziale: modello statistico $(\mathbb{R}, \mathcal{B}(\mathbb{R}), (Q_\theta)_{\theta\in\Theta})$, $\Theta\subseteq\mathbb{R}$ intervallo, t.c.

- caso discreto: $\exists\, T:\mathbb{N}\to\mathbb{R}$, $g:\mathbb{N}\to\mathbb{R}$ t.c. $\forall\theta\in\Theta$, Q_θ ha densità discreta su $\mathbb{N}$

  $p_\theta(k) = c_\theta\, g(k)\, e^{\theta T(k)}$ $\quad k\in\mathbb{N}$ $\quad$ ($c_\theta>0$ costante)

- caso assolutamente continuo: $\exists\, T:\mathbb{R}\to\mathbb{R}$, $g:\mathbb{R}\to\mathbb{R}$ boreliane t.c. $\forall\theta\in\Theta$, Q_θ ha densità

  $f_\theta(x) = c_\theta\, g(x)\, e^{\theta T(x)}$ $\quad x\in\mathbb{R}$ $\quad$ ($c_\theta>0$ costante)

Esempi:

1. leggi esponenziali: $f_\lambda(x) = \lambda e^{-\lambda x}\mathbb{1}_{(0,+\infty)}(x)$ $\quad\theta=\lambda$, $T(x)=-x$, $g(x)=\mathbb{1}_{(0,+\infty)}(x)$, $c_\lambda=\lambda$
2. leggi Poisson: $p_\lambda(k) = \frac{\lambda^k}{k!}e^{-\lambda}$ $\quad\theta=\log\lambda$, $T(k)=k$, $\quad g(k)=\frac{1}{k!}$, $c_\theta=e^{-\lambda}$ $\quad(\lambda^k = e^{k\log\lambda})$
3. leggi gaussiane $N(m,\sigma^2)$ ($\sigma^2=1$ per semplicità): $f_m(x) = \frac{1}{\sqrt{2\pi}}e^{-(x-m)^2}$ $\quad(e^{-(x-m)^2} = e^{-x^2}e^{2xm}e^{-m^2})$

   $\theta=m$, $T(x)=2x$, $g(x)=e^{-x^2}$, $c_m = \frac{1}{\sqrt{2\pi}}e^{-m^2}$
4. leggi geometriche $p_\lambda(k) = p(1-p)^{k-1}\mathbb{1}_{k\ge1}$ $\quad\theta=\log(1-p)$ $\quad T(k)=k-1$, $g(k)=\mathbb{1}_{k\ge1}$, $c_\theta=p$
5. leggi unif. su $(0,\theta)$, $f_\theta(x)=\frac{1}{\theta}\mathbb{1}_{[0,\theta]}(x)$, non è modello esponenziale

Teor: Sia $(\mathbb{R}, \mathcal{B}(\mathbb{R}), (Q_\theta)_{\theta\in\Theta})$ modello statistico t.c.

- $\forall\,\theta_1\ne\theta_2$, $Q_{\theta_1}\ne Q_{\theta_2}$
- $\Theta\subseteq\mathbb{R}$ è intervallo aperto
- $(Q_\theta)_{\theta\in\Theta}$ è modello esponenziale assol. cont.: $f_\theta(x) = c_\theta\, g(x)\, e^{\theta T(x)}$
- $x\mapsto g(x)T(x)^2e^{\theta T(x)}$ è integrabile $\forall\theta\in\Theta$

Sia $(X_1,\dots X_n)$ un campione i.i.d. di legge Q_θ. Supponiamo inoltre che

- esiste uno stimatore U_n di max verosimiglianza (a valori in Θ)

Allora lo stimatore U_n di max verosimiglianza (è unico ed) è consistente.

La dim si basa su
- legame tra U_n e "funzione di partizione" $\theta \mapsto c_\theta$
- LGN

Dim:

$$\psi(\theta) = -\log c_\theta = \log \int_{\mathbb{R}} g(x) e^{\theta T(x)} dx \quad \left(\int_{\mathbb{R}} c_\theta g(x) e^{\theta T(x)} dx = 1 \text{ da cui } \frac{1}{c_\theta} = \int_{\mathbb{R}} g(x) e^{T(x)} dx \right)$$

$$\psi'(\theta) = \frac{1}{\int g(x) e^{\theta T(x)} dx} \int g(x) T(x) e^{\theta T(x)} dx = c_\theta \int T(x) g(x) e^{\theta T(x)} dx = \int T(x) f_\theta(x) dx = E_\theta[T(X_1)]$$

$$\psi''(\theta) = -\frac{1}{\left(\int g(x) e^{\theta T(x)} dx\right)^2} \left(\int g(x) T(x) e^{\theta T(x)} dx \right)^2 + \frac{1}{\int g(x) e^{\theta T(x)} dx} \cdot \int g(x) T(x)^2 e^{\theta T(x)} dx$$

$$= -\left(\int T(x) f(x) dx \right)^2 + \int T(x)^2 f(x) dx = -E_\theta[T(X_1)]^2 + E_\theta[T(X_1)^2] = Var_\theta(T(X_1)) \geq 0$$

quindi ψ è convessa.

Mostriamo che $\psi'' > 0$. Per assurdo: $\exists \theta_0 \in \Theta$, $\psi''(\theta_0) = 0$, Quindi $T(X_1) =$ costante P_{θ_0}-q.c, cioè $T =$ costante Q_{θ_0}-q.c (cioè $\exists t \in \mathbb{R}$, $Q_{\theta_0}\{T = t\} = P_{\theta_0}\{T(X_1) = t\} = 1$)

Poiché Q_θ ha densità $c_\theta g e^{\theta T}$, deve essere $T =$ costante Lebesgue-q.o su $\{g > 0\}$ (cioè $\exists t \in \mathbb{R}$, t.c. $\{T \neq t, g > 0\}$ ha misura di Lebesgue nulla).

Ma allora ogni Q_θ ha densità $c_\theta g(x) e^{\theta t} = c\, g(x)$ indipendente (q.o.) da θ e quindi tutte le Q_θ coincidono $\lightning$.

Quindi ψ' è invertibile su Θ

$$L_\theta(x_1, \dots x_n) = f_\theta(x_1) \cdot \ldots \cdot f_\theta(x_n) = c_\theta^n g(x_1) \cdot \ldots \cdot g(x_n) e^{\theta(T(x_1) + \ldots + T(x_n))}$$

$$\log L_\theta(x_1, \dots x_n) = n\left[-\psi(\theta) + \theta \frac{1}{n} \sum_{i=1}^n T(x_i) \right] + \log(g(x_1) \cdot \ldots \cdot g(x_n))$$

$$\frac{d}{d\theta} \log L_\theta(x_1, \dots x_n) = n\left[-\psi'(\theta) + \frac{1}{n} \sum_{i=1}^n T(x_i) \right]$$

Poiché Θ è intervallo aperto, se U_n stimatore di max ver esiste (a valori in θ), U_n soddisfa

$$\frac{d}{d\theta} \log L_{U_n}(X_1, \dots X_n) = 0 \quad (\forall \omega \in \Omega)$$

quindi $\psi'(U_n) = \frac{1}{n} \sum_{i=1}^n T(X_i)$ da cui, per invertibilità di ψ' in Θ

$$U_n = (\psi')^{-1}\left(\frac{1}{n} \sum_{i=1}^n T(X_i) \right).$$

Per LGN, $\frac{1}{n} \sum_{i=1}^n T(X_i) \xrightarrow{P_\theta} E_\theta[T(X_1)] = \psi'(\theta) \quad \forall \theta$

quindi, per continuità di $(\psi')^{-1}$, $U_n = (\psi')^{-1}\left(\frac{1}{n} \sum_{i=1}^n T(X_i) \right) \xrightarrow{P_\theta} (\psi')^{-1}(\psi'(\theta)) = \theta$ (vedi lemma sotto)

Lemma: Y_n v.a. reali, $Y_n \xrightarrow{P} \ell$, $\varphi: \mathbb{R} \to \mathbb{R}$ continua $\Rightarrow \varphi(Y_n) \xrightarrow{P} \varphi(\ell)$

Dim:

$\forall \varepsilon > 0$, $\exists \delta > 0$, $|x - \ell| \leq \delta \Rightarrow |\varphi(x) - \varphi(\ell)| \leq \varepsilon$, quindi $\{|\varphi(X_n) - \varphi(\ell)| > \varepsilon\} \subseteq \{|X_n - \ell| > \delta\}$

$P\{|\varphi(X_n) - \varphi(\ell)| > \varepsilon\} \leq P\{|X_n - \ell| > \delta\} \to 0$ per $n \to \infty$.

Oss: Abbiamo usato il teor di deriv. sotto il segno di integrale:

data $F(t)=\int f(t,x)\,dx$, $t_0\in\mathbb{R}$ t.c.
- $x\mapsto f(t,x)$ è integrabile $\forall t\in$ intorno $J(t_0)$ di t_0
- $t\mapsto f(t,x)$ è differenziabile in $J(t_0)$ per q.o. x
- $\left|\frac{\partial}{\partial t}f(t,x)\right|\le g(x)$ con g integrabile, $\forall t\in J(t_0)$, per q.o. x

allora $\frac{d}{dt}F(t_0)=\int\frac{\partial f}{\partial t}(t_0,x)\,dx$

Oss: Il Teor. vale anche nel caso discreto, con enunciato del tutto analogo (l'ipotesi di integrabilità diventa $\sum_{k=0}^{\infty}g(k)T(k)^2e^{\theta T(k)}<\infty \quad \forall\theta\in\Theta$)

Oss: Il teor vale anche per $\Theta\subseteq\mathbb{R}^d$ aperto convesso, $T:\mathbb{R}\to\mathbb{R}^d$ boreliana:

$f_\theta(x)=c_\theta\, g(x)\exp(\theta\cdot T(x))$

Stimatori di massima verosimiglianza di media e varianza di gaussiana

$(S,\mathcal{S},(Q_\theta)_{\theta\in\Theta})=(\mathbb{R},\mathcal{B}(\mathbb{R}),\mathcal{N}(m,\sigma^2)_{m\in\mathbb{R},\,\sigma^2\in(0,+\infty)})\qquad(\Theta=\mathbb{R}\times(0,+\infty))$

$L_{(m,\sigma^2)}(x_1,\dots x_n)=f_{(m,\sigma^2)}(x_1)\cdot\ldots\cdot f_{(m,\sigma^2)}(x_n)=\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-m)^2\right)$

$\log L_{(m,\sigma^2)}(x_1\dots x_n)=-\frac{n}{2}\log(2\pi)-n\log\sigma-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-m)^2$

$$\left.\begin{aligned}&\frac{\partial}{\partial m}\log L_{(m,\sigma^2)}(x_1\dots x_n)=\frac{1}{\sigma^2}\sum_{i=1}^n(x_i-m)\\&\frac{\partial}{\partial\sigma}\log L_{(m,\sigma^2)}(x_1\dots x_n)=-\frac{n}{\sigma}+\frac{1}{\sigma^3}\sum_{i=1}^n(x_i-m)^2\end{aligned}\right\}=0\iff\begin{cases}m=\bar{x}=\frac{1}{n}\sum_{i=1}^n x_i\\\sigma^2=\frac{1}{n}\sum_{i=1}^n(x_i-\bar{x})^2=\frac{n-1}{n}s^2\end{cases}$$

e si verifica che $(\bar{x},\frac{n-1}{n}s^2)$ è un pt di max

Quindi $(\overline{X},\frac{n-1}{n}S^2)$ è lo stimatore di max verosimiglianza per (m,σ^2)

Oss: Con gli stessi conti si verifica, per $Q_\theta=\mathcal{N}(m,\sigma^2)$:
- se σ^2 è nota (cioè $\theta=m\in\Theta=\mathbb{R}$), $\overline{X}$ è lo stimatore di max verosimiglianza per m
- se m è nota (cioè $\theta=\sigma^2\in\Theta=(0,+\infty)$), $\frac{1}{n}\sum_{i=1}^n(X_i-m)^2$ è lo stimatore di max verosimiglianza per σ^2

23

lunedì 8 maggio 2023 20:49

Intervalli di fiducia

Esempio/motivazione:

n lanci di moneta, con p = prob di testa

$\bar{X}$ = freq relativa campionaria di teste $\left(= \frac{\#\text{teste}}{n}\right)$ è uno stimatore per p

vogliamo misurare l'incertezza della stima

intuitivamente, più n è grande minore dovrebbe essere l'incertezza

$(S, \mathcal{S}, (Q_\theta)_{\theta\in\Theta})$ modello statistico, $X_1, \dots X_n$ campione i.i.d. di legge Q_θ, def su $(\Omega, \mathcal{F}, (P_\theta)_\theta)$

Def: Dato $\alpha \in (0,1)$, una regione di fiducia a livello $1-\alpha$ per il parametro θ è un insieme aleatorio $D(\omega) \subseteq \Theta$, $\omega\in\Omega$ (cioè una mappa $\Omega \to \mathcal{P}(\Theta)$, $\omega \mapsto D(\omega)$) t.c.

$$P_\theta\{\theta \in D\} \geq 1-\alpha \quad \forall \theta\in\Theta$$

(dove $\{\theta\in D\} = \{\omega\in\Omega \mid \theta \in D(\omega)\} \subseteq \Omega$ e si sottointende che $\{\theta\in D\}\in\mathcal{F}$ $\forall\theta\in\Theta$)

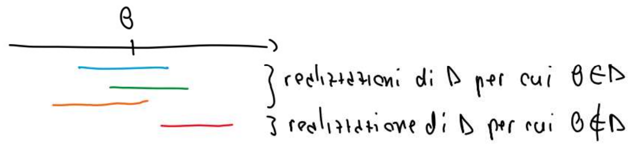
Oss: · α è piccolo, tipicamente $\alpha = 0.05$, 0.01

· D solitamente è una funzione del campione $X_1, \dots X_n$

Significato: dato un campione $X_1, \dots X_n$, possiamo determinare D t.c., con prob. alta, $\theta\in D$

Oss: nell'evento $\{\theta\in D\}$, l'aleatorietà è in D (che dipende da $X_1, \dots X_n$), non in θ, che è non noto ma deterministico

$P\{\theta\in D\} \geq 1-\alpha$ va inteso intuitivamente come: "in almeno una frazione $1-\alpha$ di tutti i campioni, θ cade in D(campione)"



Quantili di una prob. P su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, con FdR F

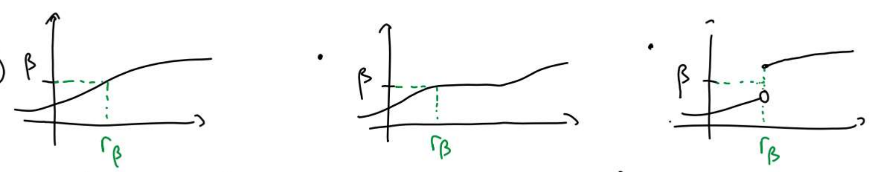
Def: Dato $\beta\in(0,1)$, quantile di ordine β (o β-quantile) di P (o di F): il numero

$r_\beta = \inf\{x\in\mathbb{R} \mid F(x) \geq \beta\}$

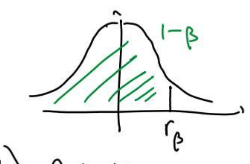
Oss: se F, ristretta ad un intervallo (a,b), è invertibile da (a,b) a $(0,1)$, allora $r_\beta = F^{-1}(\beta)$.

Esempi

· (F invertibile)



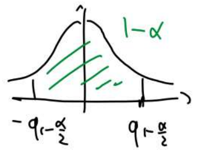
I quantili di $\mathcal{N}(0,1)$ si denotano con q_β o z_β



Per simmetria, · $q_{1-\beta} = -q_\beta$

· $P\{-q_{1-\frac{\alpha}{2}} \leq Z \leq q_{1-\frac{\alpha}{2}}\} = 1-\alpha$ per $Z\sim\mathcal{N}(0,1)$, $0<\alpha<1$

queste due proprietà valgono in generale se Z ha densità f pari, e più in generale se $Z \overset{(d)}{=} -Z$



Intervalli di fiducia per la media di una popol. normale (varianza nota)

- $(S, \mathcal{J}, (Q_\theta)_\Theta) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), (\mathcal{N}(m,\sigma^2))_{m\in\mathbb{R}}$ $(\theta = m \in \Theta \in \mathbb{R})$, σ^2 nota (sia $X \sim \mathcal{N}(m,\sigma^2)$))

  $(X_1, \dots X_n)$ campione i.i.d. di $\mathcal{N}(m,\sigma^2)$

  Poiché $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ è uno stimatore di m, è ragionevole cercare D intervallo centrato in $\bar{X}$,

  $$D = [\bar{X} - d, \bar{X} + d] =: [\bar{X} \pm d]$$

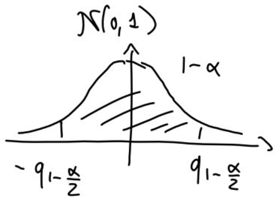
  Notiamo che, per riproducibilità e invarianza delle gaussiane per trasformazioni lineari, $\bar{X}$, combinazione lineare di gaussiane indip., è gaussiana, con $E[\bar{X}] = m$, $Var(\bar{X}) = \frac{\sigma^2}{n}$, cioè

  $\bar{X} \sim \mathcal{N}(m, \frac{\sigma^2}{n})$ (sotto P_m)

  cioè $\frac{\sqrt{n}}{\sigma}(\bar{X}-m) \sim \mathcal{N}(0,1)$ (sotto P_m)

  Usiamo questo fatto per calcolare $P_m\{m \in [\bar{X} \pm d]\}$

  $$P_m\{m \in [\bar{X} \pm d]\} = P_m\{|\bar{X}-m| \le d\} = P_m\left\{\frac{\sqrt{n}}{\sigma}|\bar{X}-m| \le \frac{\sqrt{n}}{\sigma} d\right\}$$
  $$= P\left\{|Z| \le \frac{\sqrt{n}}{\sigma} d\right\} \quad \text{con } Z \sim \mathcal{N}(0,1)$$
  $$= \Phi\left(\frac{\sqrt{n}}{\sigma} d\right) - \Phi\left(-\frac{\sqrt{n}}{\sigma} d\right) = 2\Phi\left(\frac{\sqrt{n}}{\sigma} d\right) - 1$$



  Imponiamo $P_m\{m \in [\bar{X} \pm d]\} = 1-\alpha$ (= per avere il più piccolo intervallo possibile)

  $2\Phi\left(\frac{\sqrt{n}}{\sigma} d\right) - 1 = 1 - \alpha$, cioè $\Phi\left(\frac{\sqrt{n}}{\sigma} d\right) = 1 - \frac{\alpha}{2}$

  cioè $\frac{\sqrt{n}}{\sigma} d = q_{1-\frac{\alpha}{2}}$ (con q_β quantile di $\mathcal{N}(0,1)$ di ordine β)

  Quindi $\left[\bar{X} \pm \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right]$ è un intervallo di fiducia per m di livello $1-\alpha$

Oss: L'ampiezza dell'intervallo, $2\frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}$,

- è crescente in $1-\alpha$
- è crescente in σ
- è decrescente in n

Esempio: Un metodo per misurare la carica elettrica di un corpo produce misure con distrib. gaussiana di media il valore vero della carica del corpo e dev. standard 0.1 (in Coulomb). Vengono effettuate 16 misurazioni, ottenendo una media campionaria di 5.2. Cerchiamo un intervallo di fiducia di livello 95% per la carica del corpo.

$(S, \mathcal{J}, (Q_\theta)) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{N}(m,\sigma^2))_m$, $\sigma = 0.1$, (X = carica elettrica rilevata $\sim \mathcal{N}(m,\sigma^2)$) $m \in \mathbb{R}$

X_i = risultato i-esima rilevazione $i = 1, \dots n = 16$ $\quad X_1, \dots X_n$ campione i.i.d. di $\mathcal{N}(m,\sigma^2)$

livello $1-\alpha = 0.95$, $\alpha = 0.05$, $q_{1-\frac{\alpha}{2}} = q_{0.975} \approx 1.96$

int. di fiducia $= \left[\bar{X} \pm \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right] \approx \left[\bar{X} \pm \frac{0.1}{4} \cdot 1.96\right] = [\bar{X} \pm 0.049]$

Dopo le misurazioni, con $\bar{x} = 5.2$, l'int. di fiducia risulta $[5.2 \pm 0.049] = [5.151, 5.249]$

Metodo della statistica pivotale: data una statistica pivotale, cioè una statistica $ST = g(\theta, X_1, \dots X_n)$

- ST è invertibile come funzione di θ, dato il campione:
  $\forall (x_1, \dots x_n)$, $\theta \mapsto g(\theta, x_1, \dots x_n)$ è invertibile
- ST ha legge che non dipende da θ: $ST_\# P_\theta$ è indipendente da θ (cioè $P_\theta\{ST \in A\}$ non dip. da θ, $\forall A \in \mathcal{B}(\mathbb{R})$)

allora una regione di fiducia di livello $1-\alpha$ per θ è data da $D = g(\cdot, X_1, \dots X_n)^{-1}(A)$, dove A (deterministica) è t.c. $P\{ST \in A\} = 1-\alpha$: infatti

$$P\{\theta \in D\} = P\{ST \in g(\cdot, X_1, \dots X_n)(D)\} = 1-\alpha$$

Nell'esempio precedente (media di popol. normale), $ST = \frac{\sqrt{n}}{\sigma}(\bar{X} - m) \sim N(0,1)$ sotto P_m

Def: Dato $\alpha \in (0,1)$, una regione di fiducia asintotica (o approssimata) di livello $1-\alpha \in (0,1)$ per θ è una successione di insiemi aleatori $D_n(\omega) \subseteq \Theta$, $\omega \in \Omega$, $n \in \mathbb{N}$, t.c.

$$\liminf_n P_\theta\{\theta \in D_n\} \geq 1-\alpha \quad \forall \theta \in \Theta$$

Intervalli di fiducia asintotici per la media di una popolazione (var. nota), grandi campioni

- $(S, \mathcal{S}, (Q_\theta)_{\theta \in \Theta}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), (Q_m)_{m \in \mathbb{R}})$, $m = E_m[X]$, $X \sim Q_m$, $\sigma^2 = Var(X)$ nota
  $X_1, \dots X_n$ campione i.i.d. di Q_m
  Per il TCL, se n è grande, $ST = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - m)$ è approx $\sim N(0,1)$, in particolare
  $$\lim_n P\{-q_{1-\frac{\alpha}{2}} \leq ST \leq q_{1-\frac{\alpha}{2}}\} = 1-\alpha$$
  Ripetendo i passaggi del caso popol. gaussiana (ST come statistica pivotale), si ottiene che
  $[\bar{X}_n \pm \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}]$ è un intervallo di fiducia asintotico di livello $1-\alpha$ per m

Intervalli di fiducia asintotici per una proporzione (media di una popol. Bernoulli), grandi campioni

- $(S, \mathcal{S}, Q_\theta)_{\theta \in \Theta}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), (B(p))_{p \in [0,1]})$ (o $(\{0,1\}, \mathcal{P}(\{0,1\}), (B(p))_{p \in [0,1]})$)
  $p = E_p[X]$ con $X \sim B(p)$ (p = prob del successo "= proporzione")
  $X_1, \dots X_n$ campione i.i.d. di $B(p)$ ($\bar{X}_n$ = freq. relativa campionaria del successo)
  Per il TCL, se n è grande, $\sqrt{\frac{n}{p(1-p)}}(\bar{X}_n - p)$ è approx $\sim N(0,1)$
  Problema: $p \mapsto \sqrt{\frac{n}{p(1-p)}}(\bar{X}_n - p)$ non è invertibile
  Cor: Date $X_1, \dots X_n$ i.i.d. $\sim B(p)$,
  $$\sqrt{\frac{n}{\bar{X}_n(1-\bar{X}_n)}}(\bar{X}_n - p) = \frac{n\bar{X}_n - np}{\sqrt{n\bar{X}_n(1-\bar{X}_n)}} \xrightarrow{\text{in legge}} Z \sim N(0,1) \quad \text{per } n \to \infty$$
  Questo risultato si dimostra come Corollario del TCL e di $\bar{X}_n \xrightarrow{\text{in prob.}} p$ (usando però fatti più avanzati)
  Quindi usando $ST = \sqrt{\frac{n}{\bar{X}_n(1-\bar{X}_n)}}(\bar{X}_n - p)$ come statistica pivotale (asintotica), si ottiene che
  $\left[\bar{X}_n \pm \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}\, q_{1-\frac{\alpha}{2}}\right]$ è un intervallo di fiducia asintotico di livello $1-\alpha$ per p

Intervalli di fiducia asintotici per la media di una popolazione, var. non nota, grandi campioni

- $(S, \mathcal{S}, (Q_\theta)_{\theta\in\Theta}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), (Q_m)_m)$ $m = E_m[X]$, $X \sim Q_m$, $X_1, \dots X_n$ campione i.i.d di Q_m

questa volta però supponiamo $\sigma^2 = Var(X)$ non nota

Problema: $\frac{\sqrt{n}}{\sigma}(\overline{X}_n - m)$ è funzione anche di σ non nota

Cor: Date $X_1, \dots X_n$ i.i.d. con $E[X_i^2] < \infty$,

$$\frac{\sqrt{n}}{S_n}(\overline{X}_n - m) \xrightarrow{\text{in legge}} Z \sim N(0,1) \quad \text{per } n \to \infty$$

con $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$

Questo risultato si dimostra come corollario del TCL e di $S_{n-1}^2 \to \sigma^2$ (usando fatti più avanzati)

Usando $ST = \frac{\sqrt{n}}{S_n}(\overline{X}_n - m)$ come statistica pivotale (asintotica) si ottiene che

$[\overline{X}_n \pm \frac{\sqrt{n}}{S_n} q_{1-\frac{\alpha}{2}}]$ è un intervallo di fiducia asintotico di livello $1-\alpha$ per m

Esempio: Viene rilevata l'opinione di 100 persone (estratte a caso) su un certo partito politico. Di queste 100 persone, 25 mostrano gradimento per il partito. Fornire un intervallo di fiducia (approssimato) per la percentuale di persone che apprezzano il partito, di livello 95%.

$(S, \mathcal{S}, (Q_\theta)_\theta) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), B(p)_p)$ $\left(X = \begin{cases} 1 & \text{se la persona apprezza il partito} \\ 0 & \text{altrimenti} \end{cases} \sim B(p)\right)$

p = prob gradimento

$X_i = \begin{cases} 1 & \text{se l'i-sima persona estratta apprezza il partito} \\ 0 & \text{altrimenti} \end{cases}$, $i = 1, \dots n = 100$ (abbastanza grande), $X_1, \dots X_n$ campione i.i.d di $B(p)$

$1-\alpha = 0.95$, $q_{1-\frac{\alpha}{2}} = q_{0.975} \approx 1.96$

int. di fiducia approx per p $\left[\overline{X} \pm \sqrt{\frac{\overline{X}(1-\overline{X})}{n}}\, q_{1-\frac{\alpha}{2}}\right] \approx \left[\overline{X} \pm \frac{\sqrt{\overline{X}(1-\overline{X})}}{10} \cdot 1.96\right] = \left[\overline{X} \pm 0.196 \cdot \sqrt{\overline{X}(1-\overline{X})}\right]$

Dopo le rilevazioni, con $\bar{x} = \frac{25}{100} = 0.25$, l'int. di fiducia diventa $[0.25 \pm 0.196\sqrt{0.25 \cdot 0.75}] \approx [0.165, 0.335]$

Intervallo di fiducia per la media di una popolazione gaussiana, varianza non nota

- $(S, \mathcal{S}, (Q_\theta)_{\theta\in\Theta}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), N(m, \sigma^2)_{m\in\mathbb{R}})$, $X_1, \dots X_n$ campione i.i.d. di $N(m, \sigma^2)$

supponiamo σ^2 non nota (e n non necessariamente grande)

Usiamo:

$$ST = \frac{\sqrt{n}}{S}(\overline{X} - m) \sim t_{n-1} \quad \text{(t-di-Student a n-1 gradi di libertà)}$$

(con $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$, $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$)

In particolare, chiamando $t_{\beta, n-1}$ il quantile di t_{n-1} di ordine β ($P\{\overset{\sim t_{n-1}}{T} \le t_{\beta, n-1}\} = \beta$),

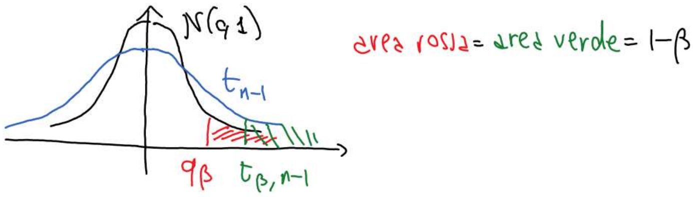
$$P\{ST \in [-t_{1-\frac{\alpha}{2}, n-1}, t_{1-\frac{\alpha}{2}, n-1}]\} = 1-\alpha$$

Quindi usando ST come statistica pivotale si ottiene che

$[\overline{X} \pm \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}, n-1}]$ è un intervallo di fiducia (esatto) di livello $1-\alpha$ per m

Oss: Poiché la densità di t_{n-1} è polinomiale, essa ha code più pesanti della densità $N(0,1)$

e quindi $t_{\beta, n-1} > q_\beta \quad \forall n, \ \forall \beta > \frac{1}{2}$



Dunque l'incertezza nel caso varianza non nota è maggiore rispetto al caso varianza nota

Oss: Si può dimostrare che, per $n \to \infty$, $t_{n-1} \xrightarrow{\text{in legge}} \mathcal{N}(0,1)$, in particolare $t_{\beta, n-1} \to q_\beta$, $\forall \beta \in (0,1)$.

Oss: È possibile considerare anche semirette come regioni di fiducia (intervalli unilateri)
Ad es., data una popolazione $\mathcal{N}(m, \sigma^2)$ di varianza σ^2 nota, possiamo cercare una regione di fiducia per la media m della forma $(-\infty, \overline{X}+d]$. Imponiamo $P_m\{m \in (-\infty, \overline{X}+d]\} = 1-\alpha$ e usiamo la statistica pivotale $ST = \frac{\sqrt{n}}{\sigma}(\overline{X}-m)$

$$P_m\{m \in (-\infty, \overline{X}+d]\} = P_m\{\overline{X}+d \geq m\} = P_m\left\{\frac{\sqrt{n}}{\sigma}(\overline{X}-m) \geq -\frac{\sqrt{n}}{\sigma} d\right\}$$
$$= \Phi\left(\frac{\sqrt{n}}{\sigma} d\right)$$
$$= 1-\alpha \iff \frac{\sqrt{n}}{\sigma} d = q_{1-\alpha}$$

e troviamo $(-\infty, \overline{X} + \frac{\sigma}{\sqrt{n}} q_{1-\alpha}]$ come una regione di fiducia di livello $1-\alpha$ per m

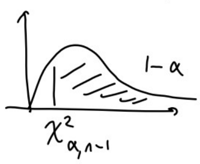
Intervalli di fiducia per la varianza di una popol. gaussiana,

- $(S, \mathcal{S}, (P_\theta)_{\theta \in \Theta}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), (\mathcal{N}(m, \sigma^2))_{\sigma^2 \in (0, +\infty)})$, $\sigma^2 = Var_{\sigma^2}(X)$ con $X \sim \mathcal{N}(m, \sigma^2)$

  $X_1, \dots X_n$ campione i.i.d.

  supponiamo m non nota

  Usiamo
  $$ST = (n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1} \quad \text{(chi-quadrato a } n-1 \text{ gradi di libertà)}$$
  In particolare, chiamando $\chi^2_{\beta, n-1}$ il quantile di ordine β di χ^2_{n-1} $(P\{Y \leq \chi^2_{\beta, n-1}\} = \beta)$, $Y \sim \chi^2_{n-1}$
  $$P\{ST \in [\chi^2_{\alpha, n-1}, +\infty)\} = 1-\alpha$$
  Quindi, usando ST come statistica pivotale $\left(ST \in [\chi^2_{\alpha, n-1}, +\infty) \iff \sigma^2 \in \left(0, \frac{(n-1)S^2}{\chi^2_{\alpha, n-1}}\right]\right)$
  si ottiene che
  $$\left(0, \frac{(n-1)S^2}{\chi^2_{\alpha, n-1}}\right]$$
  è un intervallo di fiducia (unilatero) per σ^2 di livello $1-\alpha$



Oss: Se m è nota, si usa la statistica pivotale
$$ST = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - m)^2 \sim \chi^2_n$$

24

mercoledì 10 maggio 2023 00:20

Distribuzioni legate alle v.a. gaussiane

$(\Omega, \mathcal{F}, P)$ sp di prob.

Prop: a) $Z \sim N(0,1) \Rightarrow Z^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2}) =: \chi^2_1$ (distribuzione chi-quadrato a un grado di libertà)

b) $Z_1, \dots Z_n \sim N(0,1)$ i.i.d. $\Rightarrow Z_1^2 + \dots + Z_n^2 \sim \Gamma(\frac{n}{2}, \frac{1}{2}) =: \chi^2_n$ (distrib chi-quadrato a n gradi di lib.)

Dim:

a) F_{Z^2} FdR di Z^2, dobbiamo mostrare $F_{Z^2}(u) = \int_{-\infty}^{u} f(v)dv$ con f densità $\Gamma(\frac{1}{2}, \frac{1}{2})$,

$$f(u) = \frac{1}{\sqrt{2}\,\Gamma(\frac{1}{2})} r^{-1/2} e^{-\frac{1}{2}u} \mathbb{1}_{(0,+\infty)}(u)$$

$F_{Z^2}(u) = 0 \quad \forall u < 0$. Per $u \geq 0$,

$$F_{Z^2}(u) = P\{Z^2 \leq u\} = P\{-\sqrt{u} \leq Z \leq \sqrt{u}\} = 2P\{0 \leq Z \leq \sqrt{u}\} = 2\int_0^{\sqrt{u}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

$$\begin{matrix} x^2 = v \\ 2x\,dx = dv \end{matrix} \qquad = \int_0^u \frac{1}{\sqrt{2\pi}} e^{-u/2} \frac{1}{\sqrt{v}} dv = \int_0^u f(v)dv$$

b) Z_i^2 sono indipendenti e $\sim \Gamma(\frac{1}{2}, \frac{1}{2})$. Quindi per riproducibilità di Γ, $Z_1^2 + \dots + Z_n^2 \sim \Gamma(\frac{n}{2}, \frac{1}{2})$.

Prop: $Z = (Z_1, \dots Z_n)^T$ vettore di $N(0,1)$ indip., $M \in \mathbb{R}^{n \times n}$ ortogonale $\Rightarrow MZ$ vettore di $N(0,1)$ indip.

Dim:

Notiamo che $Z_1, \dots Z_n$ sono i.i.d. $\sim N(0,1) \Leftrightarrow Z = (Z_1, \dots Z_n)$ ha densità

$$f_Z(x_1, \dots x_n) = \frac{1}{\sqrt{2\pi}} e^{-x_1^2/2} \cdot \dots \cdot \frac{1}{\sqrt{2\pi}} e^{-x_n^2/2} = \frac{1}{(2\pi)^{n/2}} e^{-|x|^2/2} \quad \left(\text{dove } |x|^2 = \sum_{i=1}^n x_i^2\right)$$

Se M è ortogonale (quindi $|Mx| = |x|$, $|\det M| = 1$), MZ ha densità

$$f_{MZ}(y_1, \dots y_n) = f_Z(M^{-1}y) |\det M^{-1}| = \frac{1}{(2\pi)^{n/2}} e^{-|M^{-1}y|^2/2} \frac{1}{|\det M|} = f_Z(y_1, \dots y_n)$$

quindi $MZ \overset{(d)}{=} Z$ è vettore di $N(0,1)$ indip.

Prop: $Z_1, \dots Z_n \sim N(0,1)$ indip., $\bar{Z} = \frac{1}{n}\sum_{i=1}^n Z_i$, $S^2 = \frac{1}{n-1}\sum_{i=1}^n (Z_i - \bar{Z})^2 \Rightarrow$

a) $\bar{Z}$, S^2 indip.

b) $\bar{Z} \sim N(0, \frac{1}{n})$, $(n-1)S_n^2 \sim \chi^2_{n-1}$

Dim:

a) Prendiamo $M \in \mathbb{R}^{n \times n}$ definita da ($e_1, \dots e_n$ base canonica di $\mathbb{R}^n$)

$$M^T e_1 = \frac{1}{\sqrt{n}}(1, \dots 1)^T = \frac{1}{\sqrt{n}}\sum_{i=1}^n e_i$$

$M^T e_2, \dots M^T e_n$ base ortonormale di $(M^T e_1)^\perp$

M^T e quindi M sono ortogonali.

Detta $Z = (Z_1, \dots Z_n)^T = \sum_{i=1}^n Z_i e_i$, MZ è vettore di $N(0,1)$ indip, in particolare

$(MZ)_1$ e $\sum_{i=1}^n (MZ)_i^2$ sono indipendenti

$$(MZ)_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i = \sqrt{n}\, \bar{Z}$$

$$\sum_{i=2}^{n} (MZ)_i^2 = |MZ|^2 - (MZ)_1^2 = |Z|^2 - n\bar{Z}^2 = \left(\sum_{i=1}^{n} Z_i^2\right) - n\bar{Z}^2 = \sum_{i=1}^{n} (Z_i - \bar{Z})^2 = (n-1) S_n^2$$

Quindi $\bar{Z}$ e S_{n-1}^2 sono indip.

b) $\bar{Z}$ è gaussiano per riproducibilità e $E[\bar{Z}]=0$, $Var(\bar{Z}) = \frac{1}{n}$

$(n-1) S_{n-1}^2 = \sum_{i=2}^{n} (MZ)_i^2$ è somma di $(n-1)$ v.a. $\mathcal{N}(0,1)$ indip, quindi $(n-1) S_{n-1}^2 \sim \chi^2_{n-1}$.

Def: Date $Z \sim \mathcal{N}(0,1)$, $Y \sim \chi^2_n$, con Z e Y indipendenti, chiamiamo distribuzione t-di-Student a n gradi di libertà (t_n) la legge di

$$T = \frac{\sqrt{n}\, Z}{\sqrt{Y}}$$

Cor: $Z_1, \dots Z_n \sim \mathcal{N}(0,1)$ i.i.d. $\Rightarrow \sqrt{n}\, \frac{\bar{Z}}{S} \sim t_{n-1}$

Dim:

$\sqrt{n}\, \frac{\bar{Z}}{S} = \frac{\sqrt{n-1} \cdot \sqrt{n}\, \bar{Z}}{\sqrt{(n-1) S^2}}$ e $\sqrt{n}\, \bar{Z} \sim \mathcal{N}(0,1)$, $(n-1) S^2 \sim \chi^2_{n-1}$ e sono indipendenti

Prop: t_n ha densità data da

$$f_{t_n}(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\, \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \qquad x \in \mathbb{R}$$

In particolare, f_{t_n} è pari.

Dim:

Dobbiamo verificare $F_{t_n}(x) = \int_{-\infty}^{x} f_{t_n}(x')\, dx'$

$$F_{t_n}(x) = \iint \mathbb{1}_{\sqrt{n} z/\sqrt{y} \le x}\, f_{\mathcal{N}(0,1)}(z)\, f_{\chi^2_n}(y)\, dy\, dz$$

$$= \iint \mathbb{1}_{\sqrt{n} z/\sqrt{y} \le x}\, c_n\, e^{-z^2/2}\, y^{-n/2} e^{-y/2}\, dy\, dz \qquad (\text{con } c_n > 0 \text{ costante})$$

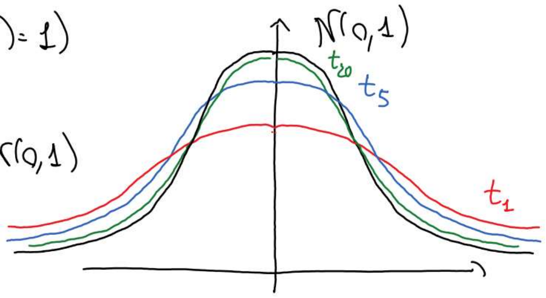
$\frac{\sqrt{n}\, z}{\sqrt{y}} = x'$, $\frac{\sqrt{n}}{\sqrt{y}} dz = dx'$

$$= \int_{-\infty}^{x} c_n' \int e^{-(x')^2 y/2n}\, e^{-y/2}\, y^{-(n-1)/2}\, dy\, dx' \qquad (\text{con } c_n' > 0 \text{ costante})$$

$\left(1 + \frac{(x')^2}{n}\right) y = \tilde{y}$, $\left(1 + \frac{(x')^2}{n}\right) dy = d\tilde{y}$

$$= \int_{-\infty}^{x} c_n' \left(1 + \frac{(x')^2}{n}\right)^{-\frac{n+1}{2}} \underbrace{\int e^{-\tilde{y}/2}\, \tilde{y}^{-\frac{n-1}{2}}\, d\tilde{y}}_{\text{cost.}}\, dx' = \int_{-\infty}^{x} c_n''\, f_{t_n}(x')\, dx'$$

e necessariamente $c_n'' = 1$ (cosicché $F(+\infty) = 1$)



Oss: • t_n ha code polinomiali, più "pesanti" delle code $\mathcal{N}(0,1)$

• per $n \to \infty$, "$t_n \to \mathcal{N}(0,1)$"

Prop: Siano $X_1, \dots X_n$ i.i.d. $\sim \mathcal{N}(m, \sigma^2)$, $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$, $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$. Allora

a) $\bar{X}$, S^2 sono indipendenti

b) $\overline{X} \sim \mathcal{N}(m, \frac{\sigma^2}{n})$, $(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$

c) $\sqrt{n}\,\frac{\overline{X}-m}{S} \sim t_{n-1}$

Dim: dai risultati precedenti e da standardizzazione.

25
giovedì 11 maggio 2023 01:02

Test statistici 1

Esempio/motivazione:

n lanci di moneta, con p = prob di testa

il banco dichiara che la moneta è equilibrata, cioè $p=\frac{1}{2}$

sulla base degli esiti degli n lanci, vogliamo verificare se l'ipotesi $p=\frac{1}{2}$ è plausibile o no:

- se su 1000 dati escono
  - 501 teste, l'ipotesi $p=\frac{1}{2}$ è plausibile
  - 900 " , " " " non plausibile
  - 540 " , " " " ?

Un test statistico è una procedura per verificare, sulla base dei dati $(x_1 .. x_n)$ del campione, se una certa ipotesi (=affermazione sul parametro) è plausibile o no.

$(S, \mathcal{S}, (Q_\theta)_{\theta\in\Theta})$ modello statistico, $(X_1, .. X_n)$ campione i.i.d. di legge Q_θ, def. su $(\Omega, \mathcal{F}, (P_\theta)_{\theta\in\Theta})$

Useremo spesso $(x_1, .. x_n)$ per indicare i dati ottenuti dal campione $(X_1, .. X_n)$

Elementi principali di un test (parametrico)

1) L'ipotesi da verificare: $\Theta = \Theta_0 \cup \Theta_1$ partizione di Θ $(\Theta_0 \cap \Theta_1 = \emptyset)$

   $\mathcal{H}_0 : \theta \in \Theta_0$ ipotesi nulla, di cui verificare se è plausibile o no

   $\mathcal{H}_1 : \theta \in \Theta_1$ " alternativa

   Nell'esempio precedente: $\mathcal{H}_0 : \theta = \frac{1}{2}$, $\Theta_0 = \{\frac{1}{2}\}$, $\Theta_1 = [0,1] \setminus \{\frac{1}{2}\}$

   Oss: asimmetria tra $\mathcal{H}_0$ e $\mathcal{H}_1$:

   - se l'esito del test, sulla base dei dati, è il rifiuto di $\mathcal{H}_0$, in favore di $\mathcal{H}_1$, allora $\mathcal{H}_0$ è poco plausibile, cioè c'è evidenza statistica per $\mathcal{H}_1$
   - se invece l'esito del test è l'accettazione di $\mathcal{H}_0$, allora $\mathcal{H}_0$ è plausibile, ma non c'è per forza evidenza per $\mathcal{H}_0$.

2) La regione critica o di rifiuto: un evento $C \in \mathcal{F}$, solitamente dipendente dal campione $X_1, .. X_n$, t.c.,
   - se l'esito ω cade in C, rifiutiamo l'ipotesi $\mathcal{H}_0$ in favore di $\mathcal{H}_1$
   - se l'esito ω non cade in C, accettiamo l'ipotesi $\mathcal{H}_0$.

   La scelta di C deve essere t.c.:
   - C sia poco probabile sotto $\mathcal{H}_0$ (poco plausibile)
   - se $\omega \in C$, allora questa è un'evidenza per $\mathcal{H}_1$

   Nell'esempio precedente ($\mathcal{H}_0: p=\frac{1}{2}$), ci aspettiamo una regione critica C della forma

   $$C = \{ |\bar{X} - \tfrac{1}{2}| > d \}$$ per un $d>0$ opportuno, per cui $P_{\frac{1}{2}}(C)$ sia piccola

   Se invece consideriamo $\mathcal{H}_0 : p \leq \frac{1}{2}$, ci aspettiamo una regione critica C della forma

   $$C = \{ \bar{X} - \tfrac{1}{2} > d \}$$ per un $d>0$ opportuno, per cui $P_p(C)$ sia piccola $\forall p \leq \frac{1}{2}$

   Oss: Nella scelta di C spesso si usa una statistica (statistica di test) ST, la cui legge non dipenda da θ $(C = \{ ST \in A \}$ per un opportuno $A)$

L'esito del test (che è aleatorio poiché dipende dai dati del campione) può essere errato, in due modi:

- errore di 1^a specie: $\mathcal{H}_0$ è vera ma l'esito del test è il rifiuto di $\mathcal{H}_0$
- " " 2^a specie: $\mathcal{H}_0$ è falsa ma l'esito del test è l'accettazione di $\mathcal{H}_0$

Def: Dato $\alpha \in (0,1)$ (piccolo, ed es. $\alpha = 0.05$, 0.01), un test è di livello α: se

$$\sup_{\theta \in \Theta_0} P_\theta(C) \leq \alpha$$

cioè la prob. dell'errore di 1^a specie (che dipende da $\theta \in \Theta_0$) è al massimo α.

Più α è piccolo, maggiore è l'evidenza statistica fornita dal test contro $\mathcal{H}_0$ in caso di rifiuto.

Def: Potenza di un test: funzione

$$\pi_C : \Theta_1 \to [0,1], \quad \pi_C(\theta) = P_\theta(C)$$

cioè $\pi_C(\theta)$ è la prob di rifiutare $\mathcal{H}_0$ quando il valore del parametro è $\theta \in \Theta_1$, quindi $\pi_C(\theta)$ rappresenta la capacità di accorgersi che l'ipotesi $\mathcal{H}_0$ è falsa

Nell'approccio "classico" ai test, si fissa un livello α piccolo e si cerca C in modo da massimizzare la potenza

Caso ipotesi semplice: $\Theta_0 = \{\theta_0\}$, $\Theta_1 = \Theta \setminus \{\theta_0\}$: legame tra test e intervalli di fiducia

- Dato $\alpha \in (0,1)$, se D è una regione di fiducia di livello $1-\alpha$ per θ, allora

  $$C = \{\theta_0 \notin D\} = \{\omega \in \Omega \mid \theta_0 \notin D(\omega)\}$$

  è una regione critica di livello α: infatti

  $$P_{\theta_0}(C) = 1 - P_{\theta_0}\{\theta_0 \in D\} \leq 1 - (1-\alpha) = \alpha$$

- Viceversa, dato $\alpha \in (0,1)$, se, $\forall \theta_0 \in \Theta$, C_{θ_0} è una regione critica di livello α per il test $\mathcal{H}_0 : \theta = \theta_0$, allora

  $$D(\omega) = \{\theta \in \Theta \mid \omega \notin C_\theta\}$$

  è una regione di fiducia per θ di livello $1-\alpha$: infatti

  $$P_{\theta_0}\{\theta_0 \in D\} = P_{\theta_0}(C_{\theta_0}^c) = 1 - P_{\theta_0}(C_{\theta_0}) \geq 1-\alpha \quad \forall \theta_0 \in \Theta$$

Test bilatero per la media di una popolazione gaussiana, varianza nota (test z)

- $(S, \mathcal{S}, (Q_\theta)_\Theta) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), N(m,\sigma^2)_{m \in \mathbb{R}})$, $\theta = m \in \Theta = \mathbb{R}$, σ^2 nota

  $X_1, \dots X_n$ campione i.i.d. di $N(m,\sigma^2)$

  Dato $m_0 \in \mathbb{R}$, consideriamo il test $\mathcal{H}_0 : m = m_0$ contro $\mathcal{H}_1 : m \neq m_0$, a livello α

  Regione critica C: poiché $\bar{X}$ è uno stimatore di m e $\mathcal{H}_1 : m \neq m_0$, è ragionevole cercare C della forma $C = \{|\bar{X} - m_0| > d\}$ per d opportuno

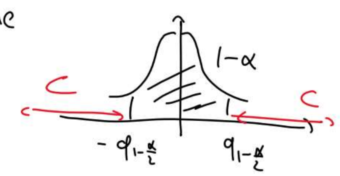
  Usiamo la statistica di test

  $$ST = \frac{\sqrt{n}}{\sigma}(\bar{X} - m) \sim N(0,1) \text{ sotto } P_m$$

e calcoliamo $P_{m_0}(C) = P_{m_0}\{\frac{\sqrt{n}}{\sigma}|\bar{X} - m_0| > \frac{\sqrt{n}}{\sigma} d\} = P\{|Z| > \frac{\sqrt{n}}{\sigma} d\}$ con $Z \sim \mathcal{N}(0,1)$

Imponiamo $P_{m_0}(C) = \alpha$, cioè $\frac{\sqrt{n}}{\sigma} d = q_{1-\frac{\alpha}{2}}$, otteniamo che

$C = \{\frac{\sqrt{n}}{\sigma}|\bar{X} - m_0| > q_{1-\frac{\alpha}{2}}\} = \{|\bar{X} - m_0| > \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\}$



è una regione critica di livello α

Oss: $\omega \in C \Leftrightarrow |\bar{X}(\omega) - m_0| > \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} \Leftrightarrow m_0 \in [\bar{X}(\omega) \pm \frac{\sigma}{\sqrt{n}} q_{1-\alpha}]$ (Int. di fiducia per m_0 di livello $1-\alpha$)

Test bilatero per una proporzione (media di popol. Bernoulli), grandi campioni (test Z approssimato)

- $(S, \mathcal{S}, (Q_\theta)_\Theta) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), B(p)_{p\in[0,1]})$ $\quad \theta = p \in \Theta = [0,1]$

  $X_1, \ldots X_n$ campione i.i.d. di $B(p)$, assumiamo n grande

  Dato $p_0 \in (0,1)$, consideriamo il test $H_0: p = p_0$ contro $H_1: p \neq p_0$ di livello approssimativamente α

  Regione critica C: è ragionevole cercare C della forma $C = \{|\bar{X} - p_0| > d\}$

  Usiamo la statistica di test

  $$ST = \sqrt{\frac{n}{p(1-p)}}(\bar{X} - p) \text{ ; per il TLC } \overset{\text{approx}}{\sim} \mathcal{N}(0,1)$$

  Otteniamo che

  $$C = \{\sqrt{\frac{n}{p_0(1-p_0)}}|\bar{X} - p_0| > q_{1-\frac{\alpha}{2}}\} = \{\frac{|n\bar{X} - np_0|}{\sqrt{np_0(1-p_0)}} > q_{1-\frac{\alpha}{2}}\}$$

  è una regione critica di livello approx. α

  Oss: qui non abbiamo bisogno di considerare $\bar{X}(1-\bar{X})$ al posto di $p(1-p)$, poiché conosciamo, sotto H_0, $p = p_0$

Analogamente agli intervalli di fiducia, e con le stesse statistiche pivotali, si possono formulare i test bilateri per media di popol. gaussiana con varianza non nota (test t), media nel caso grandi campioni (test Z approssimato), varianza di popol. gaussiana (test χ^2)

Metodo del p-value (soglia di accettazione:

Spesso si ha una famiglia di regioni critiche $(C_\alpha)_{\alpha\in(0,1)}$ con

- C_α regione critica a livello α, $\forall\alpha$
- se $\alpha \leq \alpha'$, allora $C_\alpha \subseteq C_{\alpha'}$ (ragionevole: minore è $\alpha \geq \sup_{\theta\in\Theta_0} P_\theta(C)$, più piccola è C_α)
- $\bigcap_\alpha C_\alpha = \emptyset$, $\bigcup_\alpha C_\alpha = \Omega$

(si può prendere $\Omega = \{(x_1..x_n) \mid x_i \in S\} = S^n$)

Quindi, $\forall \omega \in \Omega$ (esito del campione), $\exists\, \bar{\alpha} = \bar{\alpha}(\omega) \in (0,1)$

- se $\alpha < \bar{\alpha}$, allora $\omega \notin C_\alpha$ e quindi accettiamo H_0 se l'esito del campione è ω
- se $\alpha > \bar{\alpha}$, allora $\omega \in C_\alpha$ e quindi rifiutiamo H_0 " " " " " ω

$\bar{\alpha} = \bar{\alpha}(\omega)$ è detto p-value di ω (ed è funzione dell'esito del campione, solitamente $\bar{\alpha} = \bar{\alpha}(X_1..X_n)$) e fornisce il "grado di plausibilità" di H_0 se si verifica ω (più $\bar{\alpha}$ è basso, minore è il livello

α per cui rifiuto, quindi meno plausibile è $\mathcal{H}_0$).

Informalmente, $\bar{\alpha}(\omega)$ è "la prob., sotto $\mathcal{H}_0$, di avere dati più estremi (rispetto a $\mathcal{H}_0$) di ω"

Esempi:

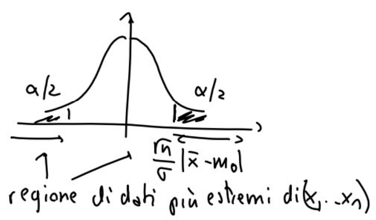
- media di popol. normale $\mathcal{N}(m,\sigma^2)$, varianza nota, $\mathcal{H}_0 : m = m_0$, $\mathcal{H}_1 : m \neq m_0$

  $C_\alpha = \left\{ \frac{\sqrt{n}}{\sigma} |\bar{X} - m_0| > q_{1-\frac{\alpha}{2}} \right\}$

  $\bar{\alpha}(x_1 \dots x_n)$ è t.c. $\frac{\sqrt{n}}{\sigma} |\bar{x} - m_0| = q_{1-\frac{\bar{\alpha}}{2}}$, cioè $\frac{\bar{\alpha}}{2} = P\left\{ Z > \frac{\sqrt{n}}{\sigma} |\bar{x} - m_0| \right\}$

  con $Z \sim \mathcal{N}(0,1)$

  $= P\left\{ |Z| > \frac{\sqrt{n}}{\sigma} |\bar{x} - m_0| \right\}$



- media di popol. $B(p)$, n grande, $\mathcal{H}_0 : p = p_0$, $\mathcal{H}_1 : p \neq p_0$

  $C_\alpha = \left\{ \sqrt{\frac{n}{p_0(1-p_0)}} |\bar{X} - p_0| > q_{1-\frac{\alpha}{2}} \right\}$ (approx.)

  $\bar{\alpha}(x_1 \dots x_n)$ è t.c. $\frac{\bar{\alpha}}{2} = P\left\{ Z > \sqrt{\frac{n}{p_0(1-p_0)}} |\bar{x} - p_0| \right\}$ (approx.)

  con $Z \sim \mathcal{N}(0,1)$

Esempi:

- Un certo modello teorico afferma che la carica elettrica di un dato corpo (in Coulomb) è 5. Dai risultati di 16 misurazioni della carica sul corpo, risulta una media campionaria di 5.2. Supponiamo che le misurazioni siano gaussiane con media il valore reale della carica e dev. standard pari a 0.1. C'è evidenza che il modello teorico proposto sia errato? Effettuare un test di livello 0.01 e calcolare il p-value per i dati forniti.

  $(S, \mathcal{S}, (Q_\theta)_\Theta) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{N}(m, \sigma^2)_{m \in \mathbb{R}})$, $\sigma = 0.1$ (X = carica rilevata in una misur. $\sim \mathcal{N}(m, \sigma^2)$)

  X_i = i-sima misurazione, $i = 1, \dots n = 16$, campione i.i.d. di $\mathcal{N}(m, \sigma^2)$

  $\mathcal{H}_0 : m = 5 (= m_0)$, $\mathcal{H}_1 : m \neq 5$

  Statistica di Test $Z = \frac{\sqrt{n}}{\sigma}(\bar{X} - m) \sim \mathcal{N}(0,1)$

  Regione critica di livello $\alpha = 0.01$ ($q_{1-\frac{\alpha}{2}} = q_{0.995} \approx 2.576$)

  $C = \left\{ \frac{\sqrt{n}}{\sigma} |\bar{X} - m_0| > q_{1-\frac{\alpha}{2}} \right\} = \left\{ 8 \cdot |\bar{X} - 5| > 2.576 \right\}$

  Per $\bar{x} = 5.2$, $8 \cdot |\bar{x} - 5| = 1.6 \leq 2.576$ $\therefore (x_1 \dots x_n) \notin C$, accettiamo H_0: non c'è evidenza statistica dei dati, a livello 0.01, contro il lavoro teorico

  p-value di $\bar{x} = 5.2$: $\bar{\alpha}(\bar{x} = 5.2) = 2P\{Z > 8|\bar{x} - 5|\} = 2(1 - \Phi(1.6)) \approx 2 \cdot (1 - 0.945) = 0.11$

  accettiamo $\mathcal{H}_0$ per $\alpha < 0.11$, rifiutiamo $\mathcal{H}_0$ per $\alpha > 0.11$

- Il banco dichiara che una certa moneta è equilibrata. Lanciamo 1000 volte la moneta, ottenendo 540 teste. C'è evidenza statistica contro l'affermazione del banco? Effettuare un test di livello 0.05 e calcolare il p-value sulla base dei dati.

$(S, \mathcal{J}, (Q_\theta)_\Theta) = (\mathbb{N}, \mathcal{B}(\mathbb{N}), B(p)_{p\in[0,1]})$ p = prob di testa

$X_i = \begin{cases} 1 & \text{se testa al lancio i-simo} \\ 0 & \text{altrimenti} \end{cases}$ $i = 1, \dots$ $n = 1000$ (grande), campione i.i.d. di $B(p)$

$H_0: p = \frac{1}{2} (= p_0)$ (affermazione del banco) $H_1: p \neq \frac{1}{2}$

Statistica di test: $ST = \sqrt{\frac{n}{p(1-p)}}(\bar{X} - p) \overset{\text{approx}}{\sim} N(0,1)$

Regione critica di livello approx. $\alpha = 0.05$ $(q_{0.975} \approx 1.96)$

$C = \left\{ \sqrt{\frac{n}{p_0(1-p_0)}} |\bar{X} - p_0| > q_{1-\frac{\alpha}{2}} \right\} = \left\{ 63.25 \cdot |\bar{X} - 0.5| > 1.96 \right\}$

Per $\bar{x} = \frac{540}{1000} = 0.54$, $63.25 \cdot |\bar{x} - 0.5| = 2.53 > 1.96$: $(x_1, \dots x_n) \in C$, rifiutiamo H_0:

dai dati c'è evidenza, a livello $\alpha = 0.05$, contro $p = \frac{1}{2}$

p-value (approx.) di $\bar{x} = 0.54$: $\bar{\alpha}(\bar{x} = 0.54) = 2P\left\{ Z > \sqrt{\frac{n}{p_0(1-p_0)}} |\bar{x} - p_0| \right\}$

$= 2P\{ Z > 2.53 \} = 2(1 - \Phi(2.53)) \approx 2(1 - 0.9943) = 0.0114$

martedì 16 maggio 2023 22:29

Test statistici 2

Caso ipotesi composta, test unilateri: $\Theta_0 = (-\infty, \theta_0]$ ($H_0: \theta \le \theta_0$), $\Theta_1 = (\theta_0, +\infty)$ ($H_1: \theta > \theta_0$)
(analogamente per $\Theta_0 = [\theta_0, +\infty)$, $\Theta_1 = (-\infty, \theta_0)$)

Esempio: per una moneta con prob. di testa p, ($\theta = p \in \Theta = [0,1]$, $Q_\theta = B(p)$), $H_0: p \le \frac{1}{2}$, $H_1: p > \frac{1}{2}$

È ragionevole prendere una regione critica della forma $\{\bar{X} > d\}$

Dato α livello del test, per determinare d dobbiamo imporre $\sup_{\theta \le \theta_0} P_\theta\{\bar{X} > d\} \le \alpha$,
intuitivamente ci aspettiamo che il sup è realizzato per $\theta = \theta_0$

Domande: a) è vero che il sup è realizzato per $\theta = \theta_0$?
b) la regione $\{\bar{X} > d\}$ è effettivamente la scelta "migliore"? precisamente, fissato α (e quindi d), essa massimizza la potenza del test?

D'ora in poi, consideriamo regioni critiche della forma $C = \{(X_1 .. X_n) \in \tilde{C}\}$ (con $\tilde{C} \in \mathcal{F}^{\otimes n}$)

Def: Dati $(S, \mathcal{J}, (Q_\theta)_{\theta \in \Theta})$ modello statistico con $\Theta \subseteq \mathbb{R}$ intervallo, $(X_1 .. X_n)$ campione i.i.d. di legge $(Q_\theta)_{\theta \in \Theta}$, definito su $(\Omega, \mathcal{F}, (P_\theta)_{\theta \in \Theta})$, $T: \Omega \to \mathbb{R}$ v.a.,
diciamo che il modello è a rapporto di verosimiglianza crescente rispetto a T: se $\forall \theta_1 < \theta_2$, esiste una funzione $f_{\theta_1, \theta_2}: \mathbb{R} \to (0, +\infty)$ strettamente crescente t.c.

$$\frac{L(\theta_2, X_1 .. X_n)}{L(\theta_1, X_1 .. X_n)} = f_{\theta_1, \theta_2}(T)$$

(cioè $L(\theta_2, X_1 .. X_n)/L(\theta_1, X_1 .. X_n)$ è funz. strett. cresc. di T)
(Si sottointende che, se $L(\theta_1, x_1 .. x_n) = 0$, allora $L(\theta_2, x_1 .. x_n) = 0$

Esempi:

- Modello gaussiano $(S, \mathcal{J}, (Q_\theta)_\Theta) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), N(m, \sigma^2)_m)$

$$L_m(x_1 .. x_n) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sigma^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2\right)$$

$$e^{-\frac{1}{2\sigma^2}(m_2 - m_1)(2n\bar{x} + (m_2 - m_1))}$$

$$\frac{L_{m_2}(x_1 .. x_n)}{L_{m_1}(x_1 .. x_n)} = \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n \left((x_i - m_2)^2 - (x_i - m_1)^2\right)\right) = \exp\left(-\frac{1}{2\sigma^2}(m_1 - m_2) \sum_{i=1}^n (2x_i - m_1 - m_2)\right)$$

$$= \exp\left(-\frac{1}{2\sigma^2}(m_2^2 - m_1^2)\right) \exp\left(\frac{n(m_2 - m_1)}{\sigma^2} \bar{x}\right)$$

per $m_2 > m_1$, il rapporto è crescente in $\bar{x}$

- Modello Bernoulli $(S, \mathcal{J}, (Q_\theta)_\Theta) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), B(p)_p)$

$$L_p(x_1 .. x_n) = p^{n\bar{x}} (1-p)^{n(1-\bar{x})}$$

$$\frac{L_{p_2}(x_1 .. x_n)}{L_{p_1}(x_1 .. x_n)} = \left(\frac{p_2}{p_1}\right)^{n\bar{x}} \left(\frac{1-p_2}{1-p_1}\right)^{n(1-\bar{x})}$$

per $p_2 > p_1$, il rapporto è crescente in $\bar{x}$

Teor: Siano $(S, \mathcal{S}, (Q_\theta)_{\theta\in\Theta})$ modello statistico con $\Theta \subseteq \mathbb{R}$ intervallo, $(X_1, \dots X_n)$ campione i.i.d. di legge Q_θ, supponiamo che il modello sia a rapporto di ver. cresc. rispetto a T.

Consideriamo il test $\mathcal{H}_0: \theta \leq \theta_0$ contro $\mathcal{H}_1: \theta > \theta_0$, sia

$$C = \{T \geq d\}$$

Allora:

(i) $\sup_{\theta\leq\theta_0} P_\theta(C) = P_{\theta_0}(C)$

(ii) il test di regione critica C è il più potente tra i test di livello $P_{\theta_0}(C)$

Lemma (Neyman-Pearson): Siano $(S, \mathcal{S}, (Q_\theta)_{\theta\in\Theta})$ modello statistico con $\Theta = \{\theta_0, \theta_1\}$, consideriamo il test $\mathcal{H}_0: \theta = \theta_0$, $\mathcal{H}_1: \theta = \theta_1$, per $c > 0$ sia

$$C = \{L_{\theta_0}(X_1, \dots X_n) \leq c L_{\theta_1}(X_1, \dots X_n)\}$$

Allora:

a) il test di regione critica C è il più potente tra i test di livello $P_{\theta_0}(C)$

b) $P_{\theta_0}(C) \leq P_{\theta_1}(C)$

Dim (del lemma di N.-P.): Per semplicità, consideriamo il caso assol. cont.

a) Notiamo che, se il test ha regione critica D, la sua potenza è $P_{\theta_1}(D)$, vogliamo quindi dim. $P_{\theta_1}(D) \leq P_{\theta_1}(C)$.

Scriviamo $C = \{(X_1, \dots X_n) \in \tilde{C} = \{L_{\theta_0} \leq c L_{\theta_1}\}\}$, $D = \{(X_1, \dots X_n) \in \underbrace{\tilde{D}}_{\subseteq \mathbb{R}^n}\}$

Notiamo $P_\theta(D) = P_\theta\{(X_1, \dots X_n) \in \tilde{D}\} = \int_{\tilde{D}} L_\theta(x_1, \dots x_n)\, dx_1 \dots dx_n$

Per def di $\tilde{C} = \{L_{\theta_0} \leq c L_{\theta_1}\}$

$$(\mathbb{1}_{\tilde{C}} - \mathbb{1}_{\tilde{D}})(L_{\theta_0} - cL_{\theta_1}) \leq \overbrace{\mathbb{1}_{\tilde{C}}(L_{\theta_0} - cL_{\theta_1})}^{\leq 0} - \overbrace{\mathbb{1}_{\tilde{D}\setminus\tilde{C}}(L_{\theta_0} - cL_{\theta_1})}^{\geq 0} \leq 0$$

Integriamo su $\mathbb{R}^n$ (risp. alla misura di Lebesgue):

$$\underbrace{\int (\mathbb{1}_{\tilde{C}} - \mathbb{1}_{\tilde{D}}) L_{\theta_0}(x_1 \dots x_n)\, dx_1 \dots dx_n}_{= P_{\theta_0}(C) - P_{\theta_0}(D)} \leq \underbrace{c\int (\mathbb{1}_{C} - \mathbb{1}_{D}) L_{\theta_1}(x_1, \dots x_n)\, dx_1 \dots dx_n}_{= c(P_{\theta_1}(C) - P_{\theta_1}(D))}$$

Poiché D ha livello $P_{\theta_0}(C)$, $P_{\theta_0}(D) \leq P_{\theta_0}(C)$, quindi $P_{\theta_1}(C) - P_{\theta_1}(D) \geq 0$

b) $(\mathbb{1}_{\tilde{C}} - P_{\theta_0}(C))(L_{\theta_0} - cL_{\theta_1}) = \underbrace{(1 - P_{\theta_0}(C))}_{\geq 0} \mathbb{1}_{\tilde{C}} \underbrace{(L_{\theta_0} - cL_{\theta_1})}_{\leq 0} - P_{\theta_0}(C) \underbrace{\mathbb{1}_{\tilde{C}^c}(L_{\theta_0} - cL_{\theta_1})}_{\geq 0} \leq 0$

Integrando come prima,

$$c(P_{\theta_1}(C) - P_{\theta_0}(C)) \geq 0$$

Dim del teor:

i) Basta dim. che $\theta_1 \leq \theta_2 \leq \theta_0 \Rightarrow P_{\theta_1}(C) \leq P_{\theta_2}(C)$

$$C = \{T \geq d\} \underset{\text{rapporto a ver cresc}}{=} \{\underbrace{f_{\theta_1,\theta_2}(T)}_{= \frac{L_{\theta_2}(X_1 \dots X_n)}{L_{\theta_1}(X_1 \dots X_n)}} \geq \underbrace{f_{\theta_1,\theta_2}(d)}_{=: 1/c}\} = \{L_{\theta_1}(X_1 \dots X_n) \leq c L_{\theta_2}(X_1 \dots X_n)\}$$

quindi per Lemma di N-P. (b) (con $\Theta=\{\theta_1,\theta_2\}$), $P_{\theta_1}(C)\le P_{\theta_2}(C)$

ii) Dobbiamo dim. che, se D è un'altra regione critica di livello $P_{\theta_0}(C)$, $\forall \theta>\theta_0$, $P_\theta(D)\le P_\theta(C)$

Poiché $P_{\theta_0}(D)\le P_{\theta_0}(C)$, per lemma di N-P. (a) (con $\Theta=\{\theta_0,\theta\}$), $P_\theta(D)\le P_\theta(C)$.

Test unilatero per la media di una popol. gaussiana, varianza nota

· $(S,\mathcal{I},(Q_\theta))=(\mathbb{R},\mathcal{B}(\mathbb{R}),\mathcal{N}(m,\sigma^2)_{m\in\mathbb{R}})$ $\quad\theta=m\in\mathbb{R}$, σ^2 nota

$X_1,\dots X_n$ campione i.i.d.

Consideriamo il test $H_0: m\le m_0$ contro $H_1: m>m_0$, a livello α

Regione critica C: $C=\{\bar X>d\}$, ottimale per il teor.

Usiamo la statistica di test

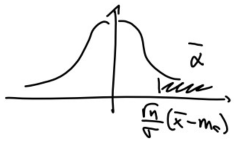
$$ST=\frac{\sqrt n}{\sigma}(\bar X-m)\sim\mathcal{N}(0,1) \text{ sotto } P_m$$

Imponiamo $\sup_{m\le m_0} P_m\{\bar X>d\}\underset{\text{teor}}{=}P_{m_0}\{\bar X>d\}=\alpha$

$$\alpha=P_{m_0}\{\bar X>d\}=P_{m_0}\left\{\frac{\sqrt n}{\sigma}(\bar X-m_0)>\frac{\sqrt n}{\sigma}d\right\}=P\left\{Z>\frac{\sqrt n}{\sigma}d\right\} \text{ con } Z\sim\mathcal{N}(0,1)$$

$$\Leftrightarrow \frac{\sqrt n}{\sigma}d=q_{1-\alpha}$$

Quindi $C=\left\{\frac{\sqrt n}{\sigma}(\bar X-m_0)>q_{1-\alpha}\right\}$



Il p-value di $(x_1,\dots x_n)$ è $\bar\alpha=P\left\{Z>\frac{\sqrt n}{\sigma}(\bar x-m_0)\right\}$ con $Z\sim\mathcal{N}(0,1)$

Oss: La potenza del test in $m>m_0$ è

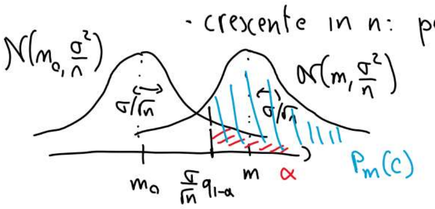
$$P_m(C)=P_m\left\{\frac{\sqrt n}{\sigma}(\bar X-m)>q_{1-\alpha}-\frac{\sqrt n}{\sigma}(m-m_0)\right\}=P\{\underbrace{Z}_{\sim\mathcal{N}(0,1)}>q_{1-\alpha}-\frac{\sqrt n}{\sigma}(m-m_0)\}$$

$$=1-\Phi\left(q_{1-\alpha}-\frac{\sqrt n}{\sigma}(m-m_0)\right)$$

$P_m(C)$ è: · crescente in m (tenendo fissi gli altri parametri)

· crescente in α: se si vuole abbassare α, purtroppo anche la potenza si abbassa

· crescente in n: per abbassare il livello α e la potenza, è possibile aumentare n



Test unilatero per una proporzione (media di popol. Bernoulli), grandi campioni

· $(S,\mathcal{I},(Q_\theta)_\Theta)=(\mathbb{R},\mathcal{B}(\mathbb{R}),\mathcal{B}(p)_{p\in[0,1]})$, $\theta=p\in[0,1]$

$X_1,\dots X_n$ campione i.i.d., assumiamo n grande

Consideriamo il test $H_0: p\le p_0$ contro $H_1: p>p_0$, a livello approssimativamente α

Regione critica C: $C=\{\bar X>d\}$ ottimale per il teor.

Usiamo la statistica di test

$$ST=\sqrt{\frac{n}{p(1-p)}}(\bar X-p), \text{ per il TLC } \overset{\text{approx}}{\sim}\mathcal{N}(0,1)$$

Otteniamo come sopra

$$C = \left\{ \sqrt{\frac{n}{p_0(1-p_0)}} (\bar{X} - p_0) > q_{1-\alpha} \right\}$$ (livello approssimativamente α)

Il p-value di $(x_1..x_n)$ è (approssimativamente) $\bar{\alpha}(x_1..x_n) = P\left\{ Z > \sqrt{\frac{n}{p_0(1-p_0)}} (\bar{x} - p_0) \right\}$ con $Z \sim \mathcal{N}(0,1)$

Analogamente si possono formulare i test unilateri per media di popol. gaussiana con varianza non nota (t test), media nel caso grandi campioni (z test approssimato), varianza di popol. gaussiana (χ^2 test)

Esempio:

- Un'azienda dichiara che la percentuale di pezzi difettosi, sul totale dei pezzi prodotti, non supera il 10%. Su 1000 prodotti testati, 120 risultano difettosi. C'è evidenza contro l'affermazione dell'azienda? Effettuare un test a livello $\alpha = 0.05$

  $(S, \mathcal{F}, (P_\theta)_\theta) = (\mathbb{N}, \mathcal{B}(\mathbb{N}), \mathcal{B}(p)_{p \in [0,1]})$. p = prob pezzo difettoso

  $X_i = \begin{cases} 1 & \text{se } i\text{-simo pezzo difettoso} \\ 0 & \text{altrimenti} \end{cases}$, $i = 1, .. n = 1000$ grande   campione i.i.d. di $\mathcal{B}(p)$

  $\mathcal{H}_0 : p \leq 0.1 (= p_0)$ $\mathcal{H}_1 : p > 0.1$

  Statistica di test $ST = \sqrt{\frac{n}{p(1-p)}} (\bar{X} - p) \overset{\text{approx}}{\sim} \mathcal{N}(0,1)$

  Regione critica di livello $\alpha = 0.05$ (approx.)

  $$C = \left\{ \sqrt{\frac{n}{p_0(1-p_0)}} (\bar{x} - p_0) > q_{1-\alpha} \right\} = \left\{ 63.25 \cdot (\bar{x} - 0.1) > 1.96 \right\}$$

  Per $\bar{x} = \frac{120}{1000} = 0.12$, $63.25 \cdot (\bar{x} - 0.1) = 1.265 \leq 1.96$: $(x_1..x_n) \notin C$, non c'è evidenza contro dichiarazione azienda

  p-value di $\bar{x}$ (approx.) : $\bar{\alpha}(\bar{x} = 0.12) = P\left\{ Z > \sqrt{\frac{n}{p_0(1-p_0)}} (\bar{x} - p_0) \right\}$

  $$= P\{ Z > 1.265 \} = 1 - \phi(1.265) \approx 1 - 0.898 = 0.102$$

  si accetta per $\alpha < \bar{\alpha}$, si rifiuta per $\alpha > \bar{\alpha}$