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

$\Big[$ 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: $\Omega = \mathbb{N}$: $\{n^\circ$ pari di figli$\} = \bigcup_{n\in\mathbb{N}} \{2n\}$
$P[n^\circ$ pari di figli$] = \sum_{n\in\mathbb{N}} P\{2n\}$

Esempio fondamentale: modello uniforme (o equiprobabile):

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

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

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

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

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

Oss: P uniforme verifica la def. assiomatica di probabilità

Esempi:

- lancio di una moneta equilibrata
($\Omega = \{T, C\}$ $\mathcal{F} = \mathcal{P}(\Omega)$ P uniforme $P\{T\} = P\{C\} = \frac{1}{2}$)
- lancio di un dado equilibrato
($\Omega = \{1, 2 \dots 6\}$ $\mathcal{F} = \mathcal{P}(\Omega)$ 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) $\forall A_1,\ldots A_n\in\mathcal{F}$,

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

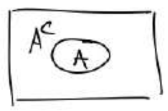
(formula di inclusione - esclusione)

Dimostrazione:

a) Già visto

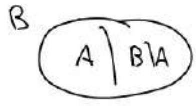
b) $\Omega = A \dot\cup A^c$ ($\dot\cup$ = unione disgiunta), quindi

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



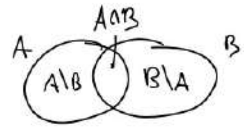
c) se $A \subseteq B$, allora $B = A \dot\cup (B \setminus A)$, quindi

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



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

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



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

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 $P(A) \leq 1 \quad \forall A \in \mathcal{F}$ e quindi P è misura di probabilità

(segue da $P(\Omega) = 1$ e monotonia)

• Se Ω è finito e P soddisfa (i), (ii), (iii') con

(iii'): $\forall A, B \in \mathcal{F}$ disgiunti, $P(A \cup B) = P(A) + P(B)$ (additività)

allora P soddisfa (iii) e quindi è misura di probabilità

(poiché $\mathcal{F}$ è finito)

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

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

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

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

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

Dim:

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

Dim:

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

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

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

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

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

Altre definizioni della probabilità:

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

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

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

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

→ più adatta per fenomeni come esperimenti fisici o biologici

* richiede di sapere a priori che il limite esiste, e non è chiara la def. rigorosa di limite
* operativamente, non sempre è possibile effettuare tante prove di un esperimento

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

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

Sequenze di esperimenti

sabato 14 marzo 2020 13:38

Modellizzare sequenza ordinate (finita o infinita) di esperimenti

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

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

  dove $\Omega_1 = \{\text{esiti del } 1° \text{ esperimento}\}$
  $\Omega_{2,\omega_1} = \{\text{esiti del } 2° \text{ "}\}$ (possibilmente dipendenti da ω_1)
  $\Omega_{3,\omega_1,\omega_2} = \{\text{esiti del } 3° \text{ "}\}$ (" " " 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)

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

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

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

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

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

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

(• 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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)?

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

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

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

Significato dell'indipendenza: A e B indipendenti $\Leftrightarrow$ 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 " "

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

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

Dim: per esercizio.

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

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

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

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

Esempio 2:

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

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

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

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

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

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)

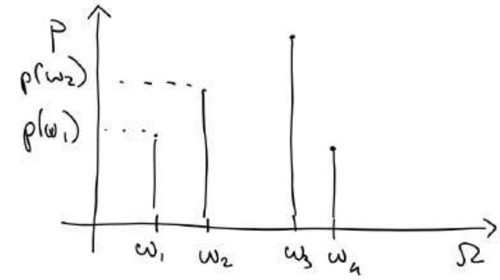
3) σ-additività: $A=\bigcup_{n\in\mathbb{N}}A_n$, A_n a due a due disgiunti

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

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

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

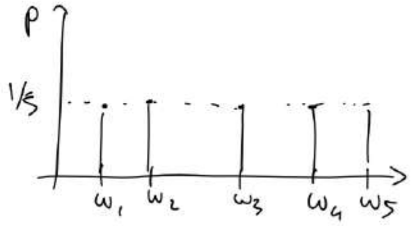
Rappresentazione della densità discreta tramite grafico a barre



(ω_i non necessariamente ordinati)

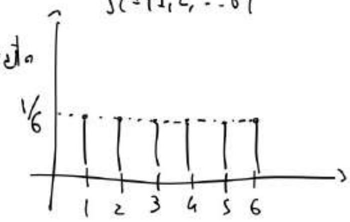
Il grafico a barre indica dove P assegna più "massa"

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



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

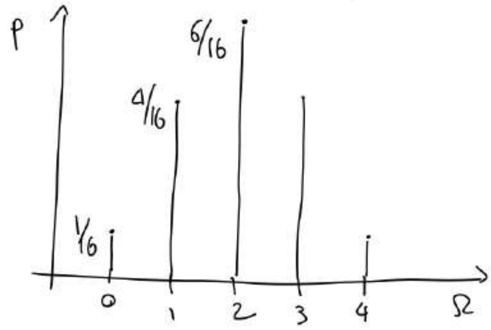
esito lancio dado equilibrato



Esempio: n° teste in 4 lanci di moneta:

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

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



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

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

Def: 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.

$P(\Omega_0) = 1$ $\quad$ (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,

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

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 $(*)$, $n \in \mathbb{N}^+ \cup \{+\infty\}$)

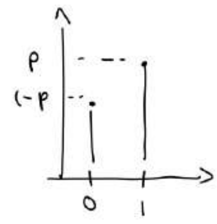
Esempi:

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

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

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

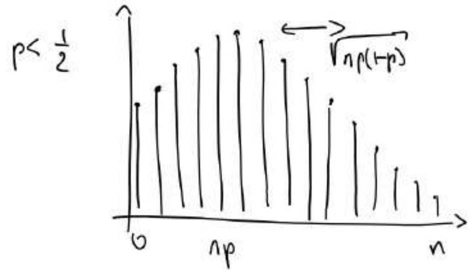


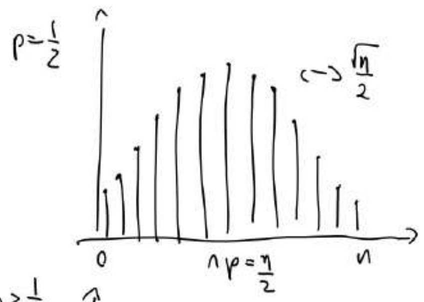
$B(p)$ rappresenta un esperimento con esito successo (1) o insuccesso (0), in cui il successo ha probabilità p (prova di Bernoulli) $(P\{1\} = \overline{P}_n\{\omega \mid \omega_i = 1\}\ \forall i)$

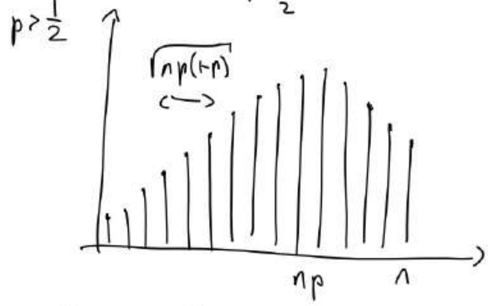
· Distribuzione binomiale di parametri $n \in \mathbb{N}^+$ e $p \in [0,1]$ $(B(n,p)$ o $Bin(n,p))$

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

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







$B(n,p)$ 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:

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

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

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

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

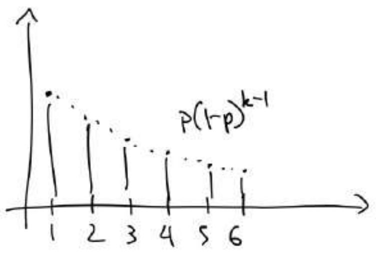
Quindi

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

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

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

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



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

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

con $T: \bar{\Omega}_\infty \to \mathbb{N}^+ \cup \{+\infty\}$, $T(\omega_1, \omega_2, \dots)$ = 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:

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

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

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

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

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

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

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

con $T_h: \bar{\Omega}_\infty \to \Omega \cup \{+\infty\}$, $T_h(\omega_1, \omega_2, \dots)$ = 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: $\tilde\Omega = \{S \subseteq \{1,..N\} \mid \#S = n\}$, $\tilde{\mathcal{F}} = \mathcal{P}(\tilde\Omega)$, $\tilde P$ 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 $P(A) = \dfrac{\#A}{\#\Omega} = p(k)$.

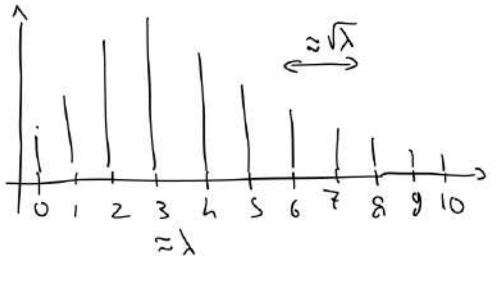
Più in generale, se gli N oggetti di un'urna sono divisi in gruppi $(a_1), ..(a_m)$, con $N_1, ..N_m$ elementi rispettivamente $(N_1 + .. + N_m = N)$, ed estraiamo n oggetti, allora

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

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

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

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



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

Prop: Sia $(p_n)_{n\in\mathbb{N}}$ successione in $(0,1)$ con $np_n \to \lambda > 0$ per $n \to \infty$
Sia $p_n(k) = \binom{n}{k} p_n^k (1-p_n)^{n-k}$ la densità discreta $Bin(n,p)$ (estesa a $k \in \mathbb{N}$ ponendo $p_n(k) = 0 \ \forall k > n$).
Allora, $\forall k \in \mathbb{N}$, $p_n(k) \to p(k) = \frac{\lambda^k}{k!} e^{-\lambda}$ densità Poisson (λ).

Dim: esercizio

Esempi di applicazione:

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

6

lunedì 2 gennaio 2023 17:41

Variabili aleatorie discrete

Def: Dato Ω spazio discreto (cioè finito o numerabile), con $\mathcal{F} = \mathcal{P}(\Omega)$, S insieme $\neq \emptyset$, una variabile aleatoria (v.a.) da Ω a S è una funzione

$X: \Omega \to S$

Tipicamente $S = \mathbb{R}$: v.a. reale

$S = \mathbb{R}^n$: vettore aleatorio

Esempio:

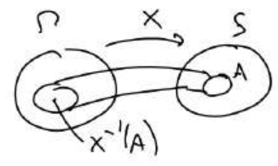
Data una sequenza di n lanci di moneta, $\Omega = \{T, C\}^n$, sono esempi di v.a. reali

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

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

Oss: Significato e utilità delle v.a:

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



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

Fatto: X^{-1} commuta rispetto alle operazioni insiemistiche:

dati A, A_i, $i \in I$, sottoinsiemi di S,

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

$X^{-1}(A^c) = X^{-1}(A)^c$ (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)

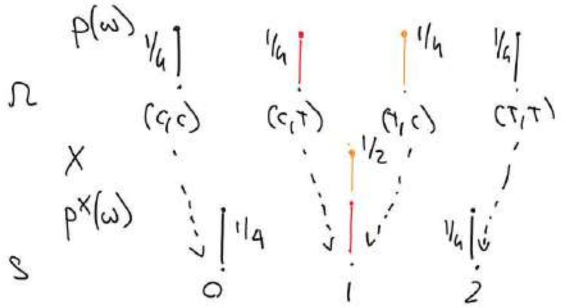
P^X (o $X_\# P$ o $P \circ X^{-1}$ o $X(P)$) : misura di probabilità su $(S_X, \mathcal{P}(S_X))$ definita da

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

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

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

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

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. $1/6$)

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

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

$P\{$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

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

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

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

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

Sì: costruzione canonica:

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

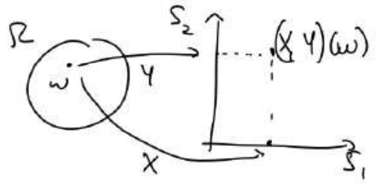
$\forall A \in \mathcal{P}(S)$, $P\{X \in A\} = P(A) = Q(A)$

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

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

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

- Se $X:\Omega\to S_1$, $Y:\Omega\to S_2$ sono v.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.

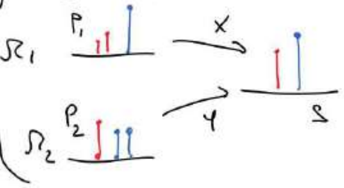


- Se $X:\Omega\to\mathbb{R}$, $Y:\Omega\to\mathbb{R}$ sono v.d. reali, allora
  $X+Y$, XY, ... sono v.d.
  Infatti $X+Y=f(X,Y)$ con $f:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$, $(x,y)\mapsto x+y$
  $XY=f(X,Y)$ $\quad f:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$, $(x,y)\mapsto xy$

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 $R_{p_1^X}=\{x\in S\mid p_1^X(x)>0\}$

Oss: Data $\mathcal{G}$ σ-algebra su S, con $\{x\}\in\mathcal{G}$ $\forall x\in S$, vale
$P_1^X=P_2^Y$ su $(S,\mathcal{P}(S)) \iff P_1^X=P_2^Y$ su $(S,\mathcal{G}) \iff p_1^X=p_2^Y$

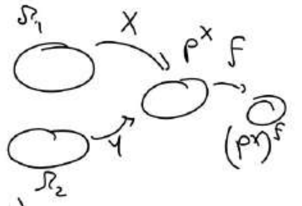
Oss: Se $X=Y$ q.c., allora $X\overset{(d)}{=}Y$, ma il viceversa non vale

Dim: · Se $X=Y$ q.c., allora, $\forall A\in\mathcal{P}(S)$,
$$\begin{aligned}P\{X\in A\}&=P\{X\in A, X=Y\}+\underbrace{P\{X\in A, X\neq Y\}}_{\le P\{X\neq Y\}=0}\\&=P\{Y\in A, X=Y\}+\overbrace{P\{Y\in A, X\neq Y\}}\\&=P\{Y\in A\}\end{aligned}$$

Esempio di X,Y v.d. non uguali q.c. ma con $X\overset{(d)}{=}Y$
$X\sim B(\frac12)$, $Y=1-X$, allora
- $Y\sim B(\frac12)$: Y ha valori in $0,1$ q.c. (cioè $P\{Y\in\{0,1\}\}=1$) e $P\{Y=1\}=P\{X=0\}=\frac12$
- $P\{X\neq Y\}=1$: $P\{X=Y\}=P\{2X-1=0\}=P\{X=\frac12\}=0$

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.

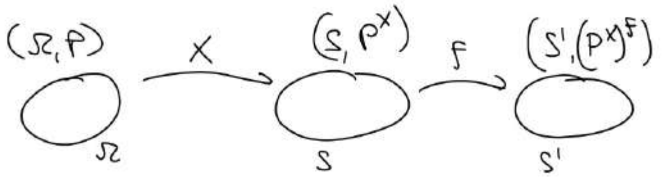


- Se $X: \Omega_1 \to S$ e $Y: \Omega_2 \to S$ hanno la stessa legge e $f: S \to S'$, allora $f(X)$ e $f(Y)$ hanno la stessa legge:

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, $i = 1 \dots n$,

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

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

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)

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

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

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

Oss: La legge congiuntà 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, $i=1,2$

$X_i \sim B(\frac{1}{2})$ $i=1,2$, $\quad Y_1 \sim B(\frac{1}{2})$, $Y_2 = Y_1$ 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.d. 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:

$$\begin{aligned} P\{X_1 \in A_1, \dots X_n \in A_n\} &= \sum_{(x_1, \dots x_n) \in A_1 \times \dots \times A_n} p_X(x_1, \dots x_n) = \sum_{x_1 \in A_1, \dots x_n \in A_n} p_X(x_1, \dots x_n) \\ &= \sum_{x_1 \in A_1, \dots x_n \in A_n} p_{X_1}(x_1) \cdot \dots \cdot p_{X_n}(x_n) \\ &= \left(\sum_{x_1 \in A_1} p_{X_1}(x_1)\right) \cdot \dots \cdot \left(\sum_{x_n \in A_n} p_{X_n}(x_n)\right) = P\{X_1 \in A_1\} \cdot \dots \cdot P\{X_n \in A_n\} \end{aligned}$$

Esempio:

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

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

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

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

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

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

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

dato $\{Y_1 = y\}$, 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 <u>non</u> 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 $X_{J_1} = (X_i)_{i\in J_1}: \Omega \to \times_{i\in J_1} S_i$, ... $X_{J_m} = (X_i)_{i \in J_m}: \Omega \to \times_{i\in J_m} S_i$ 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)

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

(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

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

Dim: a) esercizio

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

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

$P = P_1 \otimes ... \otimes P_n$

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

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

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

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

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

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

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

Significato del valore atteso:

- baricentro della distribuzione di X:

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

  media dei valori x_i assunti da X, pesata con la densità discreta p_X di X

- nell'esempio di estrazione da una popolazione Ω (P uniforme su Ω) e misurazione di una caratteristica quantitativa X,

  $E[X] = \frac{1}{\#\Omega} \sum_{\omega \in \Omega} X(\omega)$ è la media aritmetica della caratteristica su tutta la popolazione

- come vedremo (legge dei grandi numeri), dato un esperimento con esito quantitativo X, la media aritmetica dei risultati di n ripetizioni dell'esperimento converge, per $n \to \infty$, al valore atteso (se questo è finito)

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

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

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

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

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

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

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

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

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

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

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

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

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 $e^{\alpha X}\geq 0$

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

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

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

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

Lemma (proprietà del valore atteso):

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

b) X v.a integrabile / ≥ 0 q.c, $a\in\mathbb{R} \Rightarrow aX$ integrabile / ≥ 0 q.c. o ≤ 0 q.c,
$E[aX]=aE[X]$ (con la convenzione $0\cdot\infty=0$)

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

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

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

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

Dim:

a) facile esercizio

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

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

d) da oss precedente

e) Se $X\leq Y$ q.c, allora $X(\omega)p(\omega)\leq Y(\omega)p(\omega)\ \forall\omega\in\Omega$, quindi
$E[X]=\sum_{\omega}X(\omega)p(\omega)\leq\sum_{\omega}Y(\omega)p(\omega)=E[Y]$

f) $E[|X+Y|]=\sum_{\omega}|X(\omega)+Y(\omega)|\,p(\omega)\leq\sum_{\omega}(|X(\omega)|+|Y(\omega)|)p(\omega)=\sum_{\omega}|X(\omega)|p(\omega)+\sum_{\omega}|Y(\omega)|\,p(\omega)$
$<\infty$ se X e Y sono integrabili, in questo caso
$E[X+Y]=\sum_{\omega}(X(\omega)+Y(\omega))\,p(\omega)=\sum_{\omega}X(\omega)p(\omega)+\sum_{\omega}Y(\omega)p(\omega)=E[X]+E[Y]$.

Esempi notevoli:

- $X\sim B(p)$, $X=\mathbb{1}_A$ q.c con $A=\{X=1\}$, $P(A)=p$
$E[X]=E[\mathbb{1}_A]=1\cdot P(A)+0\cdot P(A^c)=P(A)=p$

- $X\sim B(n,p)$
Prendiamo $X_1,\dots X_n\sim B(p)$ 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 $\exists\, E[X] \in [0, +\infty]$

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

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

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

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

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

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

$P(C_i) = \frac{\# C_i}{\# \Omega} = \frac{(N-1)!}{N!} = \frac{1}{N}$ $\qquad$ 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

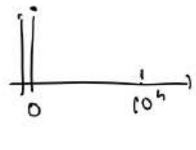
- valore atteso $E[X]$
- mediana: ogni valore $m_X \in \mathbb{R}$ t.c. $P\{X \leq m_X\} \geq \frac{1}{2}$, $P\{X \geq m_X\} \geq \frac{1}{2}$ (esercizio: esiste almeno una mediana)
- moda: ogni valore $mo_X \in \mathbb{R}$ di massimo per p_X (esercizio: esiste almeno una moda)

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

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

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

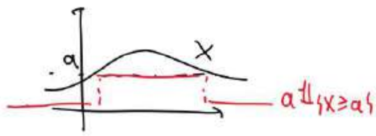
come segue da $|aX+Y|^p \le 2^{p-1}(|a|^p|X|^p + |Y|^p)$.

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 $P\{X > a\} \le e^{-\lambda a} E[e^{\lambda X}]$

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

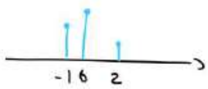
$$Var(X) = E[(X - E[X])^2] \in [0, +\infty)$$

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

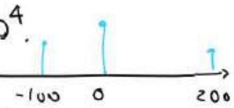


1) supponiamo di vincere 2 se esce 6, -1 se esce 1 o 2, 0 altrimenti

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) $Var(X) \ge 0$, $Var(X) = 0 \iff X$ = costante q.c.

Dim:

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

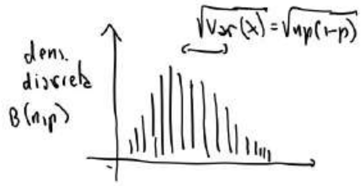
b) $Var(aX + b) = E[(aX + b - aE[X] - b)^2] = a^2 Var(X)$.

c) Poiché $(X - E[X])^2 \ge 0$, $Var(X) \ge 0$ e $= 0 \iff (X - E[X])^2 \ge 0 \iff X = E[X]$ = 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 $E[\varphi(X,Y)] = \sum_{i,j}\varphi(x_i,y_j)\,p_{(X,Y)}(x_i,y_j)$

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

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

Dim:

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

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

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

Più in generale vale

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

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

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

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

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

(ben definito per dis di Schwarz)

Proprietà di Cov:

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

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

Il viceversa non vale:

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

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

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

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

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

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

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

Dim:

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

Esempio:

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

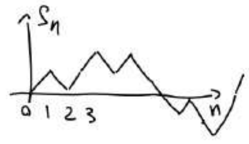
$$Var(X) = \sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

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

Esempio: passeggiata aleatoria simmetrica

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

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



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

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

Coefficiente di correlazione tra X e Y: il numero

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

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

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

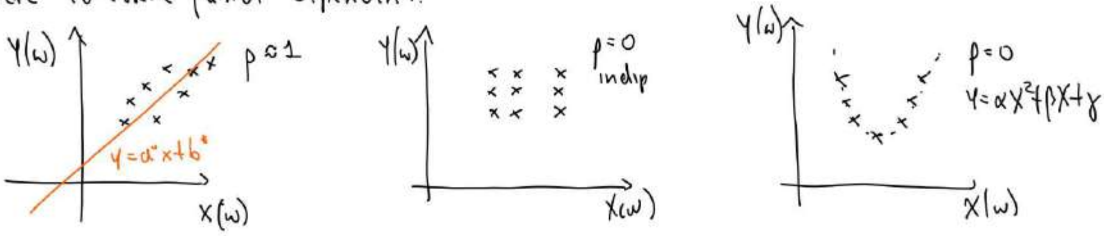
Def: La retta $y = a^* x + b^*$ è detta retta di regressione tra X e Y.

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. $\not\Rightarrow$ indipendenza corrisponde a:

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



Dim, sketch:

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

Oss: Come abbiamo visto,

- se un problema coinvolge una sola v.a., tutte le quantità di interesse ($P\{X\in A\}$, $E[X]$, $Var(X)$ ..) dipendono solo 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 congiunte è garantita dalla costruzione canonica

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

10

domenica 19 marzo 2023 11:49

Teoremi limite (LGN, TCL)

Esempio/motivazione:

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

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

Oss:
- queste affermazioni valgono per un grande numero di esperimenti 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

$(\Omega_{(n)} = \Omega_0^n,\ P_{(n)} = P_0^{\otimes n} = \text{uniforme},\ X_i = X \circ \pi_i)$

$(X_1, \dots X_n)$ 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)

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

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

Per Chebyshev, $\forall \varepsilon > 0$

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

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

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

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

Esempio:

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

Se gioco tante volte, stima della vincita netta?

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

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

X_i = vincita i-sima giocata

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

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

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

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

non conviene giocare!

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

Esempio importante: LGN per frequenza empirica / binomiale

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

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

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

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

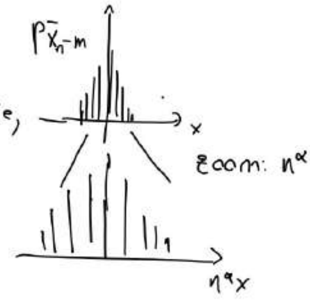
Notiamo che la freq assoluta $Y_n = X_1 + \dots + X_n$ è $\sim B(n,p)$, quindi la LGN dà

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

Ad es, in n lanci di monete 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

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

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, $\forall -\infty \leq a < b \leq +\infty$,

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

equivalentemente, $P\{a \leq \frac{\sqrt{n}}{\sigma}(\bar{X}_n - m) \leq b\} \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ 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 empirica / binomiale (teor. di De Moivre-Laplace)

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

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

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

Infatti $Y_n \overset{(d)}{=} X_1 + \ldots + X_n$ 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 $\left(\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1 \text{ e}\right)$

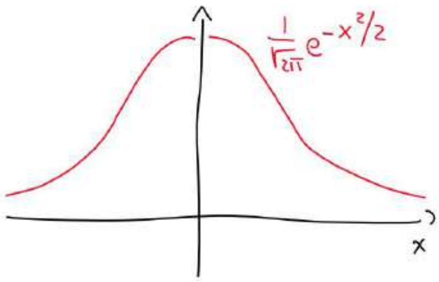
$$\int_{-3}^3 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \simeq 0.997$$

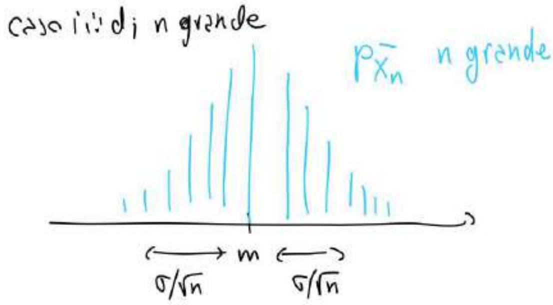
e dunque, per n grande, la prob che $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - m) \in [-3, 3]$, cioè

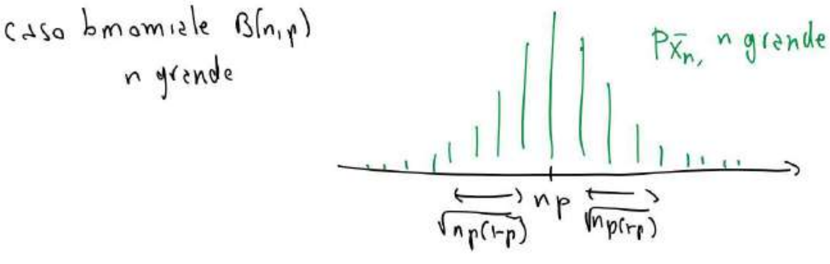
$\bar{X}_n \in \left[m - \frac{3\sigma}{\sqrt{n}}, m + \frac{3\sigma}{\sqrt{n}}\right]$ (o equiv. $n\bar{X}_n \in [mn - 3\sigma\sqrt{n}, mn + 3\sigma\sqrt{n}]$)

è ≈ 0.997.

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







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

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

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

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

dove abbiamo usato

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

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

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

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

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

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

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

$(\Omega, \mathcal{P}(\Omega), P)$ discreto (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
$P_Y(A \mid X=x) = P(Y \in A \mid X=x) \quad \forall A \in \mathcal{P}(S_2)$

$P_Y(\cdot \mid X=x)$ ha densità discreta
$$p_{Y|X}(y|x) = P\{Y=y \mid X=x\} = \frac{P\{X=x, Y=y\}}{P\{X=x\}} = \frac{p_{(X,Y)}(x,y)}{p_X(x)}$$
con p_X, $p_{(X,Y)}$ densità discrete rispettivamente di X e (X,Y)

In particolare, $P_Y(\cdot \mid X=x)$ dipende solo dalla legge congiunta di (X,Y).

(Si estende $p_{Y|X}(y|x)$ a $x \in S_1$ ponendo $p_{Y|X}(y|x) = 0$ $\forall x \in S_1 \setminus R_{p_X}$)

Prop (della legge condizionata alla legge)
$$P_Y(A) = \sum_{x \in R_{p_X}} P_Y(A \mid X=x)\, P_X(x) \quad \forall A \in \mathcal{P}(S_2)$$

Dim: Dalla formula della partizione applicata alla partizione $\{X=x\}_{x \in R_{p_X}}$:
$$P_Y(A) = P\{Y \in A\} = \sum_{x \in R_{p_X}} P\{Y \in A \mid X=x\}\, P\{X=x\}$$

Oss: In realtà $\{X=x\}_{x \in R_{p_X}}$ non è una partizione di Ω, poiché $\{X \notin R_{p_X}\}$ può essere non vuoto, benché di misura P nulla. Però ci possiamo restringere a $\Omega_0 = \{X \in R_{p_X}\}$, che ha $P(\Omega_0) = 1$:
$\hookrightarrow P|_{\mathcal{P}(\Omega_0)}$ 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 $P_M(\cdot|N=n)=B(n,p)$

$$p_M(k)=\sum_{n=0}^{\infty} p_{M|N}(k|n)\,p_N(n)=\sum_{n=0}^{\infty}\binom{n}{k}p^k(1-p)^{n-k}\mathbb{1}_{k\le n}\frac{\lambda^n}{n!}e^{-\lambda}=\sum_{n=k}^{\infty}\dots$$

Def: Date X, Y come sopra, $\varphi: S_2\to\mathbb{R}$, se $\varphi(Y)$ è ≥ 0 o integrabile, valore atteso condizionale di $\varphi(Y)$ dato $X=x$, $x\in R_{p_X}$,

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

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

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

Valore atteso condizionale di $\varphi(Y)$ dato X:

funzione di x, valutata in $x=X(\omega)$

$$E[\varphi(Y)|X]=\sum_{x\in R_{p_X}}E[\varphi(Y)|X=x]\,\mathbb{1}_{\{X=x\}} \text{ `` }= E[\varphi(Y)|X=x]\Big|_{x=X}\text{ ''}$$

Oss: $E[\varphi(Y)|X=x]=:f(x)$ è una funzione di $x\in R_{p_X}$ (estesa a S_1 ponendo $E[\varphi(Y)|X=x]=0 \ \forall x\in S_1\setminus R_{p_X}$)

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

$$E[M]=\sum_{n=0}^{\infty}E[M|N=n]\,P\{N=n\}\underset{M\sim B(n,p)\text{ sotto }N=n}{=}\sum_{n=0}^{\infty}np\frac{\lambda^n}{n!}e^{-\lambda}=p\,E[N]=\lambda p$$

12
sabato 1 aprile 2023 17:18

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 $\mathcal{C} = \bigcap\limits_{\substack{\mathcal{F}\ \sigma\text{-algebra su } \Omega \\ \mathcal{F} \supseteq \mathcal{C}}} \mathcal{F}$

Oss: Intersezione di una famiglia di σ-algebre è σ-algebra (esercizio)
in particolare $\sigma(\mathcal{C}) = \bigcap\limits_{\mathcal{F}\ \sigma\text{-alg} \supseteq \mathcal{C}} \mathcal{F}$ 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) $\forall x \leq y$, $F(x) = P((-\infty, x]) \overset{\text{monotonia}}{\leq} P((-\infty, y]) = F(y)$

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, $F(x_n) \uparrow P(\mathbb{R}) = 1$

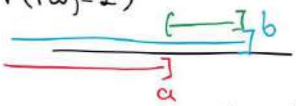
Prop (prob $\overset{1-1}{\longleftrightarrow}$ FdR):

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: $(F(-\infty) = 0,\ F(+\infty) = 1)$



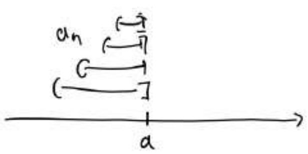
- $P((a,b]) = F(b) - F(a)$ $\quad \forall -\infty \leq a < b \leq +\infty$

  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

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



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

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

$P((a,b)) = F(b-) - F(a)$

$P([a,b)) = F(b-) - F(a-)$

} (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. $P(\Omega_0)=1$ $\quad (\Omega_0=\{x_1,x_2,\dots\})$

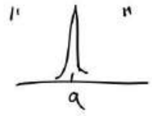
senza perdita di generalità, possiamo prendere $\Omega_0 = R_P$ il range di P

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 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$:

$$\delta_a(A) = \begin{cases} 1 & \text{se } a \in A \\ 0 & \text{se } a \notin A \end{cases}$$

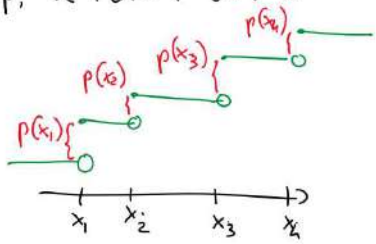
"  " massa concentrata in a

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

Data P discreta, con range $R_P=\{x_1,x_2,\dots\}$ e densità discreta p, la FdR F di P è

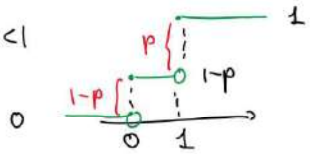
$$F(x)=\sum_{i,\, x_i \le x} p(x_i)$$

infatti $\quad F(x)=P((-\infty,x])=\sum_{x_i \in (-\infty,x]} p(x_i)$



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"

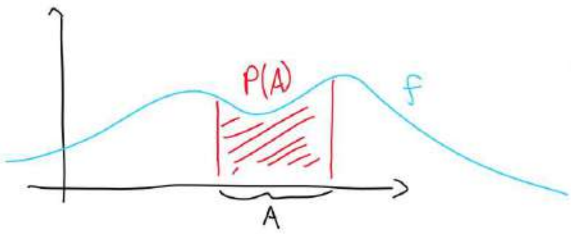
2) Probabilità assolutamente continue (rispetto alla misura di Lebesgue):

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., $h_n \uparrow h \implies \int h_n\,dx \uparrow \int h\,dx$

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

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

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

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

quindi

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

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

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

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 $P\{f=0\} = \int_{\{f=0\}} f\,dx = 0$

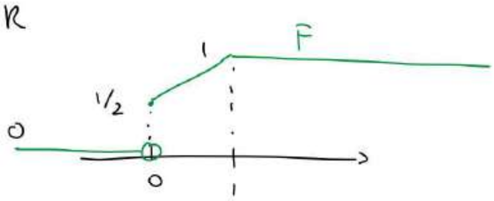
Def: Data P prob su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, supporto (topologico) di P: $\operatorname{supp} P \subseteq \mathbb{R}$ chiuso t.c.
$$(\operatorname{supp} P)^c = \bigcup_{A \text{ aperto} \subseteq \mathbb{R},\ P(A)=0} A$$

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

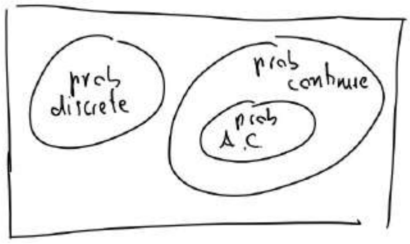
Esistono prob P su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ che non sono né discrete né continue

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

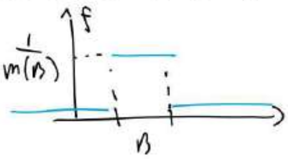


Esempi notevoli di probabilità assolutamente continue

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 $(\mathcal{U}(B))$

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

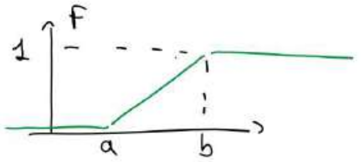


$$\forall A \in \mathcal{B}(\mathbb{R}),\ P(A) = \int_A \frac{1}{m(B)} \mathbb{1}_B(x)\,dx = \frac{m(A\cap B)}{m(B)}$$

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

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

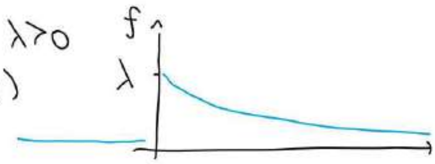
FdR: $$F(x) = \begin{cases} 0 & x<a \\ \frac{x-a}{b-a} & a \le x < b \\ 1 & x \ge b \end{cases}$$



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"

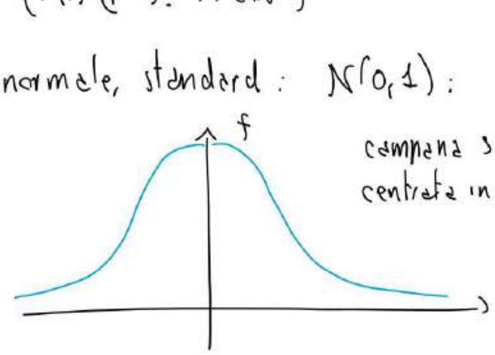
- Distribuzione Gamma di parametri $r>0$ e $\lambda > 0$ $(\Gamma(r,\lambda))$

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

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: $\mathcal{N}(0,1)$:

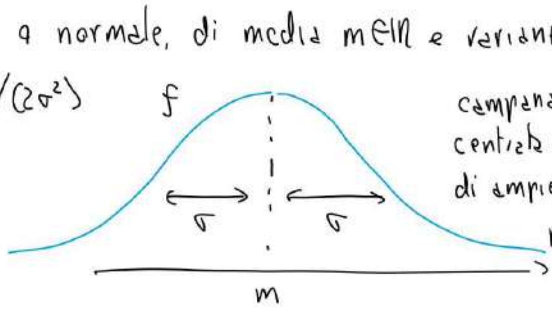
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$



campana simmetrica centrata in $x=0$

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

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



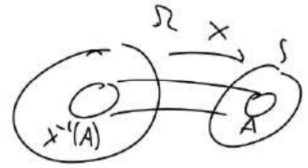
campana simmetrica centrata in $x=m$
di ampiezza proporzionale a σ
pt di flesso in $m \pm \sigma$

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

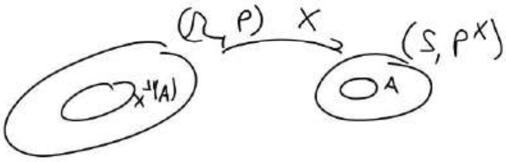
Ricordiamo: dato $A \subseteq S$, $X^{-1}(A) := \{X \in A\} := \{\omega \in \Omega \mid X(\omega) \in A\}$
X^{-1} commuta con le operazioni insiemistiche $\cup, \cap, {}^c$

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

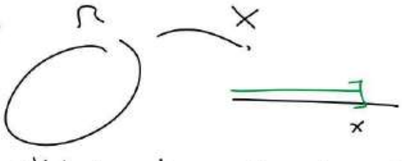
$$P^X(A) = P\{X \in A\} = P(X^{-1}(A)), \quad A \in \mathcal{S}$$



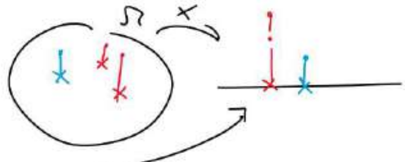
Lemma: P^X è una probabilità su $(S, \mathcal{S})$
Dim come nel caso discreto.

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

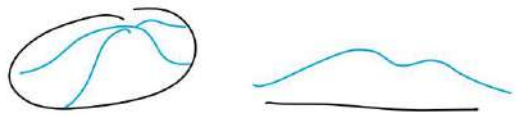


Oss: non è detto che Ω sia discreta

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

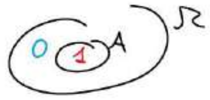
· diciamo che X è assolutamente continua con densità f, se P^X è assol. cont. con dens. f

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



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.

mercoledì 5 aprile 2023 11:46

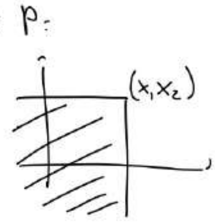
Distribuzioni congiunte e indipendenza per v.a. generali

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.
  $m([a_1, b_1] \times \ldots \times [a_d, b_d]) = (b_1 - a_1) \cdot \ldots \cdot (b_d - a_d)$ $\forall a_1, b_1, \ldots a_d, b_d$, 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 $F(x_1, \ldots x_d) = P((-\infty, x_1] \times \ldots \times (-\infty, x_d])$



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

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
$\mathcal{C} = \{(-\infty, x_1] \times \ldots \times (-\infty, x_d] \mid (x_1, \ldots x_d) \in \mathbb{R}^d\}$. Poiché $\mathcal{C}$ è chiusa per inters. finite 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.
  $P(A) = \int_A f(x)\, dx$ $\forall A \in \mathcal{B}(\mathbb{R}^d)$ (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

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

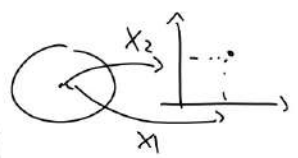
Viceversa, se P ha F che soddisfa (Δ) per qualche densità f (cioè $f: \mathbb{R}^d \to \mathbb{R}$ boreliana che soddisfa (i) e (ii)), allora P ha densità f.

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 $f(x_1 .. x_d) = \partial_{x_1} \cdots \partial_{x_d} F(x_1 \dots x_d)$"

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

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



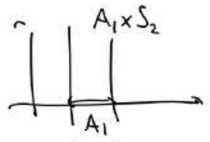
chiamiamo $(X_1, .. X_n): \Omega \to S_1 \times .. \times S_n$, $(X_1, .. X_n)(\omega) = (X_1(\omega), .. X_n(\omega))$

vettore aleatorio

Ricordiamo

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

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



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

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

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

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

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

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

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

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

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

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

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

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

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

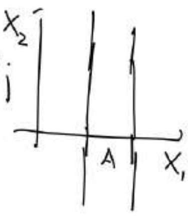
Oss: $\forall A_i \in \mathcal{J}_i$, $P\{X_1 \in A_1, \dots X_n \in A_n\} = P\{(X_1, \dots X_n) \in A_1 \times \dots \times A_n\} = P_{(X_1, \dots X_n)}(A_1 \times \dots \times A_n)$

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

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

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

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



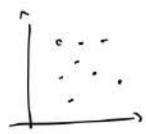
Dim: (come nel caso discreto)

La prima formula segue dall'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 $\# R_{P_{X_1}} \times \dots \times R_{P_{X_n}} \le \# \mathbb{N}$, 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

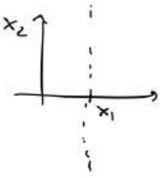
$$p_{X_i}(x) = \sum_{x_j \in \mathbb{R},\, j \neq i} p_{(X_1, \dots X_n)}(x_1, \dots x_{i-1}, x, x_{i+1}, \dots x_n) \qquad x \in \mathbb{R}$$

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

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., $(S_i, \mathcal{S}_i)$ 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

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

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

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 $i=1,..n$, 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 $X_1 .. X_n$ sono indipendenti, allora

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

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

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

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

Quindi $(X_1 .. X_n)$ e $(Y_1 .. Y_n)$ hanno la stessa FdR, quindi hanno la stessa legge.

In particolare, $X_1 .. X_n$ sono indipendenti

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

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

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

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

(come visto nel caso discreto)

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

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

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

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

$$
\begin{aligned}
P\{X_1 \in A_1, \dots X_n \in A_n\} &= \int_{A_1 \times \dots \times A_n} f(x_1, \dots x_n)\, dx_1 \dots dx_n \\
\text{ipotesi} \quad &= \int_{A_1 \times \dots \times A_n} f(x_1) \dots f(x_n)\, dx_1 \dots dx_n \\
\text{Fubini Tonelli} \quad &= \int_{A_1} \int_{A_2} \dots \int_{A_n} f(x_1) f(x_2) \dots f(x_n)\, dx_n \dots dx_2\, dx_1 \\
&= \int_{A_1} f(x_1)\, dx_1 \cdot \ \dots \ \cdot \int_{A_n} f(x_n)\, dx_n = P\{X_1 \in A_1\} \cdot \dots \cdot P\{X_n \in A_n\}
\end{aligned}
$$

quindi $X_1 \dots X_n$ sono indipendenti

b) Usiamo il criterio per FdR

$$
\begin{aligned}
F(x_1, \dots x_n) &\underset{\text{indip.}}{=} F(x_1) \cdot \dots \cdot F(x_n) = \\
&= \int_{-\infty}^{x_1} f(y_1)\, dy_1 \cdot \dots \cdot \int_{-\infty}^{x_n} f(y_n)\, dy_n \\
\text{Fubini Tonelli} \quad &= \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(y_1) \cdot \dots \cdot f(y_n)\, dy_1 \dots dy_n \qquad \forall (x_1, \dots x_n) \in \mathbb{R}^n
\end{aligned}
$$

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

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

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 $P_{(X_1 \dots X_n)}(A_1 \times \dots \times A_n) = Q(A_1 \times \dots \times A_n)$ $\forall A_i \in \mathcal{B}(\mathbb{R})$

$$
\begin{aligned}
P_{(X_1 \dots X_n)}(A_1 \times \dots \times A_n) &= P\{X_1 \in A_1, \dots X_n \in A_n\} \\
\text{indip} \quad &= P\{X_1 \in A_1\} \cdot \dots \cdot P\{X_n \in A_n\} \\
&= \int_{A_1} f_1(x_1)\, dx_1 \cdot \dots \cdot \int_{A_n} f_n(x_n)\, dx_n \\
\text{Fubini Tonelli} \quad &= \int_{A_1 \times \dots \times A_n} f(x_1) \cdot \dots \cdot f(x_n)\, dx_1 \dots dx_n
\end{aligned}
$$

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

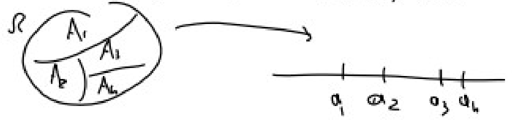
Valore atteso, momenti e varianza

$(\Omega, \mathcal{F}, P)$ 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
$$X = \sum_{i=1}^m a_i 1\!\!1_{A_i}$$



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

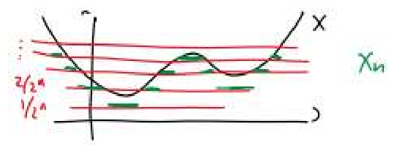
Data X semplice, $X = \sum_{i=1}^m a_i 1\!\!1_{A_i}$, definiamo
$$E[X] = \int_\Omega X \, dP := \sum_{i=1}^m a_i P(A_i)$$

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



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

$$E[\underline{\lim}_n X_n]\leq \underline{\lim}_n E[X_n]$$

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 $\Leftrightarrow \int_S |\varphi(x)| P^X(dx) < \infty$
- 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. $\varphi = 1\!\!1_A$, 2. φ semplice, 3. $\varphi \geq 0$, 4. φ generica

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

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 $\Leftrightarrow E[|\varphi(X)|] \overset{(3)}{=} \int |\varphi(x)| P^X(dx) < \infty$
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 $\Leftrightarrow E[|\varphi(X)|] = \sum_{x \in R_{p_X}} |\varphi(x)| p_X(x) < \infty$
  - 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 $\Leftrightarrow E[|\varphi(X)|] = \int_{\mathbb{R}} |\varphi(x)| f(x) dx < \infty$
  - se $\varphi(X)$ è integrabile o ≥ 0 q.c, $E[\varphi(X)] = \int_{\mathbb{R}} \varphi(x) f(x) dx$

Dim:

Secondo 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)

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

4. φ generica: le 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:

- $X \sim \mathcal{U}((a,b))$, $a<b$ $\quad \left(f(x) = \frac{1}{b-a}\mathbb{1}_{(a,b)}(x)\right)$
  X è limitata q.c., quindi $E[|X|] = \int |x| f(x)\,dx < \infty$
  $$E[X] = \int x f(x)\,dx = \frac{1}{b-a}\int_a^b x\,dx = \frac{1}{b-a}\,\frac{1}{2}(b^2-a^2) = \frac{a+b}{2}$$

- $X \sim \mathcal{E}xp(\lambda)$, $\lambda > 0$ $\quad \left(f(x) = \lambda e^{-\lambda x}\mathbb{1}_{(0,+\infty)}(x)\right)$
  $X \geq 0$ q.c., quindi $\exists\, E[X] \in [0, +\infty]$
  $$E[X] = \int_0^{+\infty} x\,\lambda e^{-\lambda x}\,dx \underset{\text{per parti}}{=} -xe^{-\lambda x}\Big|_{x=0}^{+\infty} + \int_0^{\infty} e^{-\lambda x}\,dx = 0 - \frac{1}{\lambda}e^{-\lambda x}\Big|_{x=0}^{+\infty} = \frac{1}{\lambda}$$

- $X \sim \Gamma(r,\lambda)$, $r>0$, $\lambda>0$
  $$E[X] = \frac{r}{\lambda}$$

- $X \sim \mathcal{N}(0,1)$ $\quad \left(f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\right)$
  $$E[|X|] = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} |x| e^{-x^2/2}\,dx < \infty$$
  $$E[X] = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} x e^{-x^2/2}\,dx = \frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^0 x e^{-x^2/2}\,dx + \int_0^{+\infty} x e^{-x^2/2}\,dx\right) = 0$$
  $$\int_{-\infty}^0 x e^{-x^2/2}\,dx \underset{y=-x}{=} \int_0^{+\infty} -y\, e^{-y^2/2}\,dy$$

- $X \sim \mathcal{N}(m,\sigma^2)$
  $E[X] = m$ (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 $(p-1)!! = (p-1)\cdot(p-3)\cdot \dots \cdot 3 \cdot 1$

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:

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

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

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

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

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

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

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

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

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

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

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

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

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

$\mathbb{1}_A, \mathbb{1}_B$ sono indip $\iff$ A e B sono indip.

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

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

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

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

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

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

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

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

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

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. $\iff f(x,y) = f_X(x)f_Y(y)$
con f_X e f_Y densità di X e Y risp.

$$E[|XY|] = \iint_{\mathbb{R}^2} |xy| f(x,y)\,dx\,dy \underset{\text{indip}}{=} \iint_{\mathbb{R}^2} |x| f_X(x) |y| f_Y(y)\,dx\,dy$$
$$\underset{\text{Fubini-Tonelli}}{=} \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} |x| f_X(x)\,dx\Big) |y| f_Y(y)\,dy = \int_{\mathbb{R}} |x| f_X(x)\,dx \int_{\mathbb{R}} |y| f_Y(y)\,dy$$
$$= E[|X|]\,E[|Y|] < \infty \text{ se } X, Y \text{ sono integrabili}$$

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

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

[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

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

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

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

$$=\int 1\!\!1_A(y)\,f_Y(y)\,dy$$

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):

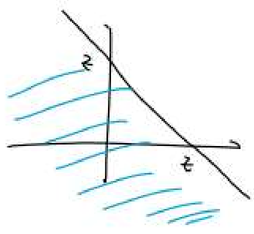
Siano $X:\Omega\to\mathbb{R}$, $Y:\Omega\to\mathbb{R}$ v.a. reali con (X,Y) assolutamente continua di densità $f_{(X,Y)}$.
Allora $X+Y$ è assolutamente continua, con densità

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

Dim:

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

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



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

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

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

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

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

V. a. gaussiane: 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

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

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

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 $-\infty \leq a < b \leq +\infty$, $P\{a \leq X \leq b\} = ?$

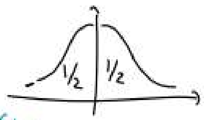
$P\{a \leq X \leq b\} = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx$ ma questo integrale non ha una forma esplicita (salvo per particolari valori di a e b)

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

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

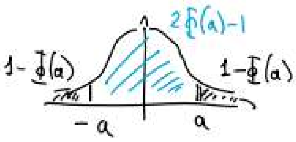
  e, per molti $x \in \mathbb{R}$, $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$ è noto con ottima approssimazione (ed è presente nei software più comuni)

  Proprietà di Φ:

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

    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 $x \leq -5$

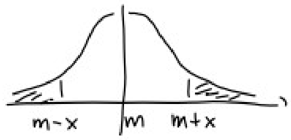
    cioè $P\{-5 \leq X \leq 5\}$ è quasi 1.
  - $P\{-3 \leq X \leq 3\} \approx 0.997$

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

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

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

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

$$= \exp\left(-\frac{1}{2(\sigma^2+1)} z^2\right) \cdot \frac{1}{2\pi\sigma} \int \exp\left(-\frac{1}{2\sigma^2}\left(\sqrt{\sigma^2+1}\,x - \frac{1}{\sqrt{\sigma^2+1}} z\right)^2\right) dx \qquad \begin{aligned} y &= \sqrt{\sigma^2+1}\,x - \tfrac{1}{\sqrt{\sigma^2+1}} z \\ dy &= \sqrt{\sigma^2+1}\,dx \end{aligned}$$

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

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

2. Caso generale

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

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

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 $X_i : \Omega \to S$, $(S, \mathcal{S})$ sp. misurabile, X_i si dicono indipendenti e identicamente distribuite (i.i.d.): se sono indipendenti e hanno la stessa legge su $(S, \mathcal{S})$

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

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

cioè $P\{|\overline{X}_n - m| > \varepsilon\} \to 0$ 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 $(S, \mathcal{S})$ 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 $\varphi = \mathbb{1}_A$, $A \in \mathcal{S}$, $(\mathbb{1}_A(X) = \mathbb{1}_{X \in A} \sim B(p)$ 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 questa 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. $\int_{(0,1)^d}|\varphi(x)|^2dx<\infty$,
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 $1\!\!1_{(0,1)^d}$)

Notare che

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

Teor (Teorema centrale del limite, TCL/TLC)

Sia $X_1, \ldots X_n, \ldots$ 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

$$\frac{\sum_{i=1}^{n}X_i-nm}{\sqrt{n}\,\sigma}=\frac{\sqrt{n}}{\sigma}(\overline{X}_n-m)$$ converge in legge a $Z\sim\mathcal{N}(0,1)$, cioè

detta F_n la FdR di $\frac{\sqrt{n}}{\sigma}(\overline{X}_n-m)$, 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

$$P\{a\leq\frac{\sqrt{n}}{\sigma}(\overline{X}_n-m)\leq b\}\to P\{a\leq Z\leq b\}\quad\forall a,b\in\mathbb{R},\ a<b$$

e analogamente per le disuguaglianze strette

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

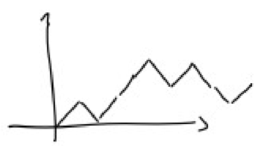
Oss: Per standardizzazione, $\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n}X_i-nm\right)=\sqrt{n}(\overline{X}_n-m)\xrightarrow{\text{in legge}}Z_\sigma$, con $Z_\sigma\sim\mathcal{N}(0,\sigma^2)$

cioè $P\{a\leq\sqrt{n}(\overline{X}_n-m)\leq b\}\to P\{a\leq Z_\sigma\leq b\}\quad\forall a,b\in\mathbb{R},\ a<b$

Applicazione: convergenza delle leggi marginali di una passeggiata aleatoria

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

$S_n=X_1+\ldots+X_n$ 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

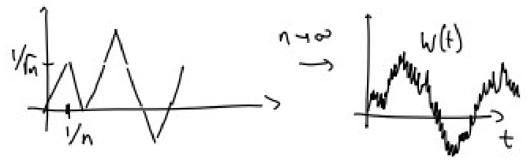
$$\left(\frac{1}{\sqrt{n}} S_{[nt_1]}, \dots \frac{1}{\sqrt{n}} S_{[nt_n]}\right) \xrightarrow{\text{"in legge"}} (W_{t_1}, \dots W_{t_n}) \qquad \forall\, t_1 < t_2 < \dots < t_n$$

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)

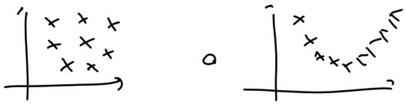
Ad es: dati 20 individui, x_i = altezza, y_i = peso dell' i-simo individuo

Alcuni indici statistici rilevanti:

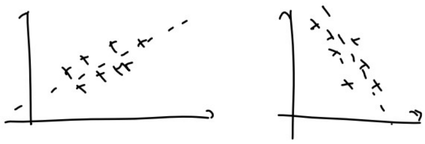
- coefficiente di correlazione campionario
  $$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\left(\sum_{i=1}^{n} (x_i - \bar{x})^2\right)^{1/2} \left(\sum_{i=1}^{n} (y_i - \bar{y})^2\right)^{1/2}} \quad \in [-1, 1]$$

indice quanto i dati sono allineati

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



- $|r| \approx 1$: 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 $x_1, \dots x_n$, $\left(x_i = \begin{cases} 1 & \text{se esce testa all'i-simo lancio} \\ 0 & \text{altrimenti} \end{cases}\right)$ vogliamo avere info sulla prob p di testa (non nota):

1) stima per p? $\rightsquigarrow$ stimatore ($\overline{x}$ = 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)? $\rightsquigarrow$ 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

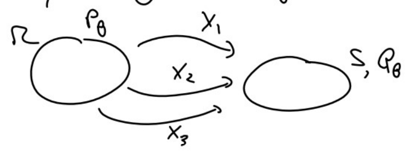
  $\theta = (m, \sigma^2) \in \mathbb{R} \times (0, +\infty) = \Theta$

  $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$

  $Q_{(m,\sigma^2)} = \mathcal{N}(m, \sigma^2)$

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 $p_\theta(x_1, \dots x_n) = q_\theta(x_1) \cdot \dots \cdot q_\theta(x_n)$ (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 $f_\theta(x_1, \dots x_n) = g_\theta(x_1) \cdot \dots \cdot g_\theta(x_n)$ (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 $p_\theta(x_1, \dots x_n) = q_\theta(x_1) \cdot \dots \cdot q_\theta(x_n)$ con q_θ densità discreta di Q_θ

  nel caso assol. cont: avente densità $f_\theta(x_1, \dots x_n) = g_\theta(x_1) \cdot \dots \cdot g_\theta(x_n)$ 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 $E^\theta[X_1^4] < \infty$ $\forall\theta\in\Theta$, $(S_{n-1}^2)_{n\in\mathbb{N}}$ è una successione di stimatori consistenti di $Var^\theta(X_1)$
  infatti per LGN,
$$S_{n-1}^2 = \underbrace{\frac{n}{n-1}}_{\to 1}\left(\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i^2}_{\xrightarrow{P_\theta} E[X_1^2]} - \underbrace{\bar{X}_n^2}_{\xrightarrow{P_\theta} E[X_1]}\right) \xrightarrow{P_\theta} E^\theta[X_1^2] - E^\theta[X_1]^2 = Var(X_1)$$
  (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 $(\mathbb{R}, \mathcal{B}(\mathbb{R}), (Q_\theta)_{\theta\in\Theta})$ 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 i.i.d. $X_1, \dots X_n$ 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 i.i.d. $X_1, \dots X_n$ 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 $\theta = \bar{x}$

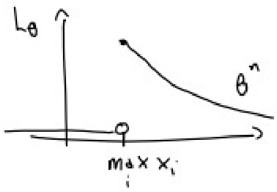
quindi $\bar{X}$ è lo stimatore di max. verosimiglianza di θ

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

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

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

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



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 $\mathbb{N}$

  $p_\theta(k) = c_\theta\, g(k)\, e^{\theta T(k)}$ $\quad k\in\mathbb{N}$ $\qquad$ ($c_\theta>0$ 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

  $f_\theta(x) = c_\theta\, g(x)\, e^{\theta T(x)}$ $\quad x\in\mathbb{R}$ $\qquad$ ($c_\theta>0$ costante)

Esempi:

1. leggi esponenziali: $f_\lambda(x) = \lambda e^{-\lambda x} 1\!\!1_{(0,+\infty)}(x)$ $\quad \theta=\lambda$, $T(x)=-x$, $g(x)=1\!\!1_{(0,+\infty)}(x)$, $c_\lambda=\lambda$
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 $\mathcal{N}(m,\sigma^2)$ ($\sigma^2=1$ 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 $p_\lambda(k) = p(1-p)^{k-1} 1\!\!1_{k\ge1}$ $\quad \theta=\log(1-p)$ $\quad T(k)=k-1$, $g(k)=1\!\!1_{k\ge1}$, $c_\theta=p$
5. leggi unif. su $(0,\theta)$, $f_\theta(x) = \frac{1}{\theta}1\!\!1_{[0,\theta]}(x)$, 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.: $f_\theta(x) = c_\theta\, g(x)\, e^{\theta T(x)}$
- $x\mapsto g(x)T(x)^2 e^{\theta T(x)}$ 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 $c_\theta g(x) e^{\theta t} = c\, g(x)$ indipendendente (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 $(\bar{X},\frac{n-1}{n}S^2)$ 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 $\theta=m\in\Theta=\mathbb{R}$), $\bar{X}$ 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: Data $\alpha \in (0,1)$, 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.

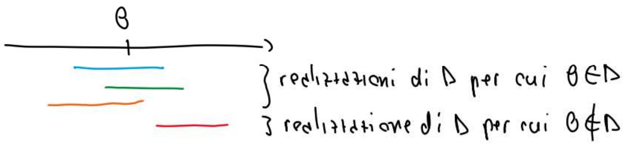
$$P_\theta\{\theta\in D\} \geq 1-\alpha \quad \forall \theta\in\Theta$$

(dove $\{\theta\in D\} = \{\omega\in\Omega \mid \theta\in D(\omega)\} \subseteq \Omega$ e si sottointende che $\{\theta \in D\} \in \mathcal{F} \quad \forall\theta\in\Theta$)

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)"



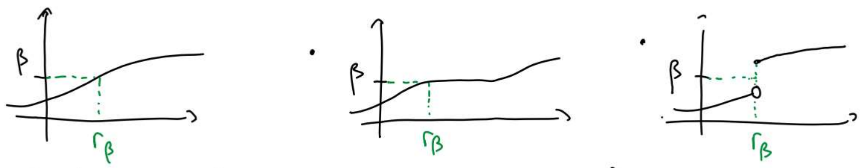
Quantili di una prob. P su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, con FdR F

Def: Dato $\beta\in(0,1)$, quantile di ordine β (o β-quantile) di P (o di F): il numero

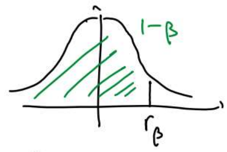
$$r_\beta = \inf\{x\in\mathbb{R} \mid F(x) \geq \beta\}$$

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)

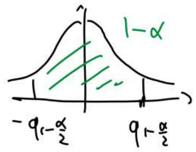


I quantili di $\mathcal{N}(0,1)$ si denotano con q_β o z_β



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

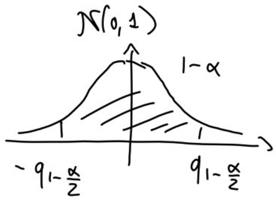
  Poiché $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ è uno stimatore di m, è ragionevole cercare D intervallo centrato in $\bar{X}$,

  $$D = [\bar{X} - d, \bar{X} + d] =: [\bar{X} \pm d]$$

  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: $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - m)$ è funzione anche di σ non nota

Cor: Date $X_1, \dots X_n$ i.i.d. con $E[X_i^2] < \infty$,

$$\frac{\sqrt{n}}{S_n}(\bar{X}_n - m) \xrightarrow{\text{in legge}} Z \sim N(0,1) \quad \text{per } n \to \infty$$

con $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$

Questo risultato si dimostra come corollario del TCL e di $S_{n-1}^2 \to \sigma^2$ (usando fatti più avanzati)

Usando $ST = \frac{\sqrt{n}}{S_n}(\bar{X}_n - m)$ come statistica pivotale (asintotica) si ottiene che

$\left[\bar{X}_n \pm \frac{\sqrt{n}}{S_n} q_{1-\frac{\alpha}{2}}\right]$ 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 $\left[\bar{X} \pm \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} q_{1-\frac{\alpha}{2}}\right] \approx \left[\bar{X} \pm \frac{\sqrt{\bar{X}(1-\bar{X})}}{10} \cdot 1.96\right] = \left[\bar{X} \pm 0.196 \cdot \sqrt{\bar{X}(1-\bar{X})}\right]$

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:

$$ST = \frac{\sqrt{n}}{S}(\bar{X} - m) \sim t_{n-1} \quad \text{(t-di-Student a n-1 gradi di libertà)}$$

(con $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$, $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$)

In particolare, chiamando $t_{\beta, n-1}$ il quantile di t_{n-1} di ordine β ($P\{\overset{\sim t_{n-1}}{T} \le t_{\beta, n-1}\} = \beta$),

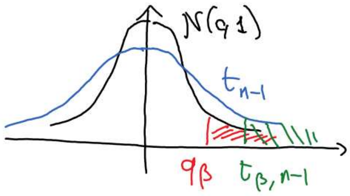
$$P\{ST \in [-t_{1-\frac{\alpha}{2}, n-1}, t_{1-\frac{\alpha}{2}, n-1}]\} = 1-\alpha$$

Quindi usando ST come statistica pivotale si ottiene che

$\left[\bar{X} \pm \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}, n-1}\right]$ 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 $t_{\beta, n-1} > q_\beta \quad \forall n, \ \forall \beta > \frac{1}{2}$



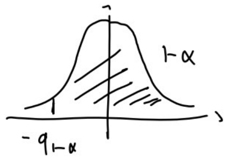
area rossa = area verde = $1-\beta$

Dunque l'incertezza nel caso varianza non nota è maggiore rispetto al caso varianza nota

Oss: Si può dimostrare che, per $n \to \infty$, $t_{n-1} \xrightarrow{\text{in legge}} N(0,1)$, 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 $N(m,\sigma^2)$ di varianza σ^2 nota, possiamo cercare una regione di fiducia per la media m della forma $(-\infty, \bar{X}+d]$. Imponiamo $P_m\{m \in (-\infty, \bar{X}+d]\} = 1-\alpha$ e usiamo la statistica pivotale $ST = \frac{\sqrt{n}}{\sigma}(\bar{X}-m)$

$$P_m\{m \in (-\infty, \bar{X}+d]\} = P_m\{\bar{X}+d \geq m\} = P_m\left\{\frac{\sqrt{n}}{\sigma}(\bar{X}-m) \geq -\frac{\sqrt{n}}{\sigma}d\right\}$$
$$= \Phi\left(\frac{\sqrt{n}}{\sigma}d\right)$$
$$= 1-\alpha \iff \frac{\sqrt{n}}{\sigma}d = q_{1-\alpha}$$



e troviamo $\left(-\infty, \bar{X} + \frac{\sigma}{\sqrt{n}} q_{1-\alpha}\right]$ come una regione di fiducia di livello $1-\alpha$ per m

Intervalli di fiducia per la varianza di una popol. gaussiana,

- $(S, \mathcal{S}, (P_\theta)_{\theta\in\Theta}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), (N(m,\sigma^2))_{\sigma^2 \in (0,+\infty)})$, $\sigma^2 = Var_{\sigma^2}(X)$ con $X \sim N(m,\sigma^2)$

  $X_1, \dots X_n$ campione i.i.d.

  supponiamo m non nota

  Usiamo

  $$ST = (n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1} \quad \text{(chi-quadrato a } n-1 \text{ gradi di libertà)}$$

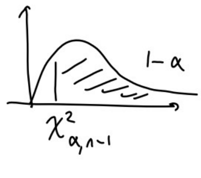
  In particolare, chiamando $\chi^2_{\beta,n-1}$ il quantile di ordine β di χ^2_{n-1} $(P\{Y \leq \chi^2_{\beta,n-1}\} = \beta)$, $Y \sim \chi^2_{n-1}$

  $$P\{ST \in [\chi^2_{\alpha,n-1}, +\infty)\} = 1-\alpha$$

  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) $Z \sim N(0,1) \Rightarrow Z^2 \sim \Gamma\left(\frac{1}{2}, \frac{1}{2}\right) =: \chi^2_1$ (distribuzione chi-quadrato a un grado di libertà)

b) $Z_1, \dots Z_n \sim N(0,1)$ i.i.d. $\Rightarrow Z_1^2 + \dots + Z_n^2 \sim \Gamma\left(\frac{n}{2}, \frac{1}{2}\right) =: \chi^2_n$ (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 $\Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$,

$$f(u) = \frac{1}{\sqrt{2}\,\Gamma\left(\frac{1}{2}\right)} u^{-1/2} e^{-\frac{1}{2}u} \mathbb{1}_{(0,+\infty)}(u)$$

$F_{Z^2}(u) = 0 \quad \forall u < 0$. 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 $\sim \Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$. Quindi per riproducibilità di Γ, $Z_1^2 + \dots + Z_n^2 \sim \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$.

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

$$f_Z(x_1, \dots x_n) = \frac{1}{\sqrt{2\pi}} e^{-x_1^2/2} \cdot \dots \cdot \frac{1}{\sqrt{2\pi}} e^{-x_n^2/2} = \frac{1}{(2\pi)^{n/2}} e^{-|x|^2/2} \qquad \left(\text{dove } |x|^2 = \sum_{i=1}^n x_i^2\right)$$

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) $\bar{Z} \sim N\left(0, \frac{1}{n}\right)$, $(n-1)S_n^2 \sim \chi^2_{n-1}$

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 $F_{t_n}(x) = \int_{-\infty}^{x} f_{t_n}(x')\,dx'$

$$F_{t_n}(x) = \iint \mathbb{1}_{\sqrt{n}\,z/\sqrt{y} \le x}\, f_{\mathcal{N}(0,1)}(z)\, f_{\chi^2_n}(y)\,dy\,dz$$

$$= \iint \mathbb{1}_{\sqrt{n}\,z/\sqrt{y} \le x}\, c_n\, e^{-z^2/2}\, y^{-n/2} e^{-y/2}\,dy\,dz \qquad \text{(con } c_n > 0 \text{ costante)}$$

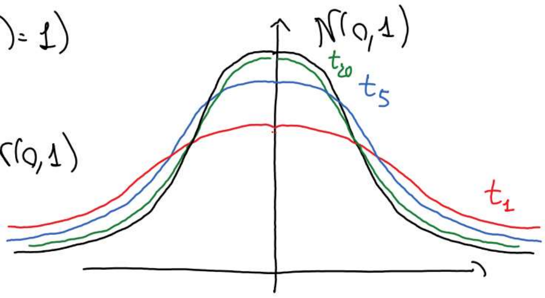
$\frac{\sqrt{n}\,z}{\sqrt{y}} = x'$, $\frac{\sqrt{n}}{\sqrt{y}}\,dz = dx'$

$$= \int_{-\infty}^{x} c_n' \int e^{-(x')^2 y/2n}\, e^{-y/2}\, y^{-(n-1)/2}\,dy\,dx' \qquad \text{(con } c_n' > 0 \text{ costante)}$$

$\left(1 + \frac{(x')^2}{n}\right) y = \tilde{y}$, $\left(1 + \frac{(x')^2}{n}\right) dy = d\tilde{y}$

$$= \int_{-\infty}^{x} c_n' \left(1 + \frac{(x')^2}{n}\right)^{-\frac{n+1}{2}} \underbrace{\int e^{-\tilde{y}/2}\, \tilde{y}^{-\frac{n-1}{2}}\,d\tilde{y}}_{\text{cost.}}\,dx' = \int_{-\infty}^{x} c_n''\, f_{t_n}(x')\,dx'$$

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 lanci escono $\begin{cases} 501 \text{ teste, l'ipotesi } p=\frac{1}{2} \text{ è plausibile} \\ 900 \text{ '' , '' '' '' non plausibile} \\ 540 \text{ '' , '' '' '' ?} \end{cases}$

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 $(\Theta_0 \cap \Theta_1 = \emptyset)$

$H_0: \theta \in \Theta_0$ ipotesi nulla, di cui verificare se è plausibile o no

$H_1: \theta \in \Theta_1$ " alternativa

Nell'esempio precedente: $H_0: \theta = \frac{1}{2}$, $\Theta_0 = \{\frac{1}{2}\}$, $\Theta_1 = [0,1] \setminus \{\frac{1}{2}\}$

Oss: asimmetria tra H_0 e H_1:

- se l'esito del test, sulla base dei dati, è il rifiuto di H_0, in favore di H_1, allora H_0 è poco plausibile, cioè c'è evidenza statistica per H_1
- se invece l'esito del test è l'accettazione di H_0, allora H_0 è plausibile, ma non c'è per forza evidenza per 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 H_0 in favore di H_1

- se l'esito ω non cade in C, accettiamo l'ipotesi H_0.

La scelta di C deve essere t.c.:
- C sia poco probabile sotto H_0 (poco plausibile)
- se $\omega \in C$, allora questa è un'evidenza per H_1

Nell'esempio precedente $(H_0: p=\frac{1}{2})$, ci aspettiamo una regione critica C della forma

$C = \{ |\bar{X} - \frac{1}{2}| > d \}$ per un $d>0$ opportuno, per cui $P_{\frac{1}{2}}(C)$ sia piccola

Se invece consideriamo $H_0: p \leq \frac{1}{2}$, ci aspettiamo una regione critica C della forma

$C = \{ \bar{X} - \frac{1}{2} > d \}$ 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: H_0 è vera ma l'esito del test è il rifiuto di H_0
- " " 2^a specie: H_0 è falsa ma l'esito del test è l'accettazione di H_0

Def: Dato $\alpha \in (0,1)$ (piccolo, ad 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 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 H_0 quando il valore del parametro è $\theta \in \Theta_1$, quindi $\pi_C(\theta)$ rappresenta la capacità di accorgersi che l'ipotesi 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 $H_0 : \theta = \theta_0$, 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 $H_0: m = m_0$ contro $H_1: m \neq m_0$, a livello α

Regione critica C: poiché $\bar{X}$ è uno stimatore di m e $H_1: m \neq m_0$, è ragionevole cercare C della forma $C = \{|\bar{X} - m_0| > d\}$ per d opportuno

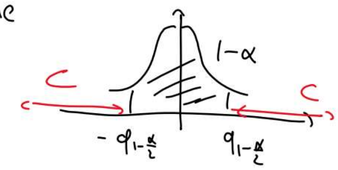
Usiamo la statistica di test

$$ST = \frac{\sqrt{n}}{\sigma}(\bar{X} - m) \sim N(0,1) \text{ sotto } P_m$$

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: $\omega \in C \iff |\bar{X}(\omega) - m_0| > \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} \iff m_0 \in [\bar{X}(\omega) \pm \frac{\sigma}{\sqrt{n}} q_{1-\alpha}]$ (Int. di fiducia per m_0 di livello $1-\alpha$)

Test bilatero per una proporzione (media di popol. Bernoulli), grandi campioni (test Z approssimato)

- $(S, \mathcal{J}, (Q_\theta)_\Theta) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), B(p)_{p \in [0,1]})$ $\quad \theta = p \in \Theta = [0,1]$

  $X_1, \dots, X_n$ 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

  $$C = \left\{\sqrt{\frac{n}{p_0(1-p_0)}}|\bar{X} - p_0| > q_{1-\frac{\alpha}{2}}\right\} = \left\{\frac{|n\bar{X} - np_0|}{\sqrt{np_0(1-p_0)}} > q_{1-\frac{\alpha}{2}}\right\}$$

  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 $\Omega = \{(x_1 \dots x_n) \mid x_i \in S\} = S^n$)

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 $\bar{\alpha} = \bar{\alpha}(X_1 \dots X_n)$) 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:

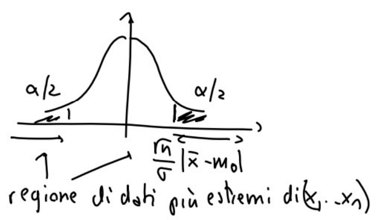
- media di popol. normale $\mathcal{N}(m,\sigma^2)$, varianza nota, $\mathcal{H}_0: m = m_0$, $\mathcal{H}_1: m \neq m_0$

  $C_\alpha = \left\{ \frac{\sqrt{n}}{\sigma} |\bar{X} - m_0| > q_{1-\frac{\alpha}{2}} \right\}$

  $\bar{\alpha}(x_1 \dots x_n)$ 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.

  $(S, \mathcal{F}, (Q_\theta)_\Theta) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{N}(m,\sigma^2)_{m \in \mathbb{R}})$, $\sigma = 0.1$ (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.

$(S, \mathcal{S}, (Q_\theta)_\theta) = (\mathbb{N}, \mathcal{B}(\mathbb{N}), \mathcal{B}(p)_{p\in[0,1]})$ p=prob di testa

$X_i = \begin{cases} 1 & \text{se testa al lancio i-simo} \\ 0 & \text{altrimenti} \end{cases}$ $i=1, .. n=1000$, campione i.i.d. di $\mathcal{B}(p)$
(grande)

$H_0: p = \frac{1}{2} (= p_0)$ (affermazione del banco) $H_1: p \neq \frac{1}{2}$

Statistica di test: $ST = \sqrt{\frac{n}{p(1-p)}}(\bar{X}-p) \overset{\text{approx}}{\sim} \mathcal{N}(0,1)$

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 $\bar{x} = \frac{540}{1000} = 0.54$, $63.25 \cdot |\bar{x} - 0.5| = 2.53 > 1.96$: $(x_1, .. x_n) \in C$, 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 $\tilde{C} \in \mathcal{J}^{\otimes n}$)

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 $(S, \mathcal{F}, (Q_\theta)_{\theta\in\Theta})$ 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 $H_0: \theta \leq \theta_0$ contro $H_1: \theta > \theta_0$, 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 $(S, \mathcal{F}, (Q_\theta)_{\theta\in\Theta})$ modello statistico con $\Theta = \{\theta_0, \theta_1\}$, consideriamo il test $H_0: \theta = \theta_0$, $H_1: \theta = \theta_1$, 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) $$(\mathbb{1}_{\tilde{C}} - P_{\theta_0}(C))(L_{\theta_0} - cL_{\theta_1}) = \underbrace{(1 - P_{\theta_0}(C))}_{\geq 0} \underbrace{\mathbb{1}_{\tilde{C}}(L_{\theta_0} - cL_{\theta_1})}_{\leq 0} - P_{\theta_0}(C) \underbrace{\mathbb{1}_{\tilde{C}^c}(L_{\theta_0} - cL_{\theta_1})}_{\geq 0} \leq 0$$

Integrando come prima,

$$c(P_{\theta_1}(C) - P_{\theta_0}(C)) \geq 0$$

Dim del teor:

i) Basta dim. che $\theta_1 \leq \theta_2 \leq \theta_0 \Rightarrow P_{\theta_1}(C) \leq P_{\theta_2}(C)$

$$C = \{T \geq d\} \underset{\substack{\text{rapporto} \\ \text{a ver cresc}}}{=} \{\underbrace{f_{\theta_1,\theta_2}(T)}_{= \frac{L_{\theta_2}(X_1 \dots X_n)}{L_{\theta_1}(X_1 \dots X_n)}} \geq \underbrace{f_{\theta_1,\theta_2}(d)}_{=: 1/c}\} = \{L_{\theta_1}(X_1 \dots X_n) \leq c L_{\theta_2}(X_1 \dots X_n)\}$$

quindi per Lemma di N.-P. (b) (con $\Theta=\{\theta_1,\theta_2\}$), $P_{\theta_1}(C) \leq P_{\theta_2}(C)$

ii) Dobbiamo dim. che, se D è un'altra regione critica di livello $P_{\theta_0}(C)$, $\forall \theta>\theta_0$, $P_\theta(D) \leq P_\theta(C)$

Poiché $P_{\theta_0}(D) \leq P_{\theta_0}(C)$, per lemma di N-P. (a) (con $\Theta=\{\theta_0,\theta\}$), $P_\theta(D) \leq P_\theta(C)$.

Test unilatero per la media di una popol. gaussiana, varianza nota

- $(S,\mathcal{S},(Q_\theta)) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{N}(m,\sigma^2)_{m\in\mathbb{R}})$ $\quad \theta=m\in\mathbb{R}$, σ^2 nota

$X_1 .. X_n$ campione i.i.d.

Consideriamo il test $H_0: m \leq m_0$ 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

$$ST = \frac{\sqrt{n}}{\sigma}(\bar{X}-m) \sim \mathcal{N}(0,1) \text{ sotto } P_m$$

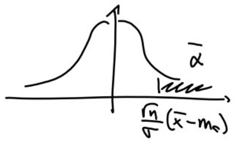
Imponiamo $\sup_{m\leq m_0} P_m\{\bar{X}>d\} \underset{\text{teor}}{=} P_{m_0}\{\bar{X}>d\} = \alpha$

$$\alpha = P_{m_0}\{\bar{X}>d\} = P_{m_0}\left\{\frac{\sqrt{n}}{\sigma}(\bar{X}-m_0) > \frac{\sqrt{n}}{\sigma} d\right\} = P\left\{Z>\frac{\sqrt{n}}{\sigma}d\right\} \text{ con } Z\sim\mathcal{N}(0,1)$$

$$\Leftrightarrow \frac{\sqrt{n}}{\sigma} d = q_{1-\alpha}$$

Quindi $C=\left\{\frac{\sqrt{n}}{\sigma}(\bar{X}-m_0)>q_{1-\alpha}\right\}$

Il p-value di $(x_1 .. x_n)$ è $\bar{\alpha} = P\left\{Z>\frac{\sqrt{n}}{\sigma}(\bar{x}-m_0)\right\}$ con $Z\sim\mathcal{N}(0,1)$

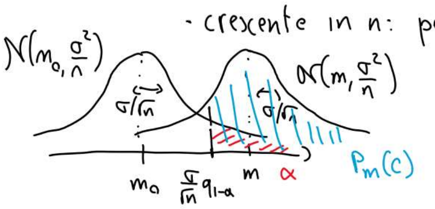


Oss: La potenza del test in $m>m_0$ è

$$P_m(C) = P_m\left\{\frac{\sqrt{n}}{\sigma}(\bar{X}-m)>q_{1-\alpha}-\frac{\sqrt{n}}{\sigma}(m-m_0)\right\} = P\left\{\overset{\sim\mathcal{N}(0,1)}{Z}>q_{1-\alpha}-\frac{\sqrt{n}}{\sigma}(m-m_0)\right\}$$

$$= 1-\Phi\left(q_{1-\alpha}-\frac{\sqrt{n}}{\sigma}(m-m_0)\right)$$

$P_m(C)$ 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

- $(S,\mathcal{S},(Q_\theta)_\Theta) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{B}(p)_{p\in[0,1]})$, $\theta=p\in[0,1]$

$X_1,.. X_n$ campione i.i.d., assumiamo n grande

Consideriamo il test $H_0: p\leq p_0$ 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)

con $Z \sim \mathcal{N}(0,1)$

Il p-value di $(x_1 .. x_n)$ è (approssimativamente) $\bar{\alpha}(x_1 .. x_n) = P\left\{ Z > \sqrt{\frac{n}{p_0(1-p_0)}} (\bar{x} - p_0) \right\}$

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 $\alpha = 0.05$

  $(S, \mathcal{F}, (Q_\theta)_\theta) = (\mathbb{N}, \mathcal{B}(\mathbb{N}), \mathcal{B}(p)_{p \in [0,1]})$. p = prob pezzo difettoso

  $X_i = \begin{cases} 1 & \text{se i-simo pezzo difettoso} \\ 0 & \text{altrimenti} \end{cases}$, $i = 1, .. \underset{\text{grande}}{n = 1000}$ campione i.i.d. di $\mathcal{B}(p)$

  $\mathcal{H}_0 : p \leq 0.1 (= p_0)$ $\quad \mathcal{H}_1 : p > 0.1$

  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 $\bar{x} = \frac{120}{1000} = 0.12$, $\quad 63.25 \cdot (\bar{x} - 0.1) = 1.265 \leq 1.96$: $(x_1 .. x_n) \notin C$, 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 $\alpha > \bar{\alpha}$

Foglio di esercizi 1

Discussione soluzioni: 07.03.2023

1. Tiriamo un dado non truccato due volte. Descrivi uno spazio degli esiti Ω e una misura di probabilità P per modellare il risultato di questo esperimento. Sia A l'evento {il secondo lancio più grande del primo}. Calcolare la probabilità $P(A)$.

2. In un gioco il giocatore ed il banco lanciano entrambi per 10 volte una moneta equilibrata. Il giocatore vince solo se il numero di teste da lui ottenuto è maggiore strettamente del numero di teste ottenuto dal banco. Qual è la probabilità che il giocatore vinca?

3. Le n cifre di un numero sono scelte in maniera casuale. Calcolare la probabilità che
   (a) non appaia il 3;
   (b) non appaiono né il 4 né il 7;
   (c) appaia almeno un 5.

   Scrivere poi un'espressione per la probabilità che nel numero il 3 appaia prima del 4.

4. Si estraggono due numeri a e b da una scatola contenente n palline numerate da 1 a n. Calcolare la probabilità che $|a - b| = 1$ nell'ipotesi che le estrazioni vengano effettuate:
   (a) senza reinserimento;
   (b) con reinserimento.

5. Due amici si trovano in coda ad uno sportello della loro banca, insieme ad altre n persone. Assumendo di non avere informazioni sul momento del loro arrivo (cioè assumendo che ogni configurazione delle persone in coda è ugualmente probabile), calcolare la probabilità che siano separati
   (a) da esattamente k persone;
   (b) da almeno 3 persone.

6. Si scrivono su 11 foglietti di carta le lettere della parola **ABRACADABRA**, una per foglietto, e le si pongono in un contenitore. Si estraggono poi, a caso, i foglietti. Qual è la probabilità che le lettere, nell'ordine estratto, diano di nuovo la stessa parola?

7. Quattro giocatori sono ad un tavolo da poker. Determinare la probabilità che, una volta distribuite le carte (ognuno dei quattro giocatori riceve 5 carte),
   (a) il primo giocatore riceva esattamente un asso;
   (b) ogni giocatore abbia esattamente un asso.

8. Siano A, B, C, D quattro eventi tali che $P(A) = 1/2$, $P(A \cap B \cap D) = 1/4$ e $P(A \cap B \cap C \cap D) = 1/9$.
   (a) Dimostrare che le ipotesi fatte non sono in contraddizione con gli assiomi di Kolmogorov.
   (b) Calcolare, se possibile, $P(A \cap [(B \cap D)^c \cup C])$.

9. (Paradosso dei compleanni). Consideriamo una classe di n persone e a ognuna di esse associamo un numero da 1 a 365 (per semplicità non si considerano gli anni bisestili) che corrisponde al numero di giorni tra il primo gennaio ed il giorno del rispettivo compleanno. Qual è la probabilità che almeno due persone abbiano il compleanno lo stesso giorno? (Dare una formula per il risultato, e tempo permettendo valutarla con l'aiuto di un computer per $n \in \{2, \dots, 100\}$)

10. Il problema di Monty Hall è un famoso problema di matematica basato su un quiz televisivo condotto da Monty Hall. Tu (il concorrente) hai davanti tre porte chiuse. C'è un premio (10^4 EUR) dietro una porta, e una capra dietro ognuna delle altre due. Ti viene chiesto di scegliere una porta tra le tre, ma non puoi ancora vedere cosa c'è dietro. Monty, che sa cosa c'è dietro ogni porta, apre una delle altre due porte per rivelare che c'è dietro una capra. Quindi ti offre una possibilità per cambiare la tua scelta. È una buona idea cambiare (assumendo di avere già tante capre a casa, quindi di volere il premio)?

    Questa domanda genera spesso molta confusione (provate con amici/famiglia!). Un un'allettante risposta intuitiva potrebbe essere: Dopo che Monty ti ha mostrato una porta senza premio, è ugualmente probabile che il premio si trovi dietro una delle altre due porte. Quindi, cambiare non fa differenza. È corretto?

    Per fare chiarezza, usiamo il modello più semplice possibile: assumiamo che tu abbia deciso se cambiare o no *prima* dell'inizio del gioco.

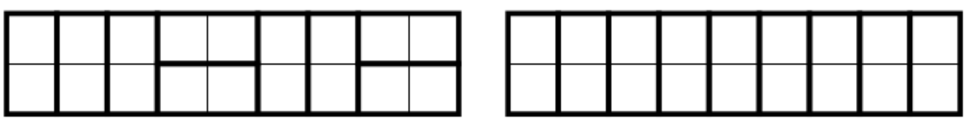
    (a) Lascia che il risultato di l'esperimento sia la porta dietro la quale si nasconde il premio. Trova lo spazio di probabilità che descrive l'esperimento (la tua decisione di cambiare o meno non fa parte del modello).

    Ora, usiamo il modello per calcolare la probabilità di vincita separatamente per i due scenari (quello dove cambi e quello dove non cambi). Per ragioni di simmetria, possiamo sistemare l'etichettatura delle porte in modo che la tua scelta iniziale sia la porta numero 1.

    (b) Supponi che la tua strategia sia di non cambiare. Qual è la probabilità di vincere il premio?[1]

    (c) Supponi che la tua strategia sia di cambiare dopo che Monty mostra una porta senza premio. Quale è la probabilità di vincere il premio?

11. Piastrelliamo una scacchiera di dimensione $2\times(2n+1)$ con $2n+1$ piastrelle di taglia 2×1 in modo che ogni casella sia coperta da esattamente una piastrella. Le piastrelle possono essere disposte orizzontalmente o verticalmente. Assumiamo di scegliere una configurazione a caso in modo che ogni configurazione distinta abbia la stessa probabilità di essere scelta. La figura qui sotto mostra due configurazioni possibili per $2n+1=9$.

    (a) Calcolare la probabilità della configurazione con tutte le piastrelle disposte verticalmente

    (b) Trovare la probabilità che ci sia una piastrella verticale al centro della scacchiera (nella posizione $n+1$)

12. (Opzionale) Un sacchetto contiene 90 gettoni numerati da 1 a 90. Tre giocatori, detti A, B, C, estraggono (senza reinserimento) 2 gettoni a testa, ed ognuno dei giocatori sceglie il gettone con il valore più grande tra i suoi gettoni. Si stila poi una classifica, in base al valore dei gettoni dal più grande al più piccolo, dei tre giocatori.

    (a) Dire quante sono le estrazioni possibili (ovvero gli insiemi di coppie di gettoni in possesso dei tre giocatori prima della selezione del massimo). Dire se le estrazioni sono equiprobabili, giustificando la propria risposta.

    (b) Dire quanti sono i risultati possibili (ovvero le terne di gettoni dei tre giocatori dopo la selezione del massimo). Dire se i risultati sono equiprobabili, giustificando la propria risposta.

    (c) Calcolare le probabilità che A si classifichi primo e che A si classifichi secondo.

    (d) Calcolare le probabilità del quesito precedente nel caso in cui inizialmente A estragga 2 gettoni, B ne estragga 3 e C ne estragga 4.

[1] Considera, per ciascuna porta i, se vinci o meno quando il premio è dietro la porta i.

1. Tiriamo un dado non truccato due volte. Descrivi uno spazio degli esiti Ω e una misura di probabilità P per modellare il risultato di questo esperimento. Sia A l'evento {il secondo lancio più grande del primo}. Calcolare la probabilità $P(A)$.

Prove ripetute:

- lancio primo dado $\Omega_1 = \{1,2,3,4,5,6\}$
- lancio secondo dado $\Omega_2 = \{1,2,3,4,5,6\}$

$$\Longrightarrow \Omega = \{(\omega_1,\omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$$
$$= \{(1,1),(1,2),(1,3)\dots(2,1)\dots(6,6)\}$$

$\mathbb{P}$ è la misura uniforme su Ω:

$$\mathbb{P}(\omega) = \frac{1}{|\Omega|} = \frac{1}{36} \qquad \forall \omega \in \Omega$$

$$\mathbb{P}(\underbrace{\{1^\circ \text{ lancio} > 2^\circ \text{ lancio}\}}_{A}) = \sum_{\omega \in A} \mathbb{P}(\omega) = \frac{1}{|\Omega|} \cdot |A| = \frac{6\cdot 5}{2} \cdot \frac{1}{36} = \frac{5}{12}$$

2. In un gioco il giocatore ed il banco lanciano entrambi per 10 volte una moneta equilibrata. Il giocatore vince solo se il numero di teste da lui ottenuto è maggiore strettamente del numero di teste ottenuto dal banco. Qual è la probabilità che il giocatore vinca?

$$\Omega_i = \{0,1\}, \quad \Omega = \{\underbrace{\omega_1,\dots,\omega_{10}}_{\text{giocatore}}, \underbrace{\omega_{11}\dots\omega_{20}}_{\text{banco}} : \omega_i \in \Omega_i\}$$

$$\mathbb{P}(\omega) = \frac{1}{2^{20}}$$

$$A_+ = \{\text{vittoria giocatore}\} = \left\{\sum_{i=1}^{10} \omega_i > \sum_{j=11}^{20} \omega_j\right\}$$

$$A_= = \left\{\sum_{i=1}^{10} \omega_i = \sum_{j=11}^{20} \omega_j\right\} \qquad A_- = \Omega \setminus (A_+ \cup A_=) = \left\{\sum_{j=11}^{20} \omega_j > \sum_{i=1}^{10} \omega_i\right\}$$

Notiamo che Ω può essere diviso in 3 pezzi disgiunti

$\Omega = A_+ \cup A_= \cup A_-$ e che $|A_+| = |A_-|$

quindi $\mathbb{P}(A_+) + \mathbb{P}(A_-) + \mathbb{P}(A_=) = 2\mathbb{P}(A_+) + \mathbb{P}(A_=) = 1$

$$\Longrightarrow \mathbb{P}(A_+) = \frac{1-\mathbb{P}(A_=)}{2} = \frac{1-\sum_{k=0}^{10}\binom{10}{k}^2 \cdot \frac{1}{2^{20}}}{2}$$

3. Le n cifre di un numero sono scelte in maniera casuale. Calcolare la probabilità che

   (a) non appaia il 3;

   (b) non appaiono né il 4 né il 7;

   (c) appaia almeno un 5.

   Scrivere poi un'espressione per la probabilità che nel numero il 3 appaia prima del 4.

$$\Omega_i = \{0,1,2,\dots,9\} \longrightarrow \Omega = \{(\omega_1\dots\omega_n) : \omega_i \in \Omega_i\}.$$

$$\mathbb{P}(\omega) = \frac{1}{|\Omega|} = \frac{1}{10^n} \qquad \forall \omega \in \Omega$$

a) $\mathbb{P}(\underbrace{\{\text{non appare il } 3\}}_{A_3}) = \sum_{\omega\in A} \mathbb{P}(\omega) = \frac{1}{|\Omega|}\cdot|A| = \frac{9^n}{10^n}$

b) $\mathbb{P}(\underbrace{\{\text{non appare } 4,7\}}_{A_{4,7}}) = \sum_{\omega\in B} \mathbb{P}(\omega) = \frac{1}{|\Omega|}|B| = \frac{8^n}{10^n}$

c) $\mathbb{P}(\underbrace{\{\text{appare almeno } 5\}}_{A_5^c = \{\text{non appare } 5\}^c}) = 1-\mathbb{P}(A_5) = 1-\frac{9^n}{10^n}$

$$\mathbb{P}(\{3 \text{ appare prima del } 4\}) = \mathbb{P}\left(\bigcup_{j=1}^{n}\{3 \text{ prima } 4, \text{ primo } 3 \text{ in posizione } j\}\right)$$

$$= \sum_{j=1}^{n} \mathbb{P}(\{\text{primo } 3 \text{ in posizione } j, \text{ nessun } 4 \text{ prima di } j\})$$

$$= \sum_{j=1}^{n} \mathbb{P}(\{3 \text{ in posizione } j, \text{ nessun } 3 \text{ o } 4 \text{ prima di } j\})$$

$$= \sum_{j=1}^{n} \frac{8^{j-1}\cdot 1\cdot 10^{n-j}}{10^n} = \frac{1}{10}\sum_{j=1}^{n}\left(\frac{8}{10}\right)^{j-1} = \frac{1}{10}\cdot\frac{1-(8/10)^n}{1-8/10}$$

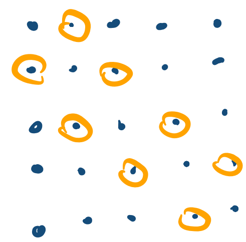
4. Si estraggono due numeri a e b da una scatola contenente n palline numerate da 1 a n. Calcolare la probabilità che $|a - b| = 1$ nell'ipotesi che le estrazioni vengano effettuate:

(a) senza reinserimento;

(b) con reinserimento.

b) $\Omega = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$ $\qquad \Omega_1 = \Omega_2 = \{1, \dots, n\}$.

$$\mathbb{P}(\omega) = \frac{1}{|\Omega|} = \frac{1}{n^2} \quad \forall \omega \in \Omega.$$



$$\longrightarrow \mathbb{P}(\{|\omega_1 - \omega_2| = 1\}) = \frac{2 \cdot (n-1)}{n^2}$$

a) $\Omega' = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2, \omega_1 \neq \omega_2\}$.

$$\mathbb{P}(\omega) = \frac{1}{|\Omega'|} = \frac{1}{\frac{n \cdot (n-1)}{2} \cdot 2}.$$

$$\mathbb{P}(\{|\omega_1 - \omega_2| = 1\}) = \frac{2 \cdot (n-1)}{n \cdot (n-1)} = \frac{2}{n}$$

6. Si scrivono su 11 foglietti di carta le lettere della parola **ABRACADABRA**, una per foglietto, e le si pongono in un contenitore. Si estraggono poi, a caso, i foglietti. Qual è la probabilità che le lettere, nell'ordine estratto, diano di nuovo la stessa parola?

$$\Omega = \{ \omega_1 \dots \omega_{11} : \omega_i \in \Omega_i , \ \omega_i \neq \omega_j \ i \neq j \}$$

$$\Omega_i = \{ A_1, A_2, A_3, A_4, A_5, B_1, B_2, C_1, D_1, R_1, R_2 \}$$

$$\mathbb{P}(\omega) = \frac{1}{|\Omega|} = \frac{1}{11!}$$

$$\mathbb{P}(\{ABRACADABRA\}) = \frac{5 \cdot 2 \cdot 2 \cdot 4 \cdot 1 \cdot 3 \cdot 1 \cdot 2 \cdot 1 \cdot 1 \cdot 1}{11!} = \frac{5!2!2!}{11!}$$

7. Quattro giocatori sono ad un tavolo da poker. Determinare la probabilità che, una volta distribuite le carte (ognuno dei quattro giocatori riceve 5 carte),

   (a) il primo giocatore riceva esattamente un asso;

   (b) ogni giocatore abbia esattamente un asso.

$$\Omega = \{ (\omega_1, \dots, \omega_{20}) : \omega_i \in \Omega_i , \ \omega_i \neq \omega_j \ \forall i, j \}$$

$$\Omega_i = \{ A♡, A♢, A♣, A♠, K♡, \dots \} \qquad |\Omega_i| = 52 .$$

a) $$\mathbb{P}(A_1) = \frac{5 \cdot 4 \cdot \cancel{48} \cdot 47 \cdot 46 \cdot 45}{52 \cdot 51 \cdot 50 \cdot 49 \cdot \cancel{48}} =$$

$A_1 = \{$esatt. un asso per $G1\}$

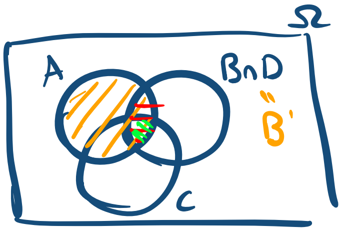
b) $$\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4) = \frac{4^4 \, 4! \, 48!/32!}{52!/32!} = \frac{5^4 \cdot 4!}{52 \cdot 51 \cdot 50 \cdot 49}$$

8. Siano A, B, C, D quattro eventi tali che $P(A) = 1/2$, $P(A \cap B \cap D) = 1/4$ e $P(A \cap B \cap C \cap D) = 1/9$.

(a) Dimostrare che le ipotesi fatte non sono in contraddizione con gli assiomi di Kolmogorov.

(b) Calcolare, se possibile, $P(A \cap [(B \cap D)^c \cup C])$.

a)



$P(A), P(A \cap B \cap D); P(A \cap B \cap C \cap D) \in [0,1]$

Consideriamo la σ-algebra generata da

$\{A, A \cap B', A \cap B' \cap D\}$

Questa σ-algebra è la stessa che è generata dalla partizione

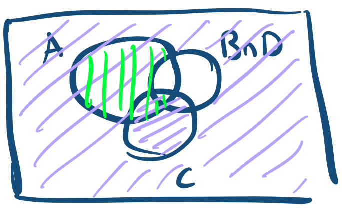
$\{A^c, A \cap (A \cap B')^c, (A \cap B') \cap (A \cap B' \cap C)^c, A \cap B' \cap C\}$

Basta calcolare che $\mathbb{P}$ di ognuno degli elementi sopra è $\in [0,1]$

$\mathbb{P}(A^c) = 1/2$, $\mathbb{P}(A \cap (A \cap B')^c) = \frac{1}{2} - \frac{1}{4}$ $\quad \mathbb{P}(A \cap B' \cap C) = \frac{1}{9}$

$\mathbb{P}((A \cap B') \cap (A \cap B' \cap C)^c) = \frac{1}{4} - \frac{1}{9}$

b)



$$\mathbb{P}(A \cap (B'^c \cup C)) = \mathbb{P}(A \cap (A \cap B')^c) + \mathbb{P}(A \cap B' \cap C)$$

$$= \frac{1}{4} + \frac{1}{9}$$

9. (Paradosso dei compleanni). Consideriamo una classe di n persone e a ognuna di esse associamo un numero da 1 a 365 (per semplicità non si considerano gli anni bisestili) che corrisponde al numero di giorni tra il primo gennaio ed il giorno del rispettivo compleanno. Qual è la probabilità che almeno due persone abbiano il compleanno lo stesso giorno? (Dare una formula per il risultato, e tempo permettendo valutarla con l'aiuto di un computer per $n \in \{2, \dots, 100\}$)

$$\Omega = \{(\omega_1, \omega_2, \dots, \omega_n) : \omega_i \in \Omega_i\} \qquad \Omega_i = \{1, 2, \dots 365\}$$

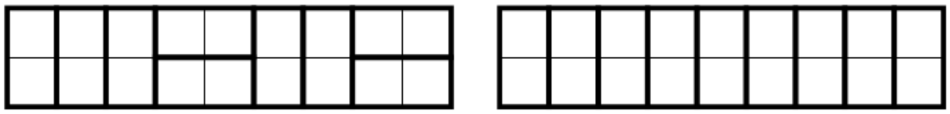
$$\mathbb{P}(\omega) = \frac{1}{|\Omega|} = \frac{1}{(365)^n}$$

$A_n = \{$due persone hanno compleanno stesso giorno$\}$

$$\mathbb{P}(A_n) = 1 - \mathbb{P}(A_n^c) = 1 - \frac{365!}{(365-n)!} \cdot \frac{1}{365^n}$$

$$\mathbb{P}(A_n^c) = \sum_{\omega \in A_n^c} \mathbb{P}(\omega) = \frac{|A_n^c|}{|\Omega|} = \frac{365 \cdot 364 \cdot \dots (365-n+1)}{365^n}$$

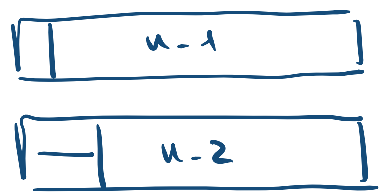
11. Piastrelliamo una scacchiera di dimensione $2\times(2n+1)$ con $2n+1$ piastrelle di taglia 2×1 in modo che ogni casella sia coperta da esattamente una piastrella. Le piastrelle possono essere disposte orizzontalmente o verticalmente. Assumiamo di scegliere una configurazione a caso in modo che ogni configurazione distinta abbia la stessa probabilità di essere scelta. La figura qui sotto mostra due configurazioni possibili per $2n+1=9$.



(a) Calcolare la probabilità della configurazione con tutte le piastrelle disposte verticalmente

(b) Trovare la probabilità che ci sia una piastrella verticale al centro della scacchiera (nella posizione $n+1$)

sia a_n il numero di configurazioni una scacchiera di lunghezza n.



Calcoliamo :
$$\begin{cases} a_1 = 1 \\ a_2 = 2 \\ a_3 = a_2 + a_1 \\ \vdots \\ a_n = a_{n-1} + a_{n-2} \end{cases}$$

a) $\mathbb{P}(\text{tutte verticali}) = \dfrac{1}{|\Omega_n|} = \dfrac{1}{a_n}$

b) $\mathbb{P}(\text{centrale verticale}) = \dfrac{|B_n|}{|\Omega_n|} = \dfrac{a^2_{\frac{n-1}{2}}}{a_n}$

$$|B_n| = a^2_{\frac{n-1}{2}}$$

10. Il problema di Monty Hall è un famoso problema di matematica basato su un quiz televisivo condotto da Monty Hall. Tu (il concorrente) hai davanti tre porte chiuse. C'è un premio (10^4 EUR) dietro una porta, e una capra dietro ognuna delle altre due. Ti viene chiesto di scegliere una porta tra le tre, ma non puoi ancora vedere cosa c'è dietro. Monty, che sa cosa c'è dietro ogni porta, apre una delle altre due porte per rivelare che c'è dietro una capra. Quindi ti offre una possibilità per cambiare la tua scelta. È una buona idea cambiare (assumendo di avere già tante capre a casa, quindi di volere il premio)?

    Questa domanda genera spesso molta confusione (provate con amici/famiglia!). Un un'allettante risposta intuitiva potrebbe essere: Dopo che Monty ti ha mostrato una porta senza premio, è ugualmente probabile che il premio si trovi dietro una delle altre due porte. Quindi, cambiare non fa differenza. È corretto?

    Per fare chiarezza, usiamo il modello più semplice possibile: assumiamo che tu abbia deciso se cambiare o no *prima* dell'inizio del gioco.

    (a) Lascia che il risultato di l'esperimento sia la porta dietro la quale si nasconde il premio. Trova lo spazio di probabilità che descrive l'esperimento (la tua decisione di cambiare o meno non fa parte del modello).

    Ora, usiamo il modello per calcolare la probabilità di vincita separatamente per i due scenari (quello dove cambi e quello dove non cambi). Per ragioni di simmetria, possiamo sistemare l'etichettatura delle porte in modo che la tua scelta iniziale sia la porta numero 1.

    (b) Supponi che la tua strategia sia di non cambiare. Qual è la probabilità di vincere il premio?[1]

    (c) Supponi che la tua strategia sia di cambiare dopo che Monty mostra una porta senza premio. Quale è la probabilità di vincere il premio?

a) $\Omega = \{1, 2, 3\}$ $\quad \mathbb{P}(\omega) = \frac{1}{|\Omega|} = \frac{1}{3}$

b) La strategia è di non cambiare, quindi

$$\mathbb{P}(\text{vincere}) = \mathbb{P}(1) = \frac{1}{3}$$

c) La strategia è di non cambiare

$$P(\text{vincere}) = 1 - P(1) = \frac{2}{3}$$

Richiamo di combinatoria

Obiettivo: calcolare "quanti modi ci sono per (...)"
aka (also known as) cardinalità di insiemi

0) Principio moltiplicativo:

Calcolare il # di modi di fare una scelta in N passi con:

- n_1 modi di fare il primo passo
- n_2 modi di fare il secondo
- ...
- n_N modi di fare l'N-esimo.

allora # modi totali $= n_1 \cdot n_2 \cdot n_3 \cdots n_N$

$$\left|\left\{(x_1, x_2, \dots x_n) : x_1 \in A_1,\ x_2 \in A_{2,x_1},\ x_3 \in A_{3x_1x_2}, \dots\ x_N \in A_{Nx_1\dots x_{N-1}}\right\}\right|$$

$$= n_1 \cdot n_2 \cdots n_N \quad \text{se} \quad \begin{cases} |A_1| = n_1 \\ |A_{2x_1}| = n_2 & \forall x_1 \in A_1 \\ \vdots \\ |A_{Nx_1\dots x_{N-1}}| = n_N & \forall x_1 \in A_1, x_2 \in A_{2x_1}, \dots \end{cases}$$

1) Permutazioni di n elementi.

"modi di ordinare n elementi $\{a_1, a_2, \dots a_n\}$"

$$= \left|\left\{(x_1, \dots x_n) : x_1, x_2, \dots, x_n \in \{a_1, a_2, \dots, a_n\},\ x_i \neq x_j\ \forall i \neq j\right\}\right|$$

$$= \left|\left\{(a_{\sigma(1)} \dots a_{\sigma(n)}) : \sigma \in S_n\right\}\right|$$

$$= n \cdot (n-1)(n-2) \cdots 2 \cdot 1 = n!$$

Dim: Costruiamo ogni n-tuple sequenzialmente in n passi.

- primo elemento in $A_1 = \{a_1 \dots a_n\} \rightarrow n_1 = n$
- secondo in $A_{2x_1} = \{a_1, \dots a_n\} \setminus \{x_1\} \rightarrow n_2 = n-1$
- ...
- ultimo in $A_{n x_1 \dots x_{n-1}} = \{a_1 \dots a_n\} \setminus \{x_1 \dots x_{n-1}\} \rightarrow n_n = 1$

$\Longrightarrow$ # permutazioni $\{a_1 \dots a_n\} = n \cdot (n-1) \dots 1 = n!$

2) Disposizioni (semplici) di k oggetti su n $(k \leq n)$.

"modi di disporre k elementi scelti in un insieme $\{a_1, \dots, a_n\}$ di n elementi tenendo conto dell'ordine"

$$= |\{(x_1, \dots, x_k) : x_1, \dots x_k \in \{a_1, \dots a_n\} \quad x_i \neq x_j \ \forall i \neq j\}|$$

$$= n \cdot (n-1) \dots (n-k+1) = \frac{n!}{(n-k)!}$$

Dim: come per permutazioni, ma fermandosi dopo k elementi.

Es: In una gara con 10 atleti, i possibili podi sono

$$\text{\# disposizioni di 3 elementi su 10} = \frac{10!}{(10-3)!} = 10 \cdot 9 \cdot 8$$

3) Disposizioni con ripetizione di k elementi scelti tra n

"modi di disporre k elementi possibilmente ripetuti scelti in un insieme $\{a_1, \dots, a_n\}$ di n el. tenendo conto dell' ordine"

$$= |\{(x_1, \dots x_k) : x_1, \dots x_k \in \{a_1, \dots, a_n\}\}| = |\{a_1, \dots, a_n\}^k|$$

$$= n^k$$

<u>Dim</u>: Costruiamo ogni k-tupla in k passi

- $x_1 \in \{a_1, \dots a_n\} \longrightarrow n_1 = n$
- $x_2 \in \{a_1, \dots, a_n\} \longrightarrow n_2 = n$
- ...

$\Longrightarrow n_1 \cdot n_2 \dots n_k = n^k$

<u>Es</u>: Lancio un dado 3 volte di fila

\# disposizioni con ripetizione di 3 oggetti tra 6 $= 6^3$

<u>Es</u>: Estrazione di 5 biglie con reinserimento da urna con 20:

\# disposizioni con ripet. di 5 elementi su 20 $= 20^5$

4) Combinazioni (semplici) di k elementi scelti tra n $(k \leq n)$

"modi di scegliere k elementi distinti in un insieme $\{a_1, \dots a_n\}$ di n elementi senza tenere conto dell' ordine"

$$= \{A \subseteq \{a_1, \dots, a_n\} : |A| = k\}$$

$$= \frac{n!}{k!(n-k)!} =: \binom{n}{k}$$

Dim: Combinazioni sono disposizioni "senza ordine".

Ad ogni combinazione corrispondono $k!$ disposizioni:

$$\left|\{(x_{\sigma(1)},\dots,x_{\sigma(k)}) \mid \sigma\in S_k\}\right| = k!$$

$$\implies \#\text{combinazioni} = \frac{\#\text{disposizioni}}{k!} = \frac{n!}{(n-k)!}\cdot\frac{1}{k!} =: \binom{n}{k}$$

Es: Possibili estrazioni del lotto (no ordine, 5 numeri su 90):

$$\#\text{combinazioni di 5 el. su } 90 = \binom{90}{5}$$

5) Combinazioni di n elementi divisi in m gruppi.

"modi di dividere un insieme $\{a_1,\dots,a_n\}$ di n elementi in m gruppi (numerati/ordinati tra loro) di cui

- il 1° gruppo con k_1 elementi
- il 2° gruppo con k_2 elementi
- ...
- l' m° gruppo con k_m elementi"

$$\underline{k_1+\dots+k_m = n}$$

$$= \frac{n!}{k_1!\,k_2!\dots k_m!}$$

Dim: $$\#\text{raggruppamenti} = \frac{\#\text{permutazioni di } n \text{ elementi}}{\#\text{permutazioni che danno lo stesso raggruppamento}}$$

$$= \frac{n!}{k_1!\,k_2!\,\dots\,k_m!}$$

per ogni permutazione ci sono $k_1!\cdot k_2!\dots k_m!$ perm. "equivalenti":

$$\underbrace{\{x_1 x_2 \dots x_{k_1}\}}_{k_1!}\ \cdot\ \underbrace{\{x_{k_1+1} \dots x_{k_1+k_2}\}}_{k_2!} \dots x_n$$

<u>Es</u>: Lancio un dado 5 volte. In quanti modi posso fare 2 volte 3, 2 volte 5 e 1 volta sei?

$= \#$ raggruppamenti di $\{1,2,3,4,5\}$ in $\begin{cases} |G_3| = 2 \\ |G_5| = 2 \\ |G_6| = 1 \end{cases}$

$$= \frac{5!}{2!\,2!\,1!}$$

<u>Nota</u>: Alternativamente, si può pensare a

$\#$ raggruppamenti di $\{1,2,3,4,5\}$ =

$= (\#$ combinazioni di posizioni "date" a $G_1) \dots$
$(\#$ combinazioni di posizioni "date" a $G_m)$

$$= \binom{n}{k_1}\cdot\binom{n-k_1}{k_2}\dots\binom{n-\sum_i^{m-1} k_i}{\ }$$

$$= \frac{n!}{k_1!\,\cancel{(n-k_1)!}}\ \frac{\cancel{(n-k_1)}!}{k_2!\,(n\cancel{-k_1-k_2})!}\ \dots\ \frac{(n-\sum k_i)!}{0!\,k_m!}$$

$$= \frac{n!}{k_1!\,k_2!\dots k_m!}$$

6) # di estrazioni (non ordinate, senza reinserimento)

- da N elementi, di cui N_1 di tipo 1, N_2 di tipo 2, .. N_m di tipo m
- di n elementi, di cui $n_1 \leq N_1$ di tipo 1, $n_2 \leq N_2$ di tipo 2, ... $n_m \leq N_m$ di tipo m

$$= \binom{N_1}{n_1}\binom{N_2}{n_2}\cdots\binom{N_i}{n_i}\cdots\binom{N_m}{n_m}$$

scelte di n_i elementi di tipo i.

Foglio di esercizi 2

Discussione soluzioni: 14.03.2022

1. Un giocatore lancia due dadi. Se il risultato del lancio del primo dado è 3, qual è la probabilità che la somma dei risultati sia almeno 6? Rispondere a questa domanda facendo uso esplicito della definizione di probabilità condizionata.

2. Due contenitori contengono rispettivamente il primo, 5 palline rosse e 7 nere, il secondo 8 palline rosse e 3 nere. Si sceglie a caso un contenitore e da esso si estraggono due palline, che risultano essere entrambe rosse. Qual è la probabilità che le palline siano state estratte dal primo contenitore?

3. Due dadi vengono tirati. Consideriamo i tre seguenti eventi:

$$\begin{aligned} A &= \text{il primo dado dà un numero dispari,} \\ B &= \text{il secondo dado dà un numero pari,} \\ C &= \text{la somma dei due risultati è pari.} \end{aligned}$$

   (a) Dire se i tre eventi A, B, C sono indipendenti.
   (b) Dire se sono a due a due indipendenti.

4. Sappiamo che il 4% della popolazione è affetto da una certa malattia. Abbiamo a disposizione un test con le seguenti caratteristiche: se la persona è malata, il test è positivo con probabilità pari a 0.95, se la persona è sana, il test è positivo con probabilità pari a 0.15.

   (a) Qual è la probabilità che una persona sia malata se è risultata positiva al test?
   (b) Qual è la probabilità che una persona sia sana se è risultata negativa al test?

5. Una coppia ha due figli.

   (a) Se almeno uno dei due è maschio, qual è la probabilità che entrambi i figli siano maschi?
   (b) Se il secondogenito è maschio, qual è la probabilità che entrambi i figli siano maschi?

6. Siano A, B, C tre eventi. Consideriamo le due affermazioni

   H_1 : "A, B sono indipendenti" e
   H_2 : "A, B sono indipendenti condizionatamente a C" (ovvero $P(A \cap B|C) = P(A|C)P(B|C)$).

   (a) Vale l'implicazione $H_1 \Rightarrow H_2$?
   (b) Vale l'implicazione $H_2 \Rightarrow H_1$?
   (c) Sotto quali ipotesi su C vale $H_1 \Leftrightarrow H_2$?

7. Mostrare che, se A, B e C sono eventi indipendenti, allora $A \cap B$ e C sono eventi indipendenti. Mostrare che il viceversa non vale. Mostrare lo stesso per $A \cup B$ e C.

8. Un mago dice di possedere una moneta magica che alterna perfettamente tra lanci risultanti in testa e croce (se la volta precedente ha dato testa, la prossima volta darà croce e vice versa con probabilità 1). Uno scettico, prima di vedere l'esperimento, pensa che ci sia solo l'1% di probabilità che la moneta abbia questa proprietà, e chiede al mago di convincerlo del contrario. Quanti lanci alternati dovranno essere osservati dallo scettico perche (secondo lui) la probabilità che la moneta abbia la proprietà professata dal mago sia maggiore del 99%?

9. Ci sono n contenitori, numerati da 1 a n. Il contenitore k-esimo contiene k palline rosse e $n-k$ palline nere. Si sceglie a caso un contenitore e da questo si estrae una pallina. Qual è la probabilità che la pallina sia rossa? Si eseguono due estrazioni, ognuna con la medesima modalità usata in precedenza. Qual è la probabilità che entrambe le palline siano rosse, se dopo la prima estrazione la pallina estratta viene rimessa nel contenitore da cui era stata estratta? Qual è la probabilità se invece la pallina non viene rimessa?

10. (Monty Hall 2.0) Risolvere il problema 10 del foglio di esercizi 1 usando la formula di Bayes: Assumiamo che tu abbia scelto la porta 1 e che Monty abbia aperto la porta 3 rivelando una capra (i numeri delle porte scelta e aperta sono scelti senza perdita di generalità). Calcolare la probabilità

$$P(\text{premio è dietro porta 1}|\text{porta 3 è aperta e contiene capra})$$

assumendo che

(a) Monty sappia dietro che porta c'è il premio. Se il premio è dietro la porta 1 sceglie a caso che porta aprire, altrimenti apre l'unica porta non scelta e che contiene la capra.

(b) Monty abbia dimenticato che porta contiene il premio (quindi apre una porta a caso)

11. Sia $s \in (1, \infty)$. La funzione zeta di Riemann è definita come segue

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s}$$

Vogliamo dimostrare che

$$\zeta(s) = \frac{1}{\prod_i (1 - p_i^{-s})}$$

dove $p_1 = 2, p_2 = 3, p_3 = 5, \dots$ sono i numeri primi (in ordine).

(a) Si consideri lo spazio di probabilità $(\mathbb{N}, 2^{\mathbb{N}}, P)$ dove

$$P(A) = \frac{1}{\zeta(s)} \sum_{n \in A} \frac{1}{n^s}$$

per ogni $A \in 2^{\mathbb{N}}$. Dimostrare che $(\mathbb{N}, 2^{\mathbb{N}}, P)$ è uno spazio di probabilità

(b) Sia p un numero primo, e $N_p := \{n \in \mathbb{N} \ : \ n \text{ è diviso da } p\}$. Calcolare $P(N_p)$

(c) Dimostrare che gli eventi $\{N_{p_i}\}_{i\geq 1}$ sono mutualmente indipendenti

(d) Calcolare $P(\cap_{i\geq 1} N_{p_i}^c)$ e dedurre il risultato desiderato

12. Ci sono n candidati per un posto di lavoro. Si ammetta che i candidati possano essere ordinati dal migliore al peggiore. L'esaminatrice incontra sequenzialmente n candidati, uno dopo l'altro, in ordine casuale. L'esaminatrice deve scegliere se accettare o rifiutare ogni candidato alla fine del colloquio corrispondente, senza possibilità di tornare indietro e cambiare la propria decisione. La strategia utilizzata (chiamata strategia k) è la seguente: si intervistano e rifiutano automaticamente k candidati e dopodiché si assume il primo candidato che è "meglio" di tutti i precedenti (inclusi i primi k). Se non c'è un tale candidato, viene assunto l'ultimo candidato. Il parametro k è scelto e fissato prima dell'inizio dei colloqui.

(a) Per n fisso, calcolare la probabilità che la strategia k porti all'assunzione del miglior candidato per $k \in \{1, \dots, n\}$,

(b) Per valori grandi di n sia k^* il valore di k che massimizza la probabilità di assumere il miglior candidato. Si esprima k^* come funzione di n. [1]

(c) Si trovi il limite k^*/n per $n \to \infty$.

[1] **Suggerimento:** Per n grande si approssimi $\frac{1}{n}\sum_{j=k+1}^{n} \frac{1}{(j-1)/n} \approx \int_{k/n}^{1} \frac{1}{x} dx$

Probabilità condizionata e indipendenza.

Def: Dato $B \in \mathcal{F}$ con $\mathbb{P}(B) > 0$ la prob. condizionata di $A \in \mathcal{F}$ dato B è

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

Prop (formula della partizione / prob. totale): Sia $\{B_i\}_{i=1}^n$ una partizione di Ω, allora $\forall A \in \mathcal{F}$:

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

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

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

Cor: $A \in \mathcal{F}$, $\{B_i\}_{i=1}^n$ partizione di Ω, allora

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\,\mathbb{P}(B_i)}{\sum_{j=1}^n \mathbb{P}(A|B_j)\,\mathbb{P}(B_j)} \qquad \forall i \in \{1, \dots n\}$$

Def: $A, B \in \mathcal{F}$ sono indipendenti $(A \perp\!\!\!\perp B)$ se $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

Oss: $A \perp\!\!\!\perp B \longrightarrow A^c \perp\!\!\!\perp B$

Def: $(A_i)_{i \in I}$ è una famiglia di eventi indipendenti se $\forall J \subseteq I$ finito

$$\mathbb{P}\Big(\bigcap_{j \in J} A_j\Big) = \prod_{j \in J} \mathbb{P}(A_j)$$

1. Un giocatore lancia due dadi. Se il risultato del lancio del primo dado è 3, qual è la probabilità che la somma dei risultati sia almeno 6? Rispondere a questa domanda facendo uso esplicito della definizione di probabilità condizionata.

$A = \{\text{primo dado} + \text{secondo dado} \geq 6\}$

$B = \{\text{primo dado} = 3\}$; $\quad C = \{\text{secondo dado} \geq 3\}$

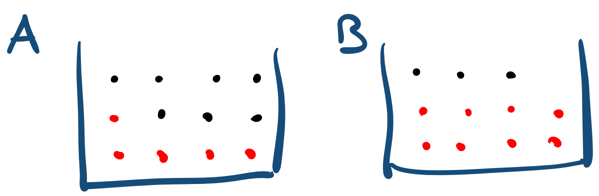
$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \cap C)}{\mathbb{P}(B)} \overset{(*)}{=} \frac{\mathbb{P}(B) \cdot \mathbb{P}(C)}{\mathbb{P}(B)}$$

$A \cap B = B \cap C$

$$= \mathbb{P}(C) = \frac{2}{3}$$

$$(*) = \mathbb{P}(B \cap C) = \frac{4}{6 \cdot 6} = \frac{1}{6} \cdot \frac{4}{6} = \mathbb{P}(B) \cdot \mathbb{P}(C)$$

2. Due contenitori contengono rispettivamente il primo, 5 palline rosse e 7 nere, il secondo 8 palline rosse e 3 nere. Si sceglie a caso un contenitore e da esso si estraggono due palline, che risultano essere entrambe rosse. Qual è la probabilità che le palline siano state estratte dal primo contenitore?



$\Omega = \{(\omega_0, \omega_1, \omega_2) : \omega_0 \in \{A, B\}\}$

$$\mathbb{P}(\{\omega_0 = A\} \mid \{(\omega_1, \omega_2) = (r, r)\}) = \frac{\mathbb{P}(\{\omega_0 = A\} \cap \{(\omega_1, \omega_2) = (r, r)\})}{\mathbb{P}(\{(\omega_1, \omega_2) = (r, r)\})}$$

$$= \frac{\mathbb{P}(\{(\omega_1, \omega_2) = (r, r)\} \mid \{\omega_0 = A\})\, \mathbb{P}(\{\omega_0 = A\})}{\mathbb{P}(\{(\omega_1, \omega_2) = (r, r)\} \mid \{\omega_0 = A\})\, \mathbb{P}(\{\omega_0 = A\}) + \mathbb{P}(\{(\omega_1, \omega_2) = (r, r)\} \mid \{\omega_0 = A\})\, \mathbb{P}(\{\omega_0 = A\})}$$

$$= \frac{\frac{5 \cdot 4}{12 \cdot 11} \cdot \frac{1}{2}}{\frac{5 \cdot 4}{12 \cdot 11} \cdot \frac{1}{2} + \frac{8 \cdot 7}{11 \cdot 10} \cdot \frac{1}{2}} = \frac{1}{1 + \frac{14}{5} \cdot \frac{6}{5}} \simeq 0.229$$

3. Due dadi vengono tirati. Consideriamo i tre seguenti eventi:

$$A = \text{il primo dado dà un numero dispari,}$$
$$B = \text{il secondo dado dà un numero pari,}$$
$$C = \text{la somma dei due risultati è pari.}$$

(a) Dire se i tre eventi A, B, C sono indipendenti.

(b) Dire se sono a due a due indipendenti.

a) Verifichiamo se si applica la definizione di indip.

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\emptyset) = 0$$

$$\mathbb{P}(A)\,\mathbb{P}(B) \cdot \mathbb{P}(C) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

$\Rightarrow \mathbb{P}(A \cap B \cap C) \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$

$\Rightarrow$ gli eventi non sono collettivamente indipendenti

b) $\Omega = \{(\omega_1, \omega_2) : a, b \in \{1,2,3,4,5,6\}\} = \{1,2,3,4,5,6\}^2$ $\qquad \mathbb{P}(\omega) = \frac{1}{36}$

- $\mathbb{P}(A \cap B) = \mathbb{P}(\{(\omega_1, \omega_2) : \omega_1 \text{ pari}, \omega_2 \text{ dispari}\}) = \frac{9}{36} = \frac{1}{4}$
- $\mathbb{P}(A \cap C) =$ ——— " ——— $= \frac{9}{36} = \frac{1}{4}$
- $\mathbb{P}(B \cap C) =$ ——— " ——— $= \frac{9}{36} = \frac{1}{4}$
- $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$

$\Rightarrow$
- $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \Rightarrow A \perp B$
- $\mathbb{P}(A \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(C) \Rightarrow A \perp C$
- $\mathbb{P}(B \cap C) = \mathbb{P}(B) \cdot \mathbb{P}(C) \Rightarrow B \perp C$

4. Sappiamo che il 4% della popolazione è affetto da una certa malattia. Abbiamo a disposizione un test con le seguenti caratteristiche: se la persona è malata, il test è positivo con probabilità pari a 0.95, se la persona è sana, il test è positivo con probabilità pari a 0.15.

   (a) Qual è la probabilità che una persona sia malata se è risultata positiva al test?

   (b) Qual è la probabilità che una persona sia sana se è risultata negativa al test?

$$\mathbb{P}(+|M) = 0.95 \qquad \mathbb{P}(+|S) = 0.15 \qquad \mathbb{P}(M) = 0.04$$

a) $$\mathbb{P}(M|+) = \frac{\mathbb{P}(+|M)\,\mathbb{P}(M)}{\mathbb{P}(+)} = \frac{\mathbb{P}(+|M)\,\mathbb{P}(M)}{\mathbb{P}(+|M)\,\mathbb{P}(M) + \mathbb{P}(+|S)\,\mathbb{P}(S)}$$

$$\mathbb{P}(+) = \mathbb{P}(+|M)\cdot\mathbb{P}(M) + \mathbb{P}(+|S)\,\mathbb{P}(S)$$

$$= 0.86$$

b) $$\mathbb{P}(S|-) = \frac{\mathbb{P}(-|S)\,\mathbb{P}(S)}{\mathbb{P}(-)} = \frac{\mathbb{P}(-|S)\,\mathbb{P}(S)}{\mathbb{P}(-|S)\,\mathbb{P}(S) + \mathbb{P}(-|M)\,\mathbb{P}(M)}$$

$$\mathbb{P}(-) = \mathbb{P}(-|S)\,\mathbb{P}(S) + \mathbb{P}(-|M)\,\mathbb{P}(M)$$

$$= 0.94$$

5. Una coppia ha due figli.

   (a) Se almeno uno dei due è maschio, qual è la probabilità che entrambi i figli siano maschi?

   (b) Se il secondogenito è maschio, qual è la probabilità che entrambi i figli siano maschi?

Assumiamo che il sesso dei figli sia indip.

a) $$\mathbb{P}(\{\text{entrambi maschi}\}|\{\text{almeno un maschio}\}) =$$

$$= \frac{\mathbb{P}(\{\text{entrambi maschi}\}\cap\{\text{almeno un maschio}\})}{\mathbb{P}(\{\text{almeno un maschio}\})} = \frac{1/4}{1-1/4} = \frac{1}{3}$$

b) $$\mathbb{P}(\{\text{entrambi maschi}\}|\{\text{secondo maschio}\}) =$$

$$= \frac{\mathbb{P}(\{\text{entrambi maschi}\}\cap\{\text{secondo maschio}\})}{\mathbb{P}(\{\text{secondo maschio}\})} = \frac{1/4}{1/2} = \frac{1}{2}$$

6. Siano A, B, C tre eventi. Consideriamo le due affermazioni

H_1 : "A, B sono indipendenti" e
H_2 : "A, B sono indipendenti condizionatamente a C" (ovvero $P(A \cap B|C) = P(A|C)P(B|C)$).

(a) Vale l'implicazione $H_1 \Rightarrow H_2$?
(b) Vale l'implicazione $H_2 \Rightarrow H_1$?
(c) Sotto quali ipotesi su C vale $H_1 \Leftrightarrow H_2$?

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) \overset{?}{\Longrightarrow} \mathbb{P}(A \cap B | C) = \mathbb{P}(A|C) \cdot \mathbb{P}(B|C)$$

a) no : $A = \{1,2\}$, $B = \{1,3\}$ $C = \{2,3\}$ $\Omega = \{1,2,3,4\}$

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

b) no : $\Omega = \{1,2,3,4\}$ $A = \{1,2\}$ $B = \{1,2\}$, $C = \{1\}$

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

c) $C \perp\!\!\!\perp A, B, A \cap B$

$$\Longrightarrow \mathbb{P}(A \cap B | C) = \mathbb{P}(A \cap B) \overset{H_1}{=} \mathbb{P}(A) \cdot \mathbb{P}(B) = \mathbb{P}(A|C) \cdot \mathbb{P}(B|C)$$

7. Mostrare che, se A, B e C sono eventi indipendenti, allora $A \cap B$ e C sono eventi indipendenti. Mostrare che il viceversa non vale. Mostrare lo stesso per $A \cup B$ e C.

a) "$\Rightarrow$" $\quad \mathbb{P}((A\cap B)\cap C) = \mathbb{P}(A\cap B\cap C) = \underbrace{\mathbb{P}(A)\cdot\mathbb{P}(B)}_{\mathbb{P}(A\cap B)}\cdot\mathbb{P}(C)$

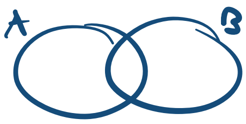
$= \mathbb{P}(A\cap B)\cdot\mathbb{P}(C)$

"$\not\Leftarrow$" Controesempio: $\Omega = \{1,2,3,4\}$, $\mathbb{P}(\omega) = \frac{1}{4}$

$A = \{2\}$, $B = \{1\}$, $C = \{2,3\}$

$\mathbb{P}(A\cap B\cap C) = \mathbb{P}(\phi) = 0 = 0\cdot\mathbb{P}(C) = \mathbb{P}(A\cap B)\,\mathbb{P}(C)$

b) "$\Rightarrow$" $\quad \mathbb{P}((A\cup B)\cap C) = \mathbb{P}(((A\cap B)\cup(A^c\cap B)\cup(A\cap B^c))\cap C)$



$= \mathbb{P}((A\cap B\cap C)\cup(A^c\cap B\cap C)\cup(A\cap B^c\cap C)) =$

$\begin{cases} p_A = \mathbb{P}(A) \\ p_B = \mathbb{P}(B) \\ p_C = \mathbb{P}(C) \end{cases}$

$= \mathbb{P}(A\cap B\cap C) + \mathbb{P}(A^c\cap B\cap C) + \mathbb{P}(A\cap B^c\cap C)$

$= p_A p_B p_C + (1-p_A)p_B p_C + p_A(1-p_B)p_C \pm p_A p_B p_C$

$= \mathbb{P}(C)\left[p_B + p_A - p_A p_B\right] = \mathbb{P}(C)\,\mathbb{P}(A\cup B)$

Alternativa: $\mathbb{P}((A\cup B)\cap C) = \mathbb{P}((A\cap C)\cup(A\cap B)) = \mathbb{P}(A\cap C) + \mathbb{P}(A\cap B) - \mathbb{P}(A\cap B\cap C) =$

"$\not\Leftarrow$" Controesempio: $\Omega = \{1,2,3,4\}$ $\quad \mathbb{P}(\omega) = \frac{1}{4}$

$A = \{3\}$ $\quad B = \{2\}$ $\quad C = \{3,4\}$

$\mathbb{P}((A\cup B)\cap C) = \mathbb{P}(\{3\}) = \frac{1}{4} = \frac{1}{2}\cdot\frac{1}{2} = \mathbb{P}(A\cup B)\cdot\mathbb{P}(C)$

$\mathbb{P}(A\cap B\cap C) = 0 \neq \mathbb{P}(A)\cdot\mathbb{P}(B)\cdot\mathbb{P}(C)$

8. Un mago dice di possedere una moneta magica che alterna perfettamente tra lanci risultanti in testa e croce (se la volta precedente ha dato testa, la prossima volta darà croce e vice versa con probabilità 1). Uno scettico, prima di vedere l'esperimento, pensa che ci sia solo l'1% di probabilità che la moneta abbia questa proprietà, e chiede al mago di convincerlo del contrario. Quanti lanci alternati dovranno essere osservati dallo scettico perche (secondo lui) la probabilità che la moneta abbia la proprietà professata dal mago sia maggiore del 99%?

$A_n = \{ n \text{ lanci alternati} \}$

$B = \{ \text{moneta sia magica} \}$.

$\mathbb{P}(A_n \mid B)$

$$0.99 \le \mathbb{P}(B \mid A_n) = \frac{\mathbb{P}(A_n \mid B) \cdot \mathbb{P}(B)}{\mathbb{P}(A_n)}$$

$$= \frac{\mathbb{P}(A_n \mid B) \cdot \mathbb{P}(B)}{\mathbb{P}(A_n \mid B) \cdot \mathbb{P}(B) + \mathbb{P}(A_n \mid B^c) \cdot \mathbb{P}(B^c)} =$$

$$= \frac{1 \cdot 0.01}{1 \cdot 0.01 + \left(\frac{1}{2}\right)^{n-1} \cdot 0.99} = 0.99$$

9. Ci sono n contenitori, numerati da 1 a n. Il contenitore k-esimo contiene k palline rosse e $n-k$ palline nere. Si sceglie a caso un contenitore e da questo si estrae una pallina. Qual è la probabilità che la pallina sia rossa? Si eseguono due estrazioni, ognuna con la medesima modalità usata in precedenza. Qual è la probabilità che entrambe le palline siano rosse, se dopo la prima estrazione la pallina estratta viene rimessa nel contenitore da cui era stata estratta? Qual è la probabilità se invece la pallina non viene rimessa?

C_1 C_2 C_n

$$\mathbb{P}(\text{rossa}) = \sum_{j=1}^{n} \mathbb{P}(\text{rossa} \mid C_j) \cdot \mathbb{P}(C_j) =$$

b) 2 rosse

Richiamo: per una lista $B_1 \dots B_n$ con $B_j \cap B_h = \emptyset$ $j \neq h$

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{j=1}^{n} A \cap B_j\right) = \sum_{j=1}^{n} \mathbb{P}(A \cap B_j) = \sum_{j=1}^{n} \mathbb{P}(A \mid B_j)\,\mathbb{P}(B_j)$$

$$= \sum_{j=1}^{n} \frac{j}{n} \cdot \frac{1}{n} = \frac{1}{n^2} \cdot \sum_{j=1}^{n} j = \frac{1}{n^{\cancel{2}}} \cdot \frac{\cancel{n}(n+1)}{2}$$

$$= \frac{n+1}{2n}$$

b) $$\mathbb{P}(2\text{ rosse}) = \sum_{j=1}^{n} \mathbb{P}(2\text{ rosse} \mid C_j)\,\underbrace{\mathbb{P}(C_j)}_{1/n}$$

$$= \sum_{j=1}^{n} \frac{j}{n} \cdot \frac{j}{n} \qquad \frac{1}{n}$$

$$= \frac{1}{n^3} \sum_{j=1}^{n} j^2 = \frac{1}{n^{\cancel{3}2}} \cdot \frac{\cancel{n}(n+1)(2n+1)}{6}$$

$$= (n+1)(2n+1)/6n^2$$

10. (Monty Hall 2.0) Risolvere il problema 10 del foglio di esercizi 1 usando la formula di Bayes: Assumiamo che tu abbia scelto la porta 1 e che Monty abbia aperto la porta 3 rivelando una capra (i numeri delle porte scelta e aperta sono scelti senza perdita di generalità). Calcolare la probabilità

$$P(\text{premio è dietro porta 1}|\text{porta 3 è aperta e contiene capra})$$

assumendo che

(a) Monty sappia dietro che porta c'è il premio. Se il premio è dietro la porta 1 sceglie a caso che porta aprire, altrimenti apre l'unica porta non scelta e che contiene la capra.

(b) Monty abbia dimenticato che porta contiene il premio (quindi apre una porta a caso)

In entrambi i casi vale che.

$$\mathbb{P}(\{\text{premio } 1\} | \{\text{porta 3 aperta}\} \cap \{\text{capra } 3\}) =$$

$$= \frac{\mathbb{P}(\{\text{premio } 1\} \cap \{\text{porta 3 aperta}\} \cap \{\text{capra } 3\})}{\mathbb{P}(\{\text{porta 3 aperta}\} \cap \{\text{capra } 3\})}$$

$$= \frac{\mathbb{P}(\{\text{premio } 1\} \cap \{\text{3 aperta}\} \cap \cancel{\{\text{capra } 3\}})}{\mathbb{P}(\{\text{3 aperta}\} \cap \cancel{\{\text{capra } 3\}} \cap \{\text{premio } 1\}) + \mathbb{P}(\{\text{3 ap}\} \cap \cancel{\{\text{capra } 3\}} \cap \{\text{premio } 2\}) + 0}$$

$$= \frac{\mathbb{P}(\{\text{3 aperta}\} | \{\text{premio } 1\}) \cdot \cancel{\mathbb{P}(\{\text{premio } 1\})}}{\mathbb{P}(\{\text{3 aperta}\} | \{\text{premio } 1\}) \cancel{\mathbb{P}(\{\text{premio } 1\})} + \mathbb{P}(\{\text{3 aperta}\} | \{\text{premio } 2\}) \cancel{\mathbb{P}(\{\text{premio } 2\})}} = (*)$$

a) $\mathbb{P}(\{\text{3 aperta}\} | \{\text{premio } 2\}) = 1$

$\mathbb{P}(\{\text{3 aperta}\} | \{\text{premio } 1\}) = 1/2 \implies (*) = \dfrac{1/2}{1 + 1/2} = \dfrac{1}{3}$

b) $\mathbb{P}(\{\text{3 aperta}\} | \{\text{premio } 2\}) = 1/2$

$\mathbb{P}(\{\text{3 aperta}\} | \{\text{premio } 1\}) = 1/2 \implies (*) = \dfrac{1/2}{1/2 + 1/2} = \dfrac{1}{2}$

11. Sia $s \in (1, \infty)$. La funzione zeta di Riemann è definita come segue

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s}$$

Vogliamo dimostrare che

$$\zeta(s) = \frac{1}{\prod_i (1 - p_i^{-s})}$$

dove $p_1 = 2, p_2 = 3, p_3 = 5, \ldots$ sono i numeri primi (in ordine).

(a) Si consideri lo spazio di probabilità $(\mathbb{N}, 2^{\mathbb{N}}, P)$ dove

$$P(A) = \frac{1}{\zeta(s)} \sum_{n \in A} \frac{1}{n^s}$$

per ogni $A \in 2^{\mathbb{N}}$. Dimostrare che $(\mathbb{N}, 2^{\mathbb{N}}, P)$ è uno spazio di probabilità

(b) Sia p un numero primo, e $N_p := \{n \in \mathbb{N} \ : \ n \text{ è diviso da } p\}$. Calcolare $P(N_p)$

(c) Dimostrare che gli eventi $\{N_{p_i}\}_{i \geq 1}$ sono mutualmente indipendenti

(d) Calcolare $P(\cap_{i \geq 1} N_{p_i}^c)$ e dedurre il risultato desiderato

a) $\Omega = \mathbb{N}$, $\mathcal{F} = 2^{\mathbb{N}}$ è una σ-algebra su $\mathbb{N}$

Verifichiamo che $\mathbb{P}$ soddisfa gli assiomi di Kolmogorov:

i) $\frac{1}{n^s} \geq 0 \implies 0 \leq \frac{1}{\zeta(s)} \sum_{n \in A} \frac{1}{n^s} \leq 1 \quad \forall A \in \mathcal{F}$.

ii) $\mathbb{P}(\Omega) = \frac{1}{\zeta(s)} \sum_{n \in \mathbb{N}} \frac{1}{n^s} = 1$ per definizione

iii) $\forall (A_n)_{n \in \mathbb{N}} \subseteq \mathcal{F}$:

$$\mathbb{P}\left(\bigcup_{m \in \mathbb{N}} A_m\right) = \frac{1}{\zeta(s)} \sum_{n \in \bigcup_{m \in \mathbb{N}} A_m} n^{-s} = \frac{1}{\zeta(s)} \sum_{m=1}^{\infty} \sum_{n \in A_m} n^{-s}$$

$$= \sum_{m=1}^{\infty} \frac{1}{\zeta(s)} \sum_{n \in A_m} n^{-s} = \sum_{m=1}^{\infty} \mathbb{P}(A_m)$$

b) $\mathbb{P}(N_p) = \frac{1}{\zeta(s)} \cdot \sum_{n \in N_p} n^{-s} = \frac{1}{\zeta(s)} \sum_{k \in \mathbb{N}} (k \cdot p)^{-s} = p^{-s} \frac{1}{\zeta(s)} \sum_{k \in \mathbb{N}} k^{-s} = p^{-s}$

c) Per ogni sottoinsieme $\{p_1, \dots p_k\}$ dei numeri primi

$$\mathbb{P}\left(\bigcap_{i=1}^{k} N_{p_i}\right) = \mathbb{P}\left(N_{\prod_{i=1}^{k} p_i}\right) = \left(\prod_{i=1}^{k} p_i\right)^{-s} = \prod_{i=1}^{k} p_i^{-s}$$

$$= \prod_{i=1}^{k} \mathbb{P}(N_{p_i})$$

d) $\bigcap_{i\geq 1} N_{p_i}^c = \{1\}$

$$\Longrightarrow \frac{1}{\zeta(s)} \cdot 1^{-s} = \mathbb{P}(\{1\}) = \mathbb{P}\left(\bigcap_{i\geq 1} N_{p_i}^c\right) =$$

$$= \prod_{i=1}^{\infty} \mathbb{P}(N_{p_i}^c) = \prod_{i=1}^{\infty} (1 - p_i^{-s})$$

$$\Longrightarrow \zeta(s) = \frac{1}{\prod_{i=1}^{\infty} (1-p_i^{-s})}$$

12. Ci sono n candidati per un posto di lavoro. Si ammetta che i candidati possano essere ordinati dal migliore al peggiore. L'esaminatrice incontra sequenzialmente n candidati, uno dopo l'altro, in ordine casuale. L'esaminatrice deve scegliere se accettare o rifiutare ogni candidato alla fine del colloquio corrispondente, senza possibilità di tornare indietro e cambiare la propria decisione. La strategia utilizzata (chiamata strategia k) è la seguente: si intervistano e rifiutano automaticamente k candidati e dopodiché si assume il primo candidato che è "meglio" di tutti i precedenti (inclusi i primi k). Se non c'è un tale candidato, viene assunto l'ultimo candidato. Il parametro k è scelto e fissato prima dell'inizio dei colloqui.

   (a) Per n fisso, calcolare la probabilità che la strategia k porti all'assunzione del miglior candidato per $k \in \{1, \dots, n\}$,

   (b) Per valori grandi di n sia k^* il valore di k che massimizza la probabilità di assumere il miglior candidato. Si esprima k^* come funzione di n. [1]

   (c) Si trovi il limite k^*/n per $n \to \infty$.

---

[1]**Suggerimento:** Per n grande si approssimi $\frac{1}{n}\sum_{j=k+1}^{n} \frac{1}{(j-1)/n} \approx \int_{k/n}^{1} \frac{1}{x} dx$

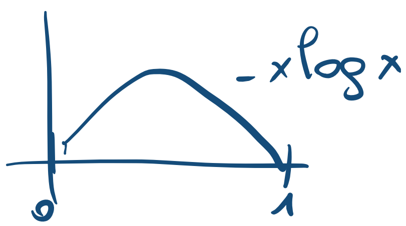
a) $\mathbb{P}(\text{miglior candidato}) =$

$$= \sum_{j=1}^{n} \mathbb{P}(j \text{ sia selezionato} \cap j \text{ sia il migliore})$$

$$= \sum_{j=k+1}^{n} \mathbb{P}(j \text{ è selezionato} \mid j \text{ è il migliore}) \underbrace{\mathbb{P}(j \text{ è il migliore})}_{1/n}$$

$$= \frac{1}{n} \sum_{j=k+1}^{n} \frac{k}{j-1} = \frac{k}{n} \sum_{j=k+1}^{n} \frac{1}{\frac{j-1}{n}} \cdot \frac{1}{n}$$

b) $$\mathbb{P}(\text{miglior candidato}) = \frac{k}{n} \frac{1}{n} \sum_{j=k+1}^{n} \frac{1}{\frac{j-1}{n}} \simeq \frac{k}{n} \int_{k/n}^{1} \frac{1}{x} dx$$



$$= -\frac{k}{n} \log \frac{k}{n} = -x \log x \Big|_{x=\frac{k}{n}}$$

$$\frac{d}{dx} x \log x = \log x + 1 = 0 \iff x = e^{-1}$$

massimo: $\frac{k}{n} = e^{-1} \implies k^* = e^{-1} \cdot n$

$$\implies \mathbb{P}(\text{miglior candidato}) = -e^{-1} \log e^{-1} = e^{-1}$$

Foglio di esercizi 3

Discussione soluzioni: 21.03.2022

1. In un contenitore ci sono 100 palline numerate da 1 a 100. Le estraiamo una dopo l'altra senza reinserimento.
   (a) Qual è la probabilità di ottenere nelle prime 10 estrazioni solo numeri ≤ 75?
   (b) Qual è la probabilità che le prime 15 palline estratte portino tutte un numero dispari?

2. Una moneta viene lanciata $2n$ volte. Sia t_n la probabilità che il numero di teste sia maggiore del numero di croci, e u_n la probabilità che il numero di teste sia pari al numero di croci.
   (a) Calcolare il limite di u_n per $n \to +\infty$.
   (b) Determinare il limite di t_n .

3. Il numero di telefonate X che arrivano ad una segreteria telefonica di un ufficio ogni 9 minuti è distribuito $X \sim \text{Pois}(6)$. Calcolare
   (a) la probabilità che arrivino almeno 5 chiamate in 9 minuti;
   (b) la probabilità che non arrivi nessuna chiamata tra le ore 9:00 e le ore 9:09

4. (distribuzione multinomiale $\mathcal{M}(n, p_1, \ldots, p_k)$). Consideriamo un esperimento di n prove indipendenti ove ogni prova può avere uno solo tra k possibili risultati, ognuno dei quali ha probabilità rispettivamente $p_1, p_2, \ldots, p_k$ (e $p_1 + \cdots + p_k = 1$) di accadere. Calcolare la probabilità $p(n_1, n_2, \ldots, n_k)$ che si abbiano n_1 esiti del tipo 1, n_2 esiti del tipo $2, \ldots, n_k$ esiti del tipo k, al variare delle k-uple $(n_1, \ldots, n_k)$ per cui $n_1 + \cdots + n_k = n$.

5. Consideriamo un esperimento a prove ripetute indipendenti con probabilità di successo p. Determinare la probabilità che
   (a) il primo successo avvenga alla prova k;
   (b) il primo successo avvenga dopo almeno k prove;
   (c) il primo successo avvenga prima della $(k+1)$-esima prova;
   (d) il primo successo avvenga in una prova dispari;
   (e) il primo successo non avvenga mai.

6. Un generatore di numeri casuali produce una successione di terne (i, j, k), con $i, j, k \in \{0, 1, 2, \ldots, 9\}$. Indichiamo con S l'evento "esce un tris" (ovvero una terna costituita da cifre tutte uguali). Calcolare la probabilità dei seguenti eventi:
   (a) tra le prime 10 terne prodotte ci sono almeno due tris
   (b) si devono produrre almeno 10 terne per ottenere due tris
   (c) si devono produrre esattamente 40 terne per avere 3 tris (ovvero il terzo tris si ha esattamente alla 40-esima terna prodotta).

7. Un collezionista ha già raccolto 60 delle 100 figurine di un album. Egli acquista una busta contenente 24 figurine (tutte diverse), tra le quali naturalmente ve ne possono essere alcune che egli già possiede. Qual'è la probabilità che tra le figurine appena comprate ve ne siano almeno 20 che già possiede?

8. Siano dati due esperimenti a prove ripetute indipendenti con probabilità di successo rispettivamente p_1 e p_2 che siano indipendenti tra loro.

(a) Qual è la probabilità che il primo successo del primo esperimento avvenga prima del primo successo del secondo?

(b) Assumiamo ora che vengano fatte 5 prove per esperimento. Calcolare la probabilità che il numero di successi nel primo gruppo sia maggiore o uguale al numero di successi nel secondo.

9. Consideriamo un'urna con N palline, di cui N_1 sono rosse, mentre $N - N_1$ sono di colore diverso dal rosso. Consideriamo l'esperimento che consiste nell'estrazione di $n \le N$ palline, in cui il "successo" è l'estrazione di una pallina rossa.

   Sia W_n la variabile aleatoria che conta il numero di palline rosse estratte tra le n. Determinare il range di W_n e la funzione di probabilità p_{W_n} associata.

10. Sia X una variabile aleatoria discreta a valori interi positivi; si dice che X gode della proprietà di perdita di memoria se

$$\mathbb{P}(X > n + k \,|\, X > k) = \mathbb{P}(X > n) \quad \text{per ogni } n, k \in \mathbb{N}^*$$

   (a) Si mostri che se $X \sim \text{Geom}'(p)$, ovvero $p_{X'}(k) = p(1-p)^{k-1}, \quad k = 1, 2, \ldots$, allora X soddisfa la proprietà di perdita di memoria.

   (b) Vale anche il viceversa, ovvero: se X è una variabile discreta a valori interi positivi che soddisfa la proprietà di perdita di memoria, allora $X \sim \text{Geom}'(p)$ per un qualche $p \in (0, 1)$. [1]

11. (Urna di Polya) Si consideri di un'urna che contiene inizialmente due palline, una rossa e una verde. Si estrare una pallina, quindi la si reimmette nell'urna, aggiungendone poi un'altra dello stesso colore di quella estratta. Si itera quindi questa procedura di estrazione/reimmissione. Sia X_n il numero di palline rosse presenti nell'urna dopo n iterazioni. Calcolare la legge di X_n per ogni $N \in \mathbb{N}$.

12. Ci sono sei tazzine da caffé con corrispondenti piattini. Due sono di colore bianco (b), due rosso (r) e due oro (o). Disponiamo i piattini in ordine sul tavolo nella sequenza *bbrroo*. Poi arriva una persona nonvedente e dispone le tazzine a caso sui piattini. Sia M il numero di tazzine il cui colore corrisponde a quello del piattino sul quale sono state disposte.

   (a) Calcolare $\mathbb{P}(M = 4)$.

   (b) (Opzionale) Calcolare la distribuzione della variabile aleatoria M.

[1] Suggerimento : se $p = prob(X = 1)$, poiché $X > 0$, allora $\mathbb{P}(X > 1) = 1 - p$

Thm: Sia $\{p_n\}$ una successione di numeri reali in $[0,1]$ tali che $\lim_{n\to\infty} n p_n = \lambda > 0$. allora per ogni $k \geq 0$

$$\lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}$$

ovvero sotto le condizioni del teorema la distribuzione binomiale converge verso quella di Poisson:

$$\lambda_n := n p_n \xrightarrow{n\to\infty} \lambda \quad \Longrightarrow \quad \mathrm{Bin}(n, p_n) \xrightarrow{n\to\infty} \mathrm{Pois}(\lambda)$$

Dim: $\displaystyle \lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} =$

$$= \lim_{n\to\infty} \frac{n\cdot(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda_n}{n}\right)^k \cdot \left(1-\frac{\lambda_n}{n}\right)^{n-k}$$

$$= \lim_{n\to\infty} \frac{n^k \cancel{+ O(n^{k-1})}}{k!} \frac{\lambda_n^k}{\cancel{n^k}} \underbrace{\left(1-\frac{\lambda_n}{n}\right)^n}_{\to e^{-\lambda}} \underbrace{\left(1-\frac{\lambda_n}{n}\right)^k}_{\to 1}$$

$$= \frac{\left(\lim_{n\to\infty} \lambda_n\right)^k}{k!} e^{-\lambda} = e^{-\lambda} \frac{\lambda^k}{k!}$$

1. In un contenitore ci sono 100 palline numerate da 1 a 100. Le estraiamo una dopo l'altra senza reinserimento.
   (a) Qual è la probabilità di ottenere nelle prime 10 estrazioni solo numeri ≤ 75?
   (b) Qual è la probabilità che le prime 15 palline estratte portino tutte un numero dispari?
2. Una moneta viene lanciata $2n$ volte. Sia t_n la probabilità che il numero di teste sia maggiore del numero di croci, e u_n la probabilità che il numero di teste sia pari al numero di croci.
   (a) Calcolare il limite di u_n per $n \to +\infty$.
   (b) Determinare il limite di t_n .
3. Il numero di telefonate X che arrivano ad una segreteria telefonica di un ufficio ogni 9 minuti è distribuito $X \sim \text{Pois}(6)$. Calcolare
   (a) la probabilità che arrivino almeno 5 chiamate in 9 minuti;
   (b) la probabilità che non arrivi nessuna chiamata tra le ore 9:00 e le ore 9:09

1) a) $\Omega = \{ S \subseteq \{1, \dots 100\},\ |S| = 10 \}$ $\qquad P(S) = \frac{1}{|\Omega|}$

$$P(\{\text{solo} \leq 75\}) = \frac{|\{\text{solo} \leq 75\}|}{|\Omega|} = \frac{75!/65!}{100!/90!} \simeq 0{,}4789$$

b) $\Omega' = \{ S \subseteq \{1, \dots 100\},\ |S| = 15 \}$ $\qquad P(S) = \frac{1}{|\Omega'|}$

$$P(\{\text{solo dispari}\}) = \frac{|\{\text{solo dispari}\}|}{|\Omega|} = \frac{50!/35!}{100!/85!} \simeq 8.89\, 10^{-6}$$

2) a) $$u_n = \binom{2n}{n}\left(\frac{1}{2}\right)^{2n} = \frac{2n!}{n!\,n!}\,\frac{1}{4^n} = \frac{\sqrt{2\pi 2n}\left(\frac{2n}{e}\right)^{2n}}{\left[\sqrt{2\pi n}\left(\frac{n}{e}\right)^{n}\right]^2} \cdot \frac{1}{4^n}$$

$$= \frac{1}{\sqrt{\pi n}}\,\frac{2^{2n}\, n^{2n} \cdot e^{-2n}}{n^{2n} \cdot e^{-2n}} \cdot \frac{1}{4^n} \sim \frac{1}{\sqrt{\pi n}} \xrightarrow{n \to \infty} 0.$$

b) $2t_n + u_n = 1 \implies t_n = \frac{1 - u_n}{2} \xrightarrow{n \to \infty} \frac{1}{2}$

3) $$P(X = k) = e^{-\mu}\,\frac{\mu^k}{k!} \qquad \mu = 6$$

a) $\implies P(X \geq 5) = 1 - P(X \leq 4) = 1 - \sum_{k=0}^{4} P(X = k) \simeq 0.7149$

b) $P(X = 0) = e^{-6}\,\frac{6^0}{0!} \simeq 0.002479$

4. (distribuzione multinomiale $\mathcal{M}(n, p_1, \dots, p_k)$). Consideriamo un esperimento di n prove indipendenti ove ogni prova può avere uno solo tra k possibili risultati, ognuno dei quali ha probabilità rispettivamente $p_1, p_2, \dots, p_k$ (e $p_1 + \cdots + p_k = 1$) di accadere. Calcolare la probabilità $p(n_1, n_2, \dots, n_k)$ che si abbiano n_1 esiti del tipo 1, n_2 esiti del tipo 2, $\dots$, n_k esiti del tipo k, al variare delle k-uple $(n_1, \dots, n_k)$ per cui $n_1 + \cdots + n_k = n$.

$$\Omega = \{1, 2, \dots, k\}^n$$

$$P((\omega_1, \omega_2, \dots, \omega_n)) = \mathbb{P}(\omega_1) \cdot \mathbb{P}(\omega_2) \cdots \mathbb{P}(\omega_n)$$

$$= p_{\omega_1} \cdot p_{\omega_2} \cdots p_{\omega_n} = \prod_{j=1}^{k} p_j^{|\{i : \omega_i = j\}|}$$

$\Longrightarrow$ per $(n_1, \dots, n_k)$ con $\sum_{j=1}^{k} n_j = n$ ogni esito

$$W_{n_1 \dots n_k} = \{\omega : |\{i : \omega_i = 1\}| = n_1, \dots, |\{i : \omega_i = k\}| = n_k\}$$

ha probabilità $\prod_{j=1}^{k} p_j^{n_j} = q(n_1 \dots n_k)$

$$p(n_1, \dots, n_k) = \mathbb{P}(W_{n_1 \dots n_k}) = \sum_{\omega \in W_{n_1 \dots n_k}} \mathbb{P}(\omega)$$

$$= |W_{n_1 \dots n_k}| \cdot q(n_1 \dots n_k)$$

$$= \frac{n!}{n_1!\, n_2! \cdots n_k!} p_1^{n_1} \cdot p_2^{n_2} \cdots p_k^{n_k}$$

$$|W_{n_1 \dots n_k}| = \binom{n}{n_1} \cdot \binom{n - n_1}{n_2} \binom{n - n_1 - n_2}{n_3} \cdots \binom{n - n_1 - \cdots - n_{k-1}}{n_k}$$

$$= \frac{n!}{n_1!\, \cancel{(n - n_1)!}} \cdot \frac{\cancel{(n - n_1)!}}{n_2!\, \cancel{(n - n_1 - n_2)!}} \cdot \frac{\cancel{(n - n_1 - n_2)!}}{n_3!\, \cancel{(n - \dots)!}}$$

$$= \frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!}$$

5. Consideriamo un esperimento a prove ripetute indipendenti con probabilità di successo p. Determinare la probabilità che
   (a) il primo successo avvenga alla prova k;
   (b) il primo successo avvenga dopo almeno k prove;
   (c) il primo successo avvenga prima della $(k+1)$-esima prova;
   (d) il primo successo avvenga in una prova dispari;
   (e) il primo successo non avvenga mai.

X = prova a cui avviene il primo successo.

a) $P(X=k) = (1-p)^{k-1} \cdot p$

b) $\boxed{\underline{\text{Nota}}:\ X \sim \text{Geom}'(p) \longrightarrow \mathbb{P}(X > n) = (1-p)^n \quad \forall n \geq 0}$

$$\mathbb{P}(X>n) = \sum_{j=n+1}^{\infty} p\,(1-p)^{j-1} = p \sum_{j'=n}^{\infty} (1-p)^{j'} = p \sum_{j''=0}^{\infty} (1-p)^n (1-p)^{j''}$$

$$= (1-p)^n \; p \sum_{j''=0}^{\infty} (1-p) = (1-p)^n \cdot p \frac{1}{1-(1-p)} = (1-p)^n$$

c) $\mathbb{P}(X < k+1) = 1 - \mathbb{P}(X > k) = 1 - (1-p)^k$

d) $$\mathbb{P}\left(\frac{X+1}{2} \in \mathbb{N}\right) = \sum_{k=0}^{\infty} P(X = 2k+1) = \sum_{k=0}^{\infty} p\,(1-p)^{2k}$$

$$= p \sum_{k=0}^{\infty} \left[(1-p)^2\right]^k = \frac{p}{1-(1-p)^2} = \frac{1}{2-p}$$

e) $\lim_{n \to \infty} P(X > n) = \lim_{n\to\infty} (1-p)^n = 0.$

6. Un generatore di numeri casuali produce una successione di terne (i,j,k), con $i,j,k \in \{0,1,2,\ldots,9\}$. Indichiamo con S l'evento "esce un tris" (ovvero una terna costituita da cifre tutte uguali). Calcolare la probabilità dei seguenti eventi:

   (a) tra le prime 10 terne prodotte ci sono almeno due tris

   (b) si devono produrre almeno 10 terne per ottenere due tris

   (c) si devono produrre esattamente 40 terne per avere 3 tris (ovvero il terzo tris si ha esattamente alla 40-esima terna prodotta).

$$p = \mathbb{P}(\text{tris}) = \sum_{j=0}^{9} P(\text{tris di } j) = 10 \cdot \frac{1}{10^3} = \frac{1}{10^2}$$

a) $Y_{10} = \#$ terne con tris su 10 terne. $\sim Bin(10, p)$

$$\longrightarrow \mathbb{P}(Y_{10} \geq 2) = 1 - \mathbb{P}(Y_{10} = 0) - \mathbb{P}(Y_{10} = 1)$$
$$= 1 - (1-10^{-2})^{10} - 10 \cdot 10^{-2}(1-10^{-2})^{9} \simeq 0{,}00426$$

b) $\{\#$ terne necessarie per il secondo successo $\geq 10\}$
$= \{\#$ successi nei primi 9 tentativi $\leq 1\} = \{X \leq 1\}$

$X = \#$ successi nei primi 9 tentativi

$$P(X \leq 1) = P(X=0) + P(X=1) = (1-p)^{10} + 10 \cdot p(1-p)^{9}$$

c) $Z = \#$ terne prodotte al terzo tris $\sim BinNeg(m, p)$

$$\mathbb{P}(Z_m = k) = p_m(k) = \binom{k-1}{m-1} p^m (1-p)^{k-m}$$ (istante m-esimo successo).

$$\Longrightarrow \mathbb{P}(Z_3 = 40) = \binom{39}{2}(10^{-2})^3(1-10^{-2})^{37} \simeq 0{,}00051$$

7. Un collezionista ha già raccolto 60 delle 100 figurine di un album. Egli acquista una busta contenente 24 figurine (tutte diverse), tra le quali naturalmente ve ne possono essere alcune che egli già possiede. Qual'è la probabilità che tra le figurine appena comprate ve ne siano almeno 20 che già possiede?

Estrazione senza reinserimento da $\Omega_0 = \{1, \ldots, 100\}$

$\Omega = \{ S \subseteq \Omega_0 \,,\, |S| = 24 \}$. $S_0 \subset \Omega_0$ figurine già raccolte, $|S_0| = 60$

$$\mathbb{P}(|S \cap S_0| \geq 20) = \sum_{j=20}^{24} \mathbb{P}(|S \cap S_0| = j)$$

$$= \sum_{j=20}^{24} \frac{\binom{60}{j}\binom{40}{24-j}}{\binom{100}{24}} \hat{=} 0.00594$$

8. Siano dati due esperimenti a prove ripetute indipendenti con probabilità di successo rispettivamente p_1 e p_2 che siano indipendenti tra loro.

(a) Qual è la probabilità che il primo successo del primo esperimento avvenga prima del primo successo del secondo?

(b) Assumiamo ora che vengano fatte 5 prove per esperimento. Calcolare la probabilità che il numero di successi nel primo gruppo sia maggiore o uguale al numero di successi nel secondo.

Sia $X_1 \sim Geom(p_1) \perp X_2 \sim Geom(p_2)$

a) $\mathbb{P}(X_1 < X_2) = \mathbb{P}(X_1 < X_2, \bigcup_{k=1}^{\infty} X_1 = k)$

$$= \mathbb{P}\left(\bigcup_{k=1}^{\infty} \{X_1 < X_2, X_1 = k\}\right) =$$

$$= \mathbb{P}\left(\bigcup_{k=1}^{\infty} \{X_2 > k, X_1 = k\}\right).$$

$$= \sum_{k=1}^{\infty} \mathbb{P}(X_2 > k, X_1 = k) = \sum_{k=1}^{\infty} \mathbb{P}(X_2 > k) \cdot \mathbb{P}(X_1 = k)$$

$$= \sum_{k=1}^{\infty} (1-p_2)^k \cdot p_1 (1-p_1)^{k-1}$$

$$= p_1(1-p_2) \sum_{k=\cancel{1}\,0}^{\infty} \left[(1-p_1)\cdot(1-p_2)\right]^{k\cancel{-1}} = \frac{p_1(1-p_2)}{1-(1-p_2)(1-p_1)}$$

b) $Y_1 \sim Bin(5, p_1) \perp Y_2 \sim Bin(5, p_2)$

$$\mathbb{P}(Y_1 \geq Y_2) = \sum_{k=0}^{5} \mathbb{P}(Y_1 \geq k, Y_2 = k) = \sum_{k=0}^{4} \mathbb{P}(Y_1 \geq k) \cdot \mathbb{P}(Y_2 = k)$$

$$= (1-(1-p_1)^5)(1-p_2)^5 +$$

$$+ (1-(1-p_1)^5 - 5(1-p_1)^4 p_1)\, 5 \cdot (1-p_2)^4 p_2$$

$$+ \left(1-(1-p_1)^5 - 5(1-p_1)^4 p_1 - \binom{5}{2}(1-p_1)^3 p_1^2\right) \binom{5}{2} (1-p_2)^3 p_2^2$$

$$+ \left(\binom{5}{4}(1-p_1) p_1^4 + p_1^5\right) \binom{5}{3} (1-p_2)^2 p_2^3$$

$$+ p_1^5 \binom{5}{4} p_2^4 (1-p_2)$$

9. Consideriamo un'urna con N palline, di cui N_1 sono rosse, mentre $N - N_1$ sono di colore diverso dal rosso. Consideriamo l'esperimento che consiste nell'estrazione di $n \leq N$ palline, in cui il "successo" è l'estrazione di una pallina rossa.

Sia W_n la variabile aleatoria che conta il numero di palline rosse estratte tra le n. Determinare il range di W_n e la funzione di probabilità p_{W_n} associata.

Notiamo che $\mathbb{P}(W_n = k) > 0$ $(k \geq 0)$ solo se ci sono almeno k palline rosse $(N_1 \geq k)$ e $n-k$ di colore diverso.

$$N - N_1 \geq n - k \iff k \geq n - (N - N_1) = N_1 - (N - n)$$

$$\implies \text{Range}(W_n) = \{k_0, k_0 + 1, \dots, N\} \qquad k_0 = \max\{0, N_1 - (N - n)\}$$

Per $k \in \text{Range}(W_n)$ otteniamo

$$p_{W_n}(k) = \mathbb{P}(W_n = k) = \frac{\binom{N_1}{k}\binom{N - N_1}{n - k}}{\binom{N}{n}}$$

10. Sia X una variabile aleatoria discreta a valori interi positivi; si dice che X gode della proprietà di perdita di memoria se

$$\mathbb{P}(X > n + k \mid X > k) = \mathbb{P}(X > n) \quad \text{per ogni } n, k \in \mathbb{N}^*$$

(a) Si mostri che se $X \sim \text{Geom}'(p)$, ovvero $p_{X'}(k) = p(1-p)^{k-1}, \quad k = 1, 2, \ldots$, allora X soddisfa la proprietà di perdita di memoria.

(b) Vale anche il viceversa, ovvero: se X è una variabile discreta a valori interi positivi che soddisfa la proprietà di perdita di memoria, allora $X \sim \text{Geom}'(p)$ per un qualche $p \in (0,1)$. [1]

<u>Nota</u>: $X \sim \text{Geom}'(p) \implies \mathbb{P}(X > n) = (1-p)^n \quad \forall n \geq 0$

$$\mathbb{P}(X > n) = \sum_{j=n+1}^{\infty} p(1-p)^{j-1} = p \sum_{j'=n}^{\infty} (1-p)^{j'} = p \sum_{j''=0}^{\infty} (1-p)^n (1-p)^{j''}$$

$$= (1-p)^n \, p \sum_{j''=0}^{\infty} (1-p) = (1-p)^n \cdot p \frac{1}{1-(1-p)} = (1-p)^n$$

a) $$\mathbb{P}(X > n+k \mid X > k) = \mathbb{P}(X > n+k, X > k) \cdot \mathbb{P}(X > k)^{-1}$$

$$= \mathbb{P}(X > n+k) \cdot \mathbb{P}(X > k)^{-1}$$

$$= (1-p)^{n+k} \cdot (1-p)^{-k} = (1-p)^n = \mathbb{P}(X > n)$$

b) Sia $f(k) = \mathbb{P}(X > k)$. $\implies$ f è decrescente

$$\mathbb{P}(X > n+k) \cdot \mathbb{P}(X > k)^{-1} = \mathbb{P}(X > n+k \mid X > k) \overset{!}{=} \mathbb{P}(X > n)$$

$$\implies f(n+k) = f(n) \cdot f(k)$$

In particolare $f(n+1) = f(n) \cdot f(1) \quad \forall n \in \mathbb{N}$, $f(0) \overset{A1}{=} 1$

$\implies f(n) = f(0) \cdot f(1)^n = 1 \cdot q^n$ per un $q \in (0,1)$

$$\implies \mathbb{P}(X_n = k) = \mathbb{P}(\{X_n > k-1\} \cap \{X_n > k\}^c) = \mathbb{P}(X_n > k-1) - \mathbb{P}(X_n > k)$$

$$= q^{k-1} - q^k = (1-q)\, q^{k-1} = p(1-p)^{k-1}$$

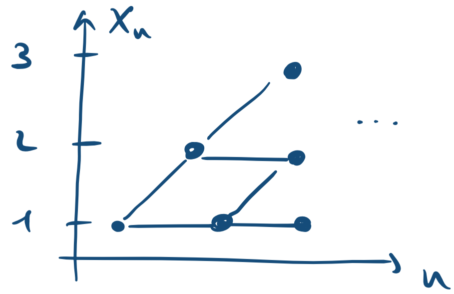
scrivo $q = 1-p$ per un $p \in (0,1)$

11. (Urna di Polya) Si consideri di un'urna che contiene inizialmente due palline, una rossa e una verde. Si estrare una pallina, quindi la si reimmette nell'urna, aggiungendone poi un'altra dello stesso colore di quella estratta. Si itera quindi questa procedura di estrazione/reimmissione. Sia X_n il numero di palline rosse presenti nell'urna dopo n iterazioni. Calcolare la legge di X_n per ogni $n \in \mathbb{N}$.

Vediamo che $\quad X_n \in \{0,1,2,\dots,n\}$

- $P(X_2=1)=1 \qquad \mathbb{P}(X_2=0)=\mathbb{P}(X_2=2)=0$
- $\mathbb{P}(X_3=1)=\mathbb{P}(X_3=2)=\frac{1}{2}, \quad \mathbb{P}(X_3=0)=\mathbb{P}(X_3=3)=0$

Meglio: $P(X_3=1)=\mathbb{P}(X_3=1 \mid X_2=1)\mathbb{P}(X_2=1)$
$$=\frac{1}{2}\cdot 1=\frac{1}{2}$$



- $P(X_4=1)=\mathbb{P}(X_4=1 \mid X_3=1)\mathbb{P}(X_3=1)=\frac{1}{2}\cdot\frac{2}{3}=\frac{1}{3}$

$\mathbb{P}(X_4=2)=\mathbb{P}(X_4=2 \mid X_3=1)\mathbb{P}(X_3=1)+\mathbb{P}(X_4=2 \mid X_3=2)\mathbb{P}(X_3=2)$
$$=\frac{1}{2}\cdot\frac{1}{3}+\frac{1}{2}\cdot\frac{1}{3}=\frac{1}{3}$$

$\mathbb{P}(X_4=3)=1-\frac{1}{3}-\frac{1}{3}=\frac{1}{3}$

$\Longrightarrow$ Claim: $\mathbb{P}(X_n=k)=\frac{1}{n-1} \qquad n\geq 2$

Proof: Induction ($n=2$ verificato)

$k \neq n-1, 0$: $\mathbb{P}(X_{n+1}=k+1)=$
$$=\mathbb{P}(X_{n+1}=k+1 \mid X_n=k)\mathbb{P}(X_n=k)+\mathbb{P}(X_{n+1}=k+1 \mid X_n=k+1)\mathbb{P}(X_n=k+1)$$
$$=\frac{k}{n}\cdot\frac{1}{n-1}+\frac{n-(k+1)}{n}\cdot\frac{1}{n-1}=\frac{n-1}{n}\cdot\frac{1}{n-1}=\frac{1}{n}$$

$k=n-1$: $\mathbb{P}(X_{n+1}=n)=\mathbb{P}(X_{n+1}=n \mid X_n=n-1)\mathbb{P}(X_n=n-1)=\frac{n-1}{n}\cdot\frac{1}{n-1}=\frac{1}{n}$

12. Ci sono sei tazzine da caffé con corrispondenti piattini. Due sono di colore bianco (b), due rosso (r) e due oro (o). Disponiamo i piattini in ordine sul tavolo nella sequenza *bbrroo*. Poi arriva una persona nonvedente e dispone le tazzine a caso sui piattini. Sia M il numero di tazzine il cui colore corrisponde a quello del piattino sul quale sono state disposte.

    (a) Calcolare $\mathbb{P}(M = 4)$.

    (b) (Opzionale) Calcolare la distribuzione della variabile aleatoria M.

$\Omega = \{$permutazioni distinguibili di $(bbrroo)\}$

$$|\Omega| = \frac{6!}{(2!)^3} = 90$$

a) rappresentiamo con (abc) un risultato dove a tazzine bianche, b rosse e c oro sono su un piattino del colore giusto

Abbiamo $\{M=4\} = \{(2,1,1), (1,2,1), (1,1,2)\}$.

(configurazione $(2,2,0)$ è impossibile)

$$\Longrightarrow \mathbb{P}(M=4) = 3 \cdot \frac{|\{\text{conf. } (2,1,1)\}|}{90} =$$

$$= 3 \frac{|\{bbroro,\ bbroor,\ bboror,\ bborro\}|}{90} = \frac{12}{90}$$

Foglio di esercizi 4

Discussione soluzioni: 28.03.2023

1. Sia X una v.a. che prende valori
$$X = \begin{cases} -1, & \text{con probabilità } 1/4; \\ 1, & \text{con probabilità } 1/8; \\ 3, & \text{con probabilità } 5/8. \end{cases}$$
   (a) Calcolare la densità discreta di X.
   (b) Calcolare la funzione di ripartizione di X.
   (c) Calcolare i valori attesi $E(X)$, $E(X^4+1)$.

2. Se X e Y sono due variabili aleatorie che rappresentano i risultati del lancio di due dadi, determinare distribuzione, media, moda e mediana di $X-2Y$. Dire infine se $X-2Y$ è indipendente da $X+Y$.

3. Si lanci un dado con risultato $N \in \{1,2,3,4,5,6\}$ e si lanci poi una moneta N volte. Sia X il numero di teste ottenute. Calcolare la densità discreta di X e il valore atteso $\mathbb{E}(X)$.

4. (a) Mostrare che, se X e Y sono due variabili aleatorie e X è costante, allora X e Y sono indipendenti.
   (b) Siano X, Y variabili aleatorie tali che X^2 e Y^2 sono indipendenti. Possiamo concludere che anche X e Y sono indipendenti?

5. Se le variabili aleatorie X_1, X_2, X_3, X_4 rappresentano i lanci successivi di un dado, calcolare
$$E[(X_1+X_2)(X_3+X_4)]$$

6. In un contenitore ci sono sei palline numerate da 1 a 6. Ne vengono estratte due senza reimmissione. Determinare legge congiunta, le leggi marginali e i corrispondenti valori di media e varianza delle seguenti variabili aleatorie:
   (a) $X =$ "massimo tra i due valori ottenuti",
   (b) $Y =$ "minimo tra i due valori ottenuti".

7. Sia X una variabile aleatoria a valori reali con densità discreta
$$\mathbb{P}\left(X = \frac{1}{n}\right) = \frac{1}{2^n} \qquad (n \geq 1)$$
   (a) Determinare la probabilità dei seguenti eventi:
      i. $X \geq 1$;
      ii. $X \geq \frac{1}{10}$;
      iii. $X \leq 0$;
      iv. $0 \leq X < \frac{1}{2}$.
   (b) Trovare $\mathbb{E}(X)$

   Ripetere il punto (b) per la variabile aleatoria Y con densità discreta
$$\mathbb{P}\left(X = \frac{1}{n}\right) = \frac{1}{2^{n+1}} \qquad (n \geq 1)$$
   e
$$\mathbb{P}(X = n) = \frac{1}{2^n} \qquad (n \geq 2)$$

8. (a) Sia X una variabile aleatoria a valori in $\mathbb{N}$. Dimostrare l'uguaglianza

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X > k)\,.$$

   (b) Ci sono b palline blu e r palline rosse in un contenitore. Le palline vengono estratte a caso una dietro l'altra fino a quando esce una pallina blu. Mostrare che il valore atteso del numero di estrazioni effettuate è $\frac{b+r+1}{b+1}$.

   (c) Generalizziamo il punto (a): sia Y una variabile aleatoria discreta a valori reali positivi con $\mathrm{range}(X) = \{x_1, x_2, \dots\}$ (potenzialmente infinito contabile), e assumiamo che $x_1 < x_2 < \dots$. Assumiamo che $\mathbb{E}(X)$ esista, dimostrare l'uguaglianza

$$\mathbb{E}(X) = \sum_{j=0}^{\infty} (x_{j+1} - x_j)\mathbb{P}(X > x_j)$$

   dove $x_0 := 0$.

9. Sia p un numero primo, si ponga $\Omega = \{1, 2, \dots, p\}$ e si consideri Ω come spazio di probabilità uniforme. Siano X e Y due variabili aleatorie indipendenti definite su Ω. Dimostrare che almeno una delle due variabile aleatoria X, Y è costante. Questo risultato è vero se p non è primo? Fornire opportuni controesempi.

10. (Opzionale) Sia $e(p) = \mathbb{E}(f(X_1, \dots, X_n))$, con $X_1, X_2, \dots, X_n$ variabili aleatorie indipendenti e con distribuzione di Bernoulli di parametro p, e $f : \{0,1\}^n \to \mathbb{R}$ una funzione crescente nel senso seguente: vale $f(x) \le f(y)$ se $x = (x_1, \dots, x_n)$ e $y = (y_1, \dots, y_n)$ soddisfano $x_k \le y_k$ per ogni $k = 1, \dots, n$. Dimostrare che $e(p_1) \le e(p_2)$ per $p_1 \le p_2$.

Claim: $X \sim \mathrm{Pois}(\mu)$, $Y \sim \mathrm{Pois}(\lambda)$, $X \perp Y$

$$\Longrightarrow Z = X + Y \sim \mathrm{Pois}(\lambda + \mu)$$

Proof:

$$\mathbb{P}(X+Y=k) = \sum_{l=0}^{k} \mathbb{P}(X+Y=k, Y=l)$$

$$= \sum_{l=0}^{k} \mathbb{P}(X=k-l, Y=l) = \sum_{l=0}^{k} \mathbb{P}(X=k-l)\mathbb{P}(Y=l)$$

$$= \sum_{l=0}^{k} e^{-\mu} \frac{\mu^{k-l}}{(k-l)!} \cdot e^{-\lambda} \frac{\lambda^{l}}{l!} \frac{k!}{k!}$$

$$= e^{-(\mu+\lambda)} \frac{1}{k!} \underbrace{\sum_{l=0}^{k} \frac{k!}{l!(k-l)!} \mu^{k-l} \lambda^{l}}_{(\lambda+\mu)^k}$$

1. Sia X una v.a. che prende valori

$$X = \begin{cases} -1, & \text{con probabilità } 1/4; \\ 1, & \text{con probabilità } 1/8; \\ 3, & \text{con probabilità } 5/8. \end{cases}$$

   (a) Calcolare la densità discreta di X.

   (b) Calcolare la funzione di ripartizione di X.

   (c) Calcolare i valori attesi $E(X)$, $E(X^4+1)$.

2. Se X e Y sono due variabili aleatorie che rappresentano i risultati del lancio di due dadi, determinare distribuzione, media, moda e mediana di $X-2Y$. Dire infine se $X-2Y$ è indipendente da $X+Y$.

3. Si lanci un dado con risultato $N \in \{1,2,3,4,5,6\}$ e si lanci poi una moneta N volte. Sia X il numero di teste ottenute. Calcolare la densità discreta di X e il valore atteso $\mathbb{E}(X)$.

1) a) $p_x(-1) = \frac{1}{4}$ $\quad p_x(1) = \frac{1}{8}$ $\quad p_x(3) = \frac{5}{8}$

b) $$F_x(x) = \begin{cases} 0 & x \in (-\infty, -1) \\ \frac{1}{4} & x \in [-1, 1) \\ \frac{3}{8} & x \in [1, 3) \\ 1 & x \in [3, \infty) \end{cases}$$

c) $$\mathbb{E}(X) = \sum_{x \in \text{range}(X)} x \cdot p_x(x) = -1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{8} + 3 \cdot \frac{5}{8} = \frac{14}{8} = \frac{7}{4}$$

$$\mathbb{E}(1+X^4) = 1 + \mathbb{E}(X^4) = 1 + \sum x^4 p_x(x) = 1 + (-1)^4 \frac{1}{4} + 1^4 \frac{1}{8} + 3^4 \frac{5}{8}$$

$$= 1 + 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{8} + 81 \cdot \frac{5}{8} = 1 + \frac{408}{8} = 1 + 21 = 22$$

2) $\text{range}(X-2Y) = \{-11, -10 \dots 3, 4\}$

$$\mathbb{P}(X-2Y=-11) = \mathbb{P}(X=1, Y=6) = \mathbb{P}(X=1) \cdot \mathbb{P}(Y=6) = \frac{1}{6^2}$$

$$\mathbb{P}(X-2Y=-10) = \mathbb{P}(X=2, Y=6) = \frac{1}{6^2}$$

$$\mathbb{P}(X-2Y=-9) = \mathbb{P}(X=3, Y=6) + \mathbb{P}(X=1, Y=5) = \frac{2}{6^2}$$

Distribuzione uniforme $\longrightarrow$ basta contare il # casi

z	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	
	1	1	2	2	3	3	3	3	3	3	3	3	2	2	1	1	36
$\mathbb{P}$	$/36$																

mediana

$$\mathbb{E}(Z) = \sum_{z \in range(Z)} z \cdot \mathbb{P}(Z=z) = -3.5$$

$$\text{modo}(Z) = \underset{z \in range(Z)}{\text{argmax}} \ \mathbb{P}(Z=z) \longrightarrow \text{non definita}$$

$$\text{mediana}(Z) = -3.5$$

$$\mathbb{P}(X+Y=12, Z=-11) = 0$$

$$\mathbb{P}(X+Y=12) = \frac{1}{36} \qquad \mathbb{P}(Z=-11) = \frac{1}{36}$$

3) Y = risultato del dado, X = # teste. $\begin{cases} range\ Y = \{1, \dots 6\} \\ range\ X = \{0, \dots 6\} \end{cases}$

$$\mathbb{P}(X=k) = \sum_{e=k}^{6} \mathbb{P}(X=k \mid Y=e)\, \mathbb{P}(Y=e)$$

$$= \sum_{e=k}^{6} \binom{e}{k} \cdot \left(\frac{1}{2}\right)^{k} \left(\frac{1}{2}\right)^{e-k} \cdot \frac{1}{6}$$

$$= \frac{1}{6} \sum_{e=k}^{6} \binom{e}{k} \frac{1}{2^e}$$

$$\mathbb{E}(X) = \sum_{k=0}^{6} k \cdot \mathbb{P}(X=k) = \sum_{k=0}^{6} \sum_{e=k}^{6} \frac{e!}{k!(e-k)!} k \frac{1}{6} \cdot \frac{1}{2^e}$$

4. (a) Mostrare che, se X e Y sono due variabili aleatorie e X è costante, allora X e Y sono indipendenti.
(b) Siano X, Y variabili aleatorie tali che X^2 e Y^2 sono indipendenti. Possiamo concludere che anche X e Y sono indipendenti?

a) Da dimostrare

$$\mathbb{P}(X=x, Y=y) = \mathbb{P}(X=x)\cdot\mathbb{P}(Y=y) \quad \forall x,y \in \underbrace{\text{range}(X)}_{=\{c\}} \times \text{range}(Y)$$

$$\mathbb{P}(X=c, Y=y) = \mathbb{P}(Y=y) = \mathbb{P}(X=c)\cdot\mathbb{P}(Y=y)$$

b) un controesempio range $(X) = \{-1,1\}$
range $(Y) = \{-1,1\}$.

$$P_{XY}(1,1) = P_{XY}(-1,-1) = \frac{1}{2} \qquad P_{XY}(1,-1) = P_{XY}(-1,1) = 0$$

$$P_X(1)\cdot P_Y(-1) = \frac{1}{4}$$

$$P_{X^2,Y^2}(1,1) = 1 = P_{X^2}(1)\cdot P_{Y^2}(1)$$

Richiamo: • $\mathbb{E}(aX+b) = a\mathbb{E}(X)+b$
• Se X,Y sono indip. $\Longrightarrow \mathbb{E}[X\cdot Y] = \mathbb{E}[X]\,\mathbb{E}[Y]$

5. Se le variabili aleatorie X_1, X_2, X_3, X_4 rappresentano i lanci successivi di un dado, calcolare

$$E[(X_1+X_2)(X_3+X_4)]$$

$$\mathbb{E}\left[(X_1+X_2)(X_3+X_4)\right] = \mathbb{E}(X_1X_3 + X_1X_4 + X_2X_3 + X_2X_4)$$

$$= \mathbb{E}(X_1X_3) + \mathbb{E}(X_1X_4) + \mathbb{E}(X_2X_3) + \mathbb{E}(X_2X_4)$$

$$= \mathbb{E}(X_1)\mathbb{E}(X_3) + \mathbb{E}(X_1)\mathbb{E}(X_4) + \mathbb{E}(X_2)\mathbb{E}(X_3) + \mathbb{E}(X_2)\mathbb{E}(X_4)$$

$$= 4\cdot(3{,}5)^2 = 49$$

6. In un contenitore ci sono sei palline numerate da 1 a 6. Ne vengono estratte due senza reimmissione. Determinare legge congiunta, le leggi marginali e i corrispondenti valori di media e varianza delle seguenti variabili aleatorie:

(a) $X =$ "massimo tra i due valori ottenuti",

(b) $Y =$ "minimo tra i due valori ottenuti".

a) $X =$ massimo tra i due valori

$$\mathbb{P}(X \leq k) = F_X(k) = \frac{k \cdot (k-1)}{6 \cdot 5}$$

$$\mathbb{P}(X = k) = \mathbb{P}(X \leq k) - \mathbb{P}(X \leq k-1)$$

$$= \frac{k(k-1)}{6 \cdot 5} - \frac{(k-1)(k-2)}{6 \cdot 5}$$

$$= \frac{(k-1)}{6 \cdot 5} \left(k - (k-2)\right) = \frac{2}{30} (k-1)$$

$$\mathbb{E}(X) = \sum_{k=2}^{6} k \cdot \mathbb{P}(X = k) = \frac{1}{15} \sum_{k=2}^{6} k \cdot (k-1).$$

b) Y minimo tra due valori

$$\mathbb{P}(Y \leq k) = 1 - \mathbb{P}(Y > k) = 1 - \frac{(6-k)(6-k-1)}{6 \cdot 5}$$

$$\mathbb{P}(Y = k) = \mathbb{P}(Y \leq k) - \mathbb{P}(Y < k) =$$

$$= \left(\cancel{1} - \frac{(6-k)(6-k-1)}{6 \cdot 5}\right) - \left(\cancel{1} - \frac{(6-(k-1))(6-(k-1)-1)}{6 \cdot 5}\right)$$

$$= \frac{(6-k+1)(6-k) - (6-k)(6-k-1)}{6 \cdot 5} = \frac{(6-k)}{15}$$

7. Sia X una variabile aleatoria a valori reali con densità discreta

$$\mathbb{P}\left(X = \frac{1}{n}\right) = \frac{1}{2^n} \qquad (n \geq 1)$$

(a) Determinare la probabilità dei seguenti eventi:

i. $X \geq 1$;

ii. $X \geq \frac{1}{10}$;

iii. $X \leq 0$;

iv. $0 \leq X < \frac{1}{2}$.

(b) Trovare $\mathbb{E}(X)$

Ripetere il punto (b) per la variabile aleatoria Y con densità discreta

$$\mathbb{P}\left(X = \frac{1}{n}\right) = \frac{1}{2^{n+1}} \qquad (n \geq 1)$$

e

$$\mathbb{P}(X = n) = \frac{1}{2^n} \qquad (n \geq 2)$$

a) i) $\mathbb{P}(X \geq 1) = \mathbb{P}(X = 1) = \frac{1}{2}$

ii) $\mathbb{P}\left(X \geq \frac{1}{10}\right) = \sum_{n=1}^{10} \mathbb{P}\left(X = \frac{1}{n}\right) = \sum_{n=1}^{10} \frac{1}{2^n} = \sum_{n=0}^{10} \frac{1}{2^n} - 1$

$$= \frac{1 - \left(\frac{1}{2}\right)^{10+1}}{1 - \frac{1}{2}} - 1 = 2\left(1 - \frac{1}{2^{11}}\right) - 1$$

iii) $\mathbb{P}(X \leq 0) = 0$

iv) $\mathbb{P}\left(X \in \left[0, \frac{1}{2}\right)\right) = 1 - \mathbb{P}\left(X \geq \frac{1}{2}\right) = 1 - \mathbb{P}(X = 1) - \mathbb{P}\left(X = \frac{1}{2}\right)$

$$= 1 - \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$

b) $\mathbb{E}(X) = \sum_{n=1}^{\infty} \frac{1}{n} \cdot \frac{1}{2^n} = \sum_{n=1}^{\infty} \int_0^{1/2} q^{n-1}\, dq = \int_0^{1/2} \sum_{n=1}^{\infty} q^{n-1}\, dq =$

$$= \int_0^{1/2} \frac{1}{1-q}\, dq = -\log\left(1 - \frac{1}{2}\right) = \log(2)$$

c) $\mathbb{E}(X) = \sum_{n=1}^{\infty} \frac{1}{n} \frac{1}{2^{n+1}} + \sum_{n=2}^{\infty} n \frac{1}{2^n} = \frac{1}{2}\log(2) + \frac{1}{2} \sum_{n=1}^{\infty} n \frac{1}{2^{n-1}} - \frac{1}{2} =$ ✗

$$= \frac{1}{2}\left(\log(2) + \frac{1}{\left(1 - \frac{1}{2}\right)^2} - \frac{1}{2}\right) = \frac{1}{2}\left(\log(2) + 4 - \frac{1}{2}\right)$$

8. (a) Sia X una variabile aleatoria a valori in $\mathbb{N}$. Dimostrare l'uguaglianza

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X > k)\,.$$

(b) Ci sono b palline blu e r palline rosse in un contenitore. Le palline vengono estratte a caso una dietro l'altra fino a quando esce una pallina blu. Mostrare che il valore atteso del numero di estrazioni effettuate è $\frac{b+r+1}{b+1}$.

(c) Generalizziamo il punto (a): sia Y una variabile aleatoria discreta a valori reali positivi con $\mathrm{range}(X) = \{x_1, x_2, \dots\}$ (potenzialmente infinito contabile), e assumiamo che $x_1 < x_2 < \dots$. Assumiamo che $\mathbb{E}(X)$ esista, dimostrare l'uguaglianza

$$\mathbb{E}(X) = \sum_{j=0}^{\infty} (x_{j+1} - x_j)\mathbb{P}(X > x_j)$$

dove $x_0 := 0$.

a) $X = \sum_{k=1}^{\infty} k \cdot \mathbb{1}_{k=X} = \sum_{k=1}^{\infty} \sum_{j=1}^{k} \mathbb{1}_{k=X} = \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \mathbb{1}_{X=k}\, \mathbb{1}_{j\le k}$

Fubini

$$= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \mathbb{1}_{X=k}\, \mathbb{1}_{k\ge j} = \sum_{j=1}^{\infty} \mathbb{1}_{X\ge j}$$

$$\mathbb{E}(X) = \mathbb{E}\Big(\sum_{j=1}^{\infty} \mathbb{1}_{X\ge j}\Big) = \mathbb{E}\Big(\sum_{e=0}^{\infty} \mathbb{1}_{X>e}\Big) =$$

$$= \sum_{e=0}^{\infty} \mathbb{E}(\mathbb{1}_{X>e}) = \sum_{e=0}^{\infty} \mathbb{P}(X>e)$$

b) $\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X>k) = \sum_{k=0} \frac{r}{b+r}\,\frac{r-1}{b+r-1} \cdots \frac{r-k+1}{b+r-k+1} =$

$$= \sum_{k=0}^{r} \frac{r!/(r-k)!}{(b+r)!/(b+r-k)!} = \frac{1}{\frac{(b+r)!}{r!\,b!}} \sum_{k=0}^{r} \frac{(b+r-k)!}{(r-k)!\,b!}$$

X

$$= \frac{1}{\binom{b+r}{r}} \sum_{k=0}^{r} \binom{b+r-k}{r-k} = \frac{1}{\binom{b+r}{r}} \sum_{j=0}^{r} \binom{b+j}{j} = \frac{1}{\binom{b+r}{r}} \binom{b+r+1}{r}$$

$$= \frac{(b+r+1)\cancel{!}\,/\,\cancel{r!}\,(b+1)\cancel{!}}{\cancel{(b+r)!}\,/\,\cancel{r!}\,\cancel{b!}} = \frac{b+r+1}{b+1}$$

c) $$\mathbb{E}(X) = \sum_{k=1}^{\infty} x_k \, \mathbb{P}(X = x_k) = \sum_{k=1}^{\infty} (x_k - x_0) \, \mathbb{P}(X = x_k)$$

$$= \sum_{k=1}^{\infty} \sum_{j=0}^{k-1} (x_{j+1} - x_j) \, \mathbb{P}(X = x_k)$$

$$= \sum_{k=1}^{\infty} \sum_{j=0}^{\infty} (x_{j+1} - x_j) \, \mathbb{P}(X = x_k) \, \mathbb{1}(j \leq k-1)$$

$$= \sum_{j=0}^{\infty} (x_{j+1} - x_j) \sum_{k=1}^{\infty} \mathbb{P}(X = x_k) \, \mathbb{1}(k \geq j+1)$$

$$= \sum_{j=0}^{\infty} (x_{j+1} - x_j) \, \mathbb{P}(X > x_j)$$

9. Sia p un numero primo, si ponga $\Omega = \{1, 2, \ldots, p\}$ e si consideri Ω come spazio di probabilità uniforme. Siano X e Y due variabili aleatorie indipendenti definite su Ω. Dimostrare che almeno una delle due variabile aleatoria X, Y è costante. Questo risultato è vero se p non è primo? Fornire opportuni controesempi.

Siccome $(X,Y)(\omega)$ $\omega \in \Omega$ e $p(\omega) = \frac{1}{p}$ $\forall \omega$

$P_{XY}(x,y) = n_{xy} \cdot \frac{1}{p}$ per ogni coppia (x,y)
$n_{xy} \in \mathbb{N}$

$\Longrightarrow P_X(x) = \sum_{y \in \text{range}(Y)} P_{XY}(x,y) = \frac{n_x}{p}$, $P_Y(y) = \frac{n_y}{p}$

Se X e Y sono indip.

$$\boxed{\frac{n_{xy}}{p} = P_{XY}(x,y) = P_X(x) \cdot P_Y(y) = \frac{n_x \cdot n_y}{p \, p}}$$

$\Longrightarrow n_{xy} = \frac{n_x \cdot n_y}{p}$ per $0 \leq n_x, n_y < p$

↯

Se p non è primo questo non vale

Esempio $p = 4$

$X = \{0, 1\}$, $Y = \{0, 1\}$ $P_{XY}(x,y) = \frac{1}{4}$

$P_{XY}(x,y) = \frac{1}{4} = P_X(x) \cdot P_Y(y)$

10. (Opzionale) Sia $e(p) = \mathbb{E}(f(X_1, \dots, X_n))$, con $X_1, X_2, \dots, X_n$ variabili aleatorie indipendenti e con distribuzione di Bernoulli di parametro p, e $f : \{0,1\}^n \to \mathbb{R}$ una funzione crescente nel senso seguente: vale $f(x) \le f(y)$ se $x = (x_1, \dots, x_n)$ e $y = (y_1, \dots, y_n)$ soddisfano $x_k \le y_k$ per ogni $k = 1, \dots, n$. Dimostrare che $e(p_1) \le e(p_2)$ per $p_1 \le p_2$.

$$\mathbb{E}_{p_1}(f(X_1,\dots,X_n)) - \mathbb{E}_{p_2}(f(X_1,\dots,X_n)) \le 0$$

$$\sum_{x\in\{0,1\}^n} f(x)\left(p_1^{\|x\|_1}(1-p_1)^{n-\|x\|_1}\right) - \sum_{x\in\{0,1\}^n} f(x)\left(p_2^{\|x\|_1}(1-p_2)^{n-\|x\|_1}\right) =$$

$$= \sum_{x\in\{0,1\}} f(x)\left(p_1^{\|x\|_1}(1-p_1)^{n-\|x\|_1} - p_2^{\|x\|_1}(1-p_2)^{n-\|x\|_1}\right)$$

$$= \sum_{d=0}^{n} \sum_{x:\|x\|_1=d} f(x)\left(p_1^{d}(1-p_1)^{n-d} - p_2^{d}(1-p_2)^{n-d}\right).$$

$$= \sum_{d=0}^{n} \left(p_1^{d}(1-p_1)^{n-d} - p_2^{d}(1-p_2)^{n-d}\right)\binom{n}{d}\left[\underbrace{\frac{1}{\binom{n}{d}}\sum_{|x|=d} f(x)}_{\text{funzione crescente in } d} - c\right]$$

$d=0 \longrightarrow$ positiva

$d=n \longrightarrow$ negativa

$\Rightarrow$ esiste d^*: $p_1^{d^*}(1-p_1)^{n-d^*} - p_2^{d^*}(1-p_2)^{n-d^*} \ge 0$

$p_1^{d^*+1}(1-p_1)^{n-d^*-1} - p_2^{d^*+1}(1-p_2)^{n-d^*-1} \le 0$

$\Rightarrow$ scegliamo c tale che

$$\forall d: \; p_1^{d}(1-p_1)^{n-d} - p_2^{d}(1-p_2)^{n-d} \overset{\ge}{\le} 0$$

$$\Rightarrow \left(\frac{1}{\binom{n}{d}}\sum_{|x|=d} f(x) - c\right) \overset{\le}{\ge} 0$$

e l'inverso $\Rightarrow$ Somma è negativa

Nota $c\sum_{j=0}^{n}\left(p_1^j(1-p_1)^{n-j}-p_2^j(1-p_2)^{n-j}\right)\binom{n}{j}.$

$$= c\left(\underbrace{\sum_{j=0}^{n}\binom{n}{j}p_1^j(1-p_1)^{n-j}}_{1}-\underbrace{\sum_{j=0}^{n}\binom{n}{j}p_2^j(1-p_2)^{n-j}}_{1}\right)$$

$$= 0$$

Foglio di esercizi 5

Discussione soluzioni: 04.04.2023

* Rappresenza gli esercizi prioritari

1. In un gioco, il concorrente gioca una somma iniziale $P > 0$. La probabilità di vittoria ad ogni turno è p e la somma giocata viene raddoppiata se vince, dimezzata se perde. Qual è la quantità attesa di denaro posseduta dal giocatore dopo l'n-esimo turno? Calcolarne il limite per $n \to \infty$.

2. * Siano X, Y e Z variabili aleatorie indipendenti tutte di legge di Poisson di parametro λ.
   (a) Qual è la legge condizionale di X dato $X + Y = n$? Si tratta di una legge nota? Quanto vale la media di questa legge condizionale?
   (b) Quanto vale la covarianza delle variabili aleatorie $X + Y$ e $X + Z$?
   (c) Quanto vale $P(X + Y = 2, X + Z = 3)$?

3. In un esperimento di Bernoulli con parametro p, siano T l'istante del primo successo e U l'istante del secondo successo.
   (a) Qual è la legge di distribuzione di U?
   (b) Dire se le variabili aleatorie T e U sono indipendenti.
   (c) Determinare la legge di distribuzione di T, sapendo che $U = n > 1$.

4. Un sacchetto contiene r gettoni rossi. Si inserisce nel sacchetto un gettone blu e poi si estrae (con reinserimento) un gettone. Si ripete questo procedimento (di inserire cioè un gettone blu e poi estrarre un gettone) fino a quando si estrae un gettone rosso. Detto N il numero di estrazioni fatte,
   (a) Mostrare che $P[N = \infty] = 0$;
   (b) Determinare la distribuzione di N ;
   (c) Calcolare media e varianza di N.

5. * Siano X e Y due variabili aleatorie reali. Definiamo la variabile aleatoria Z come X se esce testa al lancio di una moneta (indipendente da X e Y), e Y se invece esce croce. Determinare la covarianza tra X e Z in funzione del valore atteso e varianza di X e Y.

6. Sia $Z = \max\{X^2, Y^2\}$ con X, Y variabili aleatorie standardizzate[1] aventi covarianza ρ.
   (a) Si mostri che $E[Z] = 1$ nel caso particolare in cui $|\rho| = 1$.
   (b) Si mostri che in generale valgono le stime $1 \leq E[Z] \leq 1 + \sqrt{1 - \rho^2}$.
   Suggerimento: Usare che se $x, y > 0$, $\max(x, y) = \frac{1}{2}(x + y + |x - y|)$ e la disuguaglianza di Cauchy-Schwarz.

7. * Sia dato un numero $n \geq 1$. Si lancia n volte un dado regolare e siano $X_1, X_2, \ldots, X_n$ i risultati. Si dice che avviene un calo all'estrazione $k \geq 2$ se $X_k < X_{k-1}$. Sia D_n il numero di cali nei lanci effettuati.
   (a) Scrivere D_n come una somma di variabili aleatorie di Bernoulli (aka variabili indicatrici).
   (b) Calcolare il valore medio di ciascuna delle variabili di Bernoulli definite al punto precedenze e la covarianza di ogni coppia di tali variabili aleatorie.

[1] Una v.a. X si dice standardizzata se $E[X] = 0$ e $\mathrm{Var}(X) = 1$.

(c) Calcolare media e varianza di D_n usando i risultati dei punti precedenti.

(d) Calcolare la probabilità che non ci siano cali.

8. Un sacchetto contiene m gettoni, contrassegnati dai numeri da 1 a m, con $m \geq 3$. I gettoni sono estratti con la seguente modalità: ogni gettone estratto viene reinserito nel sacchetto dopo la successiva estrazione[2]. Per $n \geq 1$, sia S_n il numero di volte in cui il gettone m è estratto nelle prime n estrazioni.

   (a) Calcolare la media di S_n.

   (b) Determinare il limite in probabilità di $\frac{S_n}{n}$.

9. * Un dado viene lanciato n volte. Sia S_n il numero di volte, in n lanci, in cui si osserva che è uscito 6. Dimostrare che

   (a) $\lim_{n\to\infty} P[\frac{S_n}{n} > \frac{1}{3}] = 0$.

   (b) esiste una costante $C > 0$ tale che $P[\frac{S_n}{n} > \frac{1}{3}] \leq e^{-Cn}$.

10. Una moneta dà testa con probabilità p incognita. Per avere informazioni su p, la moneta viene tirata n volte e si osserva la percentuale di teste uscite su n lanci. Volendo approssimare p con tale percentuale,

    (a) quanto deve essere grande n affinché con probabilità maggiore di 0.99 l'errore commesso sia al più 0.1?

    (b) quanto deve essere grande n affinché con probabilità maggiore di 0.9 l'errore commesso sia al più 0.01?

11. * Sia $(X_n)_{n\geq 0}$ una successione di variabili aleatorie indipendenti di Poisson di parametro λ e poniamo $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$.

    (a) Stimare con la disuguaglianza di Chebyshev ($\eta > 0$)

    $$\mathbb{P}[|\bar{X}_n - \lambda| \geq \eta]$$

    (b) Stimare la stessa quantità facendo un'approssimazione gaussiana. [3]

    (c) Confrontare le due stime per $\lambda = 1, \eta = 10^{-2}$ e $n = 10000$.

12. Siano $\{X_i\}_{i=1}^{100}$ variabili aleatorie mutualmente indipendenti e identicamente distribuite (IID) con distribuzione geometrica di parametro $p \in (0, 1)$.

    (a) Determinare per quali valori di p si possono definire il valore atteso e la varianza di $Z_1 := e^{X_1}$.

    (b) Approssimare tramite il teorema del limite centrale $\mathbb{P}(\sum_{i=1}^{100} e^{X_1} \leq 300)$ per $p = 0.9$.

13. * Sia X una variabile aleatoria con media nulla e momento quarto finito. Dimostrare che esiste una costante $C > 0$ tale che per ogni $\epsilon > 0$ ed ogni intero positivo n

    $$P[|\bar{X}_n| \geq \epsilon] \leq \frac{C}{\epsilon^4 n^2}$$

    con $\bar{X}_n$ la media empirica di n copie indipendenti di X.

14. (Opzionale) Siano X, Y, Z variabili aleatorie discrete che assumono valori distinti con probabilità 1 e siano $a = P[X > Y]$, $b = P[Y > Z]$, $c = P[Z > X]$.

    (a) Dimostrare che $\min\{a, b, c\} \leq \frac{2}{3}$ e fornire un esempio in cui si ottiene l'uguaglianza $\min\{a, b, c\} = \frac{2}{3}$

    (b) Assumiamo che $P[X = 0] = 1$, che Y, Z siano indipendenti e tali che $P[Z = 1] = P[Y = -1] = p$ e $P[Z = -2] = P[Y = 2] = 1 - p$. Calcolare $\min\{a, b, c\}$ per $p \in [0, 1]$ fissato e calcolare $\sup_{p\in[0,1]} \min\{a, b, c\}$.

[2] Se ad esempio $m = 3$, e alla prima estrazione viene estratto, per dire, il gettone 1, allora alla seconda estrazione i gettoni nel sacchetto da cui si pesca sono 2 e 3. Se poi alla seconda estrazione viene estratto 2, i gettoni presenti nel sacchetto per la terza estrazione saranno 3 e 1 (estratto alla prima e reinserito nel sacchetto dopo la seconda estrazione).

[3] È sufficiente un argomento non rigoroso basato sul Teorema Limite Centrale.

1. In un gioco, il concorrente gioca una somma iniziale $P > 0$. La probabilità di vittoria ad ogni turno è p e la somma giocata viene raddoppiata se vince, dimezzata se perde. Qual è la quantità attesa di denaro posseduta dal giocatore dopo l'n-esimo turno? Calcolarne il limite per $n \to \infty$.

Y_n = somma al turno n $\qquad Y_0 = P$

Notiamo che possiamo scrivere

$Y_n = 2^{X_n - n}$ dove $X_n \sim Bin(n, p)$

conta # successi su n lanci

$$\Longrightarrow \mathbb{E}(Y_n) = \sum_{k=0}^{n} 2^{k-n} \binom{n}{k} p^k \cdot (1-p)^{n-k}$$

$$= 2^{-n} \sum_{k=0}^{n} \binom{n}{k} (2p)^k (1-p)^{n-k}$$

$$= 2^{-n} (2p + 1 - p)^n = \left(\frac{1+p}{2}\right)^n$$

2. * Siano X, Y e Z variabili aleatorie indipendenti tutte di legge di Poisson di parametro λ.

(a) Qual è la legge condizionale di X dato $X + Y = n$? Si tratta di una legge nota? Quanto vale la media di questa legge condizionale?

(b) Quanto vale la covarianza delle variabili aleatorie $X + Y$ e $X + Z$?

(c) Quanto vale $P(X + Y = 2, X + Z = 3)$?

a) $$\mathbb{P}(X=k \mid X+Y=n) = \frac{\mathbb{P}(X=k,\ X+Y=n)}{\mathbb{P}(X+Y=n)} =$$

$$= \frac{\mathbb{P}(X=k,\ Y=n-k)}{\mathbb{P}(X+Y=n)} = \frac{\mathbb{P}(X=k)\cdot\mathbb{P}(Y=n-k)}{\mathbb{P}(X+Y=n)}$$

$$= \frac{\cancel{e^{-\lambda}}\frac{\cancel{\lambda^k}}{k!}\cdot\cancel{e^{-\lambda}}\frac{\lambda^{n-\cancel{k}}}{(n-k)!}}{\cancel{e^{-2\lambda}}\frac{(2\lambda)^n}{n!}} = \frac{1}{2^n}\cdot\binom{n}{k}$$

b) $$\mathrm{Cov}(X+Y,\ X+Z) = \mathrm{Cov}(X, X+Z) + \mathrm{Cov}(Y, X+Z)$$

$$= \mathrm{Cov}(X,X) + \cancel{\mathrm{Cov}(X,Z)} + \cancel{\mathrm{Cov}(Y,X)} + \cancel{\mathrm{Cov}(Y,Z)}$$

$X \perp\!\!\!\perp Y$

$\mathrm{Cov}(X,Y)=0$ $= \mathrm{Var}(X) = \lambda$

c) $$\mathbb{P}(X+Y=2,\ X+Z=3) = \sum_{k=0}^{2} \mathbb{P}(X+Y=2, X+Z=3 \mid X=k)\,\mathbb{P}(X=k)$$

$$= \cdots = e^{-3\lambda}\left(\frac{1}{12}\lambda^5 + \frac{1}{2}\lambda^4 + \frac{1}{2}\lambda^3\right)$$

indipendenza X, Y, Z

4. Un sacchetto contiene r gettoni rossi. Si inserisce nel sacchetto un gettone blu e poi si estrae (con reinserimento) un gettone. Si ripete questo procedimento (di inserire cioè un gettone blu e poi estrarre un gettone) fino a quando si estrae un gettone rosso. Detto N il numero di estrazioni fatte,

(a) Mostrare che $P[N = \infty] = 0$;

(b) Determinare la distribuzione di N ;

(c) Calcolare media e varianza di N.

a) $\mathbb{P}(N=1) = \frac{1}{r+1}$

$$P(N=2) = \left(1 - \frac{1}{r+1}\right) \frac{2}{r+2}$$

$\vdots$

$$\mathbb{P}(N=n) = \prod_{i=1}^{n-1} \left(1 - \frac{i}{r+i}\right) \cdot \frac{n}{r+n} = \prod_{i=1}^{n-1} \frac{r}{r+i} \cdot \frac{n}{r+n} \qquad (*)$$

In particolare

$$\mathbb{P}(N=n) \le \left(\frac{r}{r+1}\right)^{n-1} \xrightarrow{n\to\infty} 0$$

b) vedi $(*)$

c) $$\mathbb{E}(N) = \sum_{n=1}^{\infty} n \frac{n}{r+n} \prod_{e=1}^{n-1} \frac{r}{r+e} = \sum_{n=1}^{\infty} n \frac{n}{r+n} \frac{r^{n-1}}{\frac{(r+n-1)!}{r!}}$$

$$= r! \frac{1}{r^{r+1}} \sum_{n=1}^{\infty} n^2 \frac{r^{n+r}}{(r+n)!}$$

3. In un esperimento di Bernoulli con parametro p, siano T l'istante del primo successo e U l'istante del secondo successo.

   (a) Qual è la legge di distribuzione di U?
   (b) Dire se le variabili aleatorie T e U sono indipendenti.
   (c) Determinare la legge di distribuzione di T, sapendo che $U = n > 1$.

a) $U \sim \mathrm{NegBin}(2, p)$

$$P(U = k) = \binom{k-1}{1} \cdot p^2 (1-p)^{k-2}$$

b) Non sono indipendenti:

$$\mathbb{P}(T = 2, U = 2) = 0 \neq p^2 \cdot p \cdot (1-p) = \mathbb{P}(U=2)\,\mathbb{P}(T=1)$$

c) Chiaramente se sappiamo che $U = n$ range$(T) = \{1, \dots, n-1\}$

$$\mathbb{P}(T = k \mid U = n) = \frac{\mathbb{P}(T = k, U = n)}{\mathbb{P}(U = n)} =$$

$$= \frac{(1-p)^{k-1} \cdot p \cdot (1-p)^{n-(k-1)-2}\, p}{(n-1) \cdot p^2 (1-p)^{n-2}}$$

$$= \frac{1}{n-1} \qquad \text{per } k \in \{1, \dots, n-1\}$$

$\Longrightarrow$ distribuzione uniforme.

5. * Siano X e Y due variabili aleatorie reali. Definiamo la variabile aleatoria Z come X se esce testa al lancio di una moneta (indipendente da X e Y), e Y se invece esce croce. Determinare la covarianza tra X e Z in funzione del valore atteso e varianza di X e Y.

$Z = I \cdot X + (1-I)Y$ dove $\underbrace{I \in \{0,1\}}_{\text{lancio moneta.}}$

Usando la bilinearità della covarianza

$$\mathrm{Cov}(X,Z) = \mathrm{Cov}(X, IX + (1-I)Y) =$$

$$= \mathrm{Cov}(X, IX) + \mathrm{Cov}(X, (1-I)Y)$$

$$\overset{*+**}{=} \frac{1}{2}(\mathrm{Var}(X) + \mathrm{Cov}(X,Y))$$

$$\mathrm{Cov}(X, IX) = \mathbb{E}(X^2 I) - \mathbb{E}(X)\mathbb{E}(IX) =$$

$$= (\mathbb{E}(X^2) - \mathbb{E}(X)^2)\mathbb{E}(I) = \frac{1}{2}\mathrm{Var}(X) \qquad *$$

$$\mathrm{Cov}(X, (1-I)Y) = \mathbb{E}(XY(1-I)) - \mathbb{E}(X)\mathbb{E}((1-I)Y)$$

$$= \mathbb{E}(1-I)(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y))$$

$$= \frac{1}{2}\mathrm{Cov}(X,Y) \qquad **$$

6. Sia $Z = \max\{X^2, Y^2\}$ con X, Y variabili aleatorie standardizzate[1] aventi covarianza ρ.

(a) Si mostri che $E[Z] = 1$ nel caso particolare in cui $|\rho| = 1$.

(b) Si mostri che in generale valgono le stime $1 \leq E[Z] \leq 1 + \sqrt{1-\rho^2}$.
Suggerimento: Usare che se $x, y > 0$, $\max(x, y) = \frac{1}{2}(x + y + |x - y|)$ e la disuguaglianza di Cauchy-Schwarz.

a) $$\mathrm{Cov}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{SD(X)\,SD(Y)}$$

Fatto: $\mathrm{Cov}(X,Y) \in [-1,1]$

$|\mathrm{Cov}(X,Y)| = 1 \Longrightarrow X = \alpha \cdot Y$

$$SD(X) = SD(Y) = 1$$

$$\Longrightarrow |\rho| = |\mathrm{Cov}(X,Y)| = 1 \Longrightarrow X = \alpha Y$$

$$1 \overset{!}{=} SD(Y) = SD(\alpha X) = |\alpha|\,SD(X) = |\alpha|\,1$$

$$\Longrightarrow \max(X^2, Y^2) = \max(X^2, \alpha^2 X^2) = X^2$$

$$\Longrightarrow E(Z) = E(X^2) = E(X^2) - E(X)^2 = \mathrm{Var}(X) = 1$$

b) • $\max(a,b) = \dfrac{a+b+|a-b|}{2}$ • $E(X^2) = E(Y^2) = 1$

$$E(Z) = E\left(\frac{X^2+Y^2+|X^2-Y^2|}{2}\right) = 1 + E\left(\frac{|X^2-Y^2|}{2}\right)$$

$$= 1 + \frac{1}{2} E\left[|(X-Y)(X+Y)|\right] \overset{C.S}{\leq} 1 + \frac{1}{2}\sqrt{E((X-Y)^2)}\cdot\sqrt{E((X+Y)^2)}$$

$$= 1 + \sqrt{\frac{E(X^2)+E(Y^2)-2(E(XY)-E(X)E(Y))}{2}}\,\sqrt{\frac{2+2\mathrm{Cov}(X,Y)}{2}}$$

$$\rho$$

$$= 1 + \sqrt{1-\rho}\cdot\sqrt{1+\rho} = 1 + \sqrt{1-\rho^2}$$

8. Un sacchetto contiene m gettoni, contrassegnati dai numeri da 1 a m, con $m \geq 3$. I gettoni sono estratti con la seguente modalità: ogni gettone estratto viene reinserito nel sacchetto dopo la successiva estrazione[2]. Per $n \geq 1$, sia S_n il numero di volte in cui il gettone m è estratto nelle prime n estrazioni.

   (a) Calcolare la media di S_n.

   (b) Determinare il limite in probabilità di $\frac{S_n}{n}$.

$$\mathbb{1}_j = \begin{cases} 1 & \text{se il gettone } m \text{ è estratto alla } j\text{-esima estr.} \\ 0 & \text{altrimenti} \end{cases}$$

a) $$\mathbb{E}(S_n) = \mathbb{E}\left(\sum_{j=1}^{n} \mathbb{1}_j\right) = \sum_{j=1}^{n} \mathbb{E}(\mathbb{1}_j) = \sum_{j=1}^{n} \mathbb{P}(\mathbb{1}_j = 1)$$

$$\mathbb{P}(\mathbb{1}_1 = 1) = \frac{1}{m}$$

$$\mathbb{P}(\mathbb{1}_2 = 1) = \mathbb{P}(\mathbb{1}_2 = 1, \mathbb{1}_1 = 0) = \mathbb{P}(\mathbb{1}_2 = 1 \mid \mathbb{1}_0 = 0)\mathbb{P}(\mathbb{1}_0 = 0)$$

$$= \frac{1}{m-1} \cdot \frac{m-1}{m} = \frac{1}{m}$$

$\vdots$

$$\mathbb{P}(\mathbb{1}_j = 1) = \frac{1}{m}$$

$$= n \cdot \frac{1}{m} = \frac{n}{m}$$

b) <u>Claim:</u> $\mathbb{P}\left(\left|\frac{1}{n}S_n - \frac{1}{m}\right| > \varepsilon\right) \longrightarrow 0$

<u>Proof</u>: $$\mathbb{P}\left(\left|\frac{1}{n}S_n - \frac{1}{m}\right| > \varepsilon\right) \overset{(C)}{\leq} \frac{\mathrm{Var}\left(\frac{S_n}{n}\right)}{\varepsilon^2} \leq \frac{C^2}{n\varepsilon^2} \xrightarrow{n \to \infty} 0$$

7. * Sia dato un numero $n \geq 1$. Si lancia n volte un dado regolare e siano $X_1, X_2, \ldots, X_n$ i risultati. Si dice che avviene un calo all'estrazione $k \geq 2$ se $X_k < X_{k-1}$. Sia D_n il numero di cali nei lanci effettuati.

   (a) Scrivere D_n come una somma di variabili aleatorie di Bernoulli (aka variabili indicatrici).

   (b) Calcolare il valore medio di ciascuna delle variabili di Bernoulli definite al punto precedenze e la covarianza di ogni coppia di tali variabili aleatorie.

   (c) Calcolare media e varianza di D_n usando i risultati dei punti precedenti.

   (d) Calcolare la probabilità che non ci siano cali.

a) $A_j = \{\text{calo al lancio } j\}$ $\qquad \mathbb{1}_{A_j} = \begin{cases} 1 & \text{se calo al lancio } j \\ 0 & \text{altrimenti} \end{cases}$

$D_n = \sum_{j=2}^{n} \mathbb{1}_{A_j}$

b) $\mathbb{E}(\mathbb{1}_{A_j}) \quad \mathbb{P}(A_j) = \mathbb{P}(A_2) = \frac{6 \cdot 5/2}{36} = \frac{5}{12}$

$Var(\mathbb{1}_{A_j}) = p(1-p) = \frac{5}{12} \cdot \frac{7}{12} \qquad$ dove $p = \mathbb{P}(A_2)$

<u>Claim</u>: $A_j \perp\!\!\!\perp A_k$ se $|j-k| > 1$

$$\begin{aligned}
Cov(\mathbb{1}_{A_2}, \mathbb{1}_{A_3}) &= \mathbb{E}(\mathbb{1}_{A_2} \cdot \mathbb{1}_{A_3}) - \mathbb{E}(\mathbb{1}_{A_2})\mathbb{E}(\mathbb{1}_{A_3}) \\
&= \mathbb{E}(\mathbb{1}_{A_2 \cap A_3}) - \mathbb{E}(\mathbb{1}_{A_2}) \cdot \mathbb{E}(\mathbb{1}_{A_3}) \\
&= \mathbb{P}(A_2 \cap A_3) - \mathbb{P}(A_2)\mathbb{P}(A_3) \\
&= \frac{\frac{6 \cdot 5 \cdot 4}{6}}{3} - \frac{5}{12} \cdot \frac{5}{12} \approx -0.08
\end{aligned}$$

c) $\mathbb{E}(D_n) = \sum_{j=2}^{n} \mathbb{E}(\mathbb{1}_{A_j}) = \sum_{j=2}^{n} \mathbb{P}(A_j) = (n-1)\mathbb{P}(A_2)$

$= (n-1) \cdot \frac{\frac{6 \cdot 5}{2}}{36} = (n-1)\frac{5}{12}$

$$\mathrm{Var}(D_n) = \sum_{j,k=2}^{n} \mathrm{Cov}(\mathbb{1}_{A_j}, \mathbb{1}_{A_k})$$

$$= \sum_{j=2}^{n} \underbrace{\mathrm{Cov}(\mathbb{1}_{A_j}, \mathbb{1}_{A_j})}_{\mathrm{Var}(\mathbb{1}_{A_j})} + 2\sum_{j=2}^{n-1} \mathrm{Cov}(\mathbb{1}_{A_j}, \mathbb{1}_{A_{j+1}})$$

$$= (n-1)\cdot\frac{5}{12}\cdot\frac{7}{12} - 2(n-2)\, 0.08$$

d) L'esperimento consiste in n prove indipendenti ciascuna con esito in $\Omega_0 = \{1,2,3,4,5,6\}$.

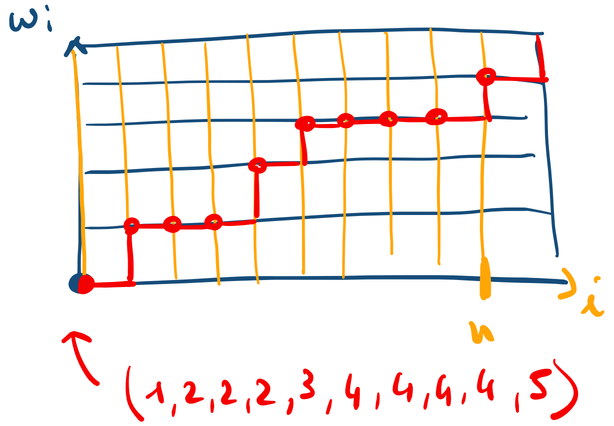
→ La distribuzione sullo spazio degli esiti è uniforme

$$\mathbb{P}(\omega) = \frac{1}{6^n} \quad , \quad \omega \in \Omega_n = \{1,2,\dots,6\}^n$$

Definiamo $A_n = \{\omega : \omega_i \le \omega_{i+1},\ i \in \{1,\dots,n-1\}\}$

$$\mathbb{P}(D_n = 0) = \sum_{\omega \in A_n} \frac{1}{6^n} = |A_n| \frac{1}{6^n} = \binom{6+n+1}{6}\cdot\frac{1}{6^n}$$

Calcoliamo $|A_n|$ rappresentando ogni esito ω come un cammino sulla griglia di larghezza $6 \times (n+1)$. Allora



$|A_n| = (\#$ cammini senza cali$)$

$= (\#$ modi di scegliere dove aument.$)$

$= (\#$scelte di 6 aumenti in $(6+n+1)$ spazi$)$

$$= \binom{6+n+1}{6}$$

9. * Un dado viene lanciato n volte. Sia S_n il numero di volte, in n lanci, in cui si osserva che è uscito 6. Dimostrare che

(a) $\lim_{n\to\infty} P[\frac{S_n}{n} > \frac{1}{3}] = 0$.

(b) esiste una costante $C > 0$ tale che $P[\frac{S_n}{n} > \frac{1}{3}] \leq e^{-Cn}$.

a) $$\mathbb{E}\left(\frac{S_n}{n}\right) = \frac{\frac{1}{6}\cdot n}{n} = \frac{1}{6} \qquad \sigma_n = SD\left(\frac{S_n}{n}\right) = \frac{\sqrt{n p(1-p)}}{n} = \frac{1}{\sqrt{n}}\cdot\frac{1}{6}\cdot\sqrt{5}$$

$$\mathbb{P}\left(\frac{S_n}{n} > \frac{1}{3}\right) = \mathbb{P}\left(\frac{S_n}{n} - \frac{1}{6} > \frac{1}{6}\right) \leq \mathbb{P}\left(\left|\frac{S_n}{n} - \frac{1}{6}\right| > \frac{1}{6}\right)$$

$$= \mathbb{P}\left(\frac{\left|\frac{S_n}{n} - \frac{1}{6}\right|}{\sigma_n} \geq \frac{1}{6\sigma_n}\right) \leq (6\sigma_n)^2 = \frac{5}{n} \quad \boxed{\xrightarrow[n\to\infty]{LLN} 0}$$

b) $$\mathbb{P}\left(\frac{S_n}{n} > \frac{1}{3}\right) = \mathbb{P}\left(\alpha S_n > \frac{\alpha n}{3}\right) = \mathbb{P}\left(e^{\alpha S_n} > e^{\alpha n/3}\right)$$

$$\overset{M}{\leq} \frac{\mathbb{E}(e^{\alpha S_n})}{e^{n/3}} = \frac{\sum_{k=1}^{n}\binom{n}{k}(1-p)^{n-k}(pe^{\alpha})^k}{e^{n/3}}$$

$$= \frac{(1-p+pe^{\alpha})^n}{e^{\alpha n/3}} = e^{(\underbrace{-\frac{\alpha}{3}}_{0.3} + \underbrace{\log(1-p+pe^{\alpha})}_{0.1})n}$$

10. Una moneta dà testa con probabilità p incognita. Per avere informazioni su p, la moneta viene tirata n volte e si osserva la percentuale di teste uscite su n lanci. Volendo approssimare p con tale percentuale,

(a) quanto deve essere grande n affinché con probabilità maggiore di 0.99 l'errore commesso sia al più 0.1?

(b) quanto deve essere grande n affinché con probabilità maggiore di 0.9 l'errore commesso sia al più 0.01?

$$I_j = \begin{cases} 1 & \text{se testa al } j\text{-esimo lancio} \\ 0 & \text{altrimenti} \end{cases} \qquad \mathbb{P}(I_j = 1) = p$$

$$\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} I_j \implies$$

$$\begin{cases} \mathbb{E}(\overline{X}_n) = \frac{1}{n}\cdot np = p \\ \operatorname{Var}(\overline{X}_n) = \operatorname{Var}\left(\frac{1}{n}\sum_{j=1}^{n} I_j\right) = \frac{1}{n^2}\cdot n \operatorname{Var}(I_j) = \frac{1}{n}\underbrace{p(1-p)}_{\sigma^2} \end{cases}$$

Chebyshev inequality

$$\mathbb{P}(|\overline{X}_n - p| \geq \varepsilon) \leq \frac{\sigma_n^2}{\varepsilon^2}$$

a) $\varepsilon = 0.1$, $\frac{\sigma_n^2}{0.01} \leq 0.01 \iff \frac{p(1-p)}{n} \leq 10^{-4}$

$\iff n \geq 10^4 p(1-p)$

b) $\varepsilon = 0.01$, $\frac{\sigma_n^2}{10^{-4}} \leq 0.1 \iff \frac{p(1-p)}{n} \leq 10^{-5}$

$\iff n \geq 10^5 p(1-p)$

11. * Sia $(X_n)_{n\geq 0}$ una successione di variabili aleatorie indipendenti di Poisson di parametro λ e poniamo $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$.

(a) Stimare con la disuguaglianza di Chebyshev ($\eta > 0$)

$$\mathbb{P}[|\bar{X}_n - \lambda| \geq \eta]$$

(b) Stimare la stessa quantità facendo un'approssimazione gaussiana. [3]

(c) Confrontare le due stime per $\lambda = 1, \eta = 10^{-2}$ e $n = 10000$.

a) $\bar{X}_n$

$$\mathbb{E}(\bar{X}_n) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\mathbb{E}(X_i)}_{\lambda} = \lambda$$

$Var(aX) = a^2 Var(X)$

$$Var(\bar{X}_n) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} Var\left(\sum_i X_i\right) = \frac{1}{n^2}\sum_i^n Var(X_i) = \frac{1}{n}\lambda$$

$$P\left(\frac{|\bar{X}_n - \lambda|}{\sqrt{Var(\bar{X}_n)}} \geq \frac{\eta}{\sqrt{Var(\bar{X}_n)}}\right) = \mathbb{P}\left(\frac{|\bar{X}_n - \lambda|}{\sqrt{\lambda/n}} \geq \underbrace{\frac{\eta}{\sqrt{\lambda/n}}}_{k}\right)$$

$$\overset{C}{\leq} \frac{1}{k^2} = \frac{\lambda}{\eta^2 n} = 1$$

c) $$\frac{\lambda}{\eta^2 n} = \frac{1}{10^{-4} \cdot 10^4} = 1$$

b) $$\mathbb{P}\left(\frac{|\bar{X}_n - \lambda|}{\sqrt{\lambda/n}} \geq \frac{\eta}{\sqrt{\lambda/n}}\right) \overset{CLT}{\simeq} 2\int_{\eta/\sqrt{\lambda/n}}^{\infty} \phi(x)\,dx = 2\left(1 - \Phi\left(\frac{\eta}{\sqrt{\lambda/n}}\right)\right)$$

$$= 2\Phi\left(-\frac{\eta}{\sqrt{\lambda/n}}\right) = 2\Phi\left(-\sqrt{\frac{\eta^2 n}{\lambda}}\right)$$

c) $$= 2\Phi(-1) = 2 \cdot 0.15 = 0.3$$

Richiamo: $X \sim Geom(p)$

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} k(1-p)^{k-1} p = p \sum_{k=0}^{\infty} \frac{d}{dq} q^k \Big|_{q=1-p} = p \frac{d}{dq} \sum_{k=0}^{\infty} q^k$$

$$= p \frac{d}{dq} \frac{1}{1-q} \Big|_{q=1-p} = p \frac{1}{(1-q)^2} \Big|_{q=1-p} = \frac{1}{p}$$

$$\mathbb{E}(X^2) = \sum_{k=0}^{\infty} k^2 (1-p)^{k-1} p = \sum_{k=0}^{\infty} k(k+1)(1-p)^{k-1} p - \sum_{k=0}^{\infty} k(1-p)^{k-1} p$$

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

$$= p\left(+\frac{1}{(1-q)^{\cancel{4}\,3}} \cdot 2\cancel{(1-q)} - \frac{1}{(1-q)^2}\right) = \frac{2}{p^2} - \frac{1}{p}$$

$$\Longrightarrow Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

12. Siano $\{X_i\}_{i=1}^{100}$ variabili aleatorie mutualmente indipendenti e identicamente distribuite (IID) con distribuzione geometrica di parametro $p \in (0, 1)$.

(a) Determinare per quali valori di p si possono definire il valore atteso e la varianza di $Z_1 := e^{X_1}$.

(b) Approssimare tramite il teorema del limite centrale $\mathbb{P}(\sum_{i=1}^{100} e^{X_i} \leq 300)$ per $p = 0.9$.

a) $$\mathbb{E}(e^{X_1}) = \sum_{k=1}^{\infty} e^k (1-p)^{k-1} \cdot p = p \cdot e \cdot \sum_{k=1}^{\infty} [e(1-p)]^{k-1} =$$

$$= p \cdot e \cdot \sum_{k=0}^{\infty} [e(1-p)]^k = \frac{p \cdot e}{1 - e(1-p)} \quad \text{se} \quad e \cdot (1-p) < 1$$

$$\Longleftrightarrow p > \frac{e-1}{e}$$

$$Var(e^{X_1}) = \mathbb{E}(e^{2X_1}) - \mathbb{E}(e^{X_1})^2$$

$$\mathbb{E}(e^{2X_1}) = \sum_{k=1}^{\infty} e^{2k} (1-p)^{1-k} p = pe^2 \sum_{k=1}^{\infty} [e^2(1-p)]^{k-1} = pe^2 \frac{1}{1-e^2(1-p)}$$

$$\text{se} \quad e^2(1-p) < 1 \Longleftrightarrow p > \frac{e^2-1}{e^2} > \frac{e-1}{e}$$

b) $p = 0.9 > \frac{e^2-1}{e^2} \Longrightarrow \mathbb{E}(e^{X_i}), \mathbb{V}\mathrm{ar}(e^{X_i})$ exist.

$$\mu = \mathbb{E}\left(\sum_{i=1}^{100} e^{X_i}\right) = 100\,\mathbb{E}\left(e^{X_1}\right) = \frac{100 \cdot p \cdot e}{1-e(1-p)} = 335,97$$

$$\mathbb{V}\mathrm{ar}\left(\sum_{i=1}^{100} e^{X_i}\right) = 100\,\mathbb{V}\mathrm{ar}\left(e^{X_1}\right) = \frac{100\, pe^2}{1-e^2(1-p)} - 100\left(\frac{pe}{1-e(1-p)}\right)^2$$

$$= 1418.25$$

$$\sigma = \sqrt{\mathbb{V}\mathrm{ar}\left(\sum_{i=1}^{100} e^{X_i}\right)} = 37.66$$

$$\mathbb{P}\left(\sum_{i=1}^{100} e^{X_i} < 300\right) = \mathbb{P}\left(\frac{\sum_{i=1}^{100} e^{X_i} - \mu}{\sigma} < -\frac{35.97}{37.66}\right)$$

$$\overset{CLT}{\approx} \Phi(-0.955)$$

$$\approx 0.171$$

13. * Sia X una variabile aleatoria con media nulla e momento quarto finito. Dimostrare che esiste una costante $C > 0$ tale che per ogni $\epsilon > 0$ ed ogni intero positivo n

$$P[|\bar{X}_n| \geq \epsilon] \leq \frac{C}{\epsilon^4 n^2}$$

con $\bar{X}_n$ la media empirica di n copie indipendenti di X.

$$\mu = 0 \qquad \sigma = \sqrt{\mathrm{Var}\, X} \implies \sigma_n = \sqrt{\mathrm{Var}\, \bar{X}_n} = \frac{1}{\sqrt{n}}\sigma$$

$$\mathbb{P}(|\bar{X}_n| \geq \varepsilon) = \mathbb{P}\left(\left|\frac{\bar{X}_n}{\sigma_n}\right| \geq \frac{\varepsilon}{\sigma_n}\right) = \mathbb{P}\left(\left|\frac{\bar{X}_n}{\sigma_n}\right|^4 \geq \frac{\varepsilon^4}{\sigma_n^4}\right)$$

$$Y_n = \left(\frac{\bar{X}_n}{\sigma_n}\right)^4 \qquad = \mathbb{P}\left(Y_n \geq \frac{\varepsilon^4}{\sigma_n^4}\right) \overset{M}{\leq} \frac{\mathbb{E}(Y_n)}{\varepsilon^4 / \frac{\sigma^4}{n^2}} = \frac{\mathbb{E}(Y_n)\sigma^4}{n^2 \varepsilon^4}$$

Dobbiamo dimostrare $\mathbb{E}(Y_n) < \infty$

$$\mathbb{E}(Y_n) = \frac{1}{\sigma_n^4} \cdot \mathbb{E}\left(\left(\frac{1}{n}\sum X_i\right)^4\right) = \frac{n^2}{\sigma^4} \cdot \frac{1}{n^4} \mathbb{E}\left(\left(\sum_{i=1}^n X_i\right)^4\right)$$

$$= \frac{1}{n^2\sigma^4} \cdot \mathbb{E}\left(\sum_{i,j,e,l=1}^n X_i X_j X_e X_l\right)$$

$$= \frac{1}{n^2\sigma^4} \mathbb{E}\left(\sum_{j=1}^n X_j^4 + \sum_{j,e=1}^n X_j^2 X_e^2\right)$$

$$= \frac{1}{n^2\sigma^4}\left(n\,\mathbb{E}(X^4) + n^2\,\mathbb{E}(X^2)\right)$$

$$\leq \frac{1}{\cancel{n^2}\sigma^4}\left(\cancel{n^2}\mathbb{E}(X^4) + \cancel{n^2}\mathbb{E}(X^2)\right) = \frac{C}{\sigma^4}$$

14. (Opzionale) Siano X, Y, Z variabili aleatorie discrete che assumono valori distinti con probabilità 1 e siano $a = P[X > Y]$, $b = P[Y > Z]$, $c = P[Z > X]$.

   (a) Dimostrare che $\min\{a, b, c\} \leq \frac{2}{3}$ e fornire un esempio in cui si ottiene l'uguaglianza $\min\{a, b, c\} = \frac{2}{3}$

   (b) Assumiamo che $P[X = 0] = 1$, che Y, Z siano indipendenti e tali che $P[Z = 1] = P[Y = -1] = p$ e $P[Z = -2] = P[Y = 2] = 1 - p$. Calcolare $\min\{a, b, c\}$ per $p \in [0, 1]$ fissato e calcolare $\sup_{p \in [0,1]} \min\{a, b, c\}$.

Foglio di esercizi 6

Discussione soluzioni: 18.04.2023

1. Un esperimento consiste nel ripetere 60 volte l'estrazione (con reinserimento) di una biglia da un'urna contenente 50 biglie, 15 delle quali sono bianche.
   (a) Calcolare approssimativamente, usando il teorema del limite centrale (TLC), la probabilità che il numero di biglie bianche estratte sia maggiore di 10 e minore o uguale di 30.
   (b) Supponendo ora che il numero totale di palline nel contenitore sia 500 (sempre con 15 biglie bianche), si calcoli approssimativamente la probabilità che il numero di biglie bianche estratte sia minore o uguale a 2 usando un'approssimazione più appropriata di quella data dal TLC.
2. Sia $X \sim \text{NegBin}(100, 1/3)$.
   (a) Maggiorare $P(X > 500)$ usando la disuguaglianza di Markov.
   (b) Maggiorare $P(X > 500)$ usando la disuguaglianza di Chebyshev.
   (c) Ricordiamo che X può essere scritta come la somma di 100 variabili aleatorie IID rcon distribuzione Geom(1/3). Usando questa rappresentazione ed il TLC, trovare un valore approssimativo per $P(X > 500)$.
3. Se non ancora fatto, svolgere l'esercizio 12 del foglio precedente.

Proprietà utili della funzione cumulativa Φ di una normale standard

- Valore limite: $\Phi(\infty) = \int_{-\infty}^{\infty} \phi(x)dx = 1$
- Simmetria: $\Phi(a) = \int_{-\infty}^{a} \phi(x)dx = \int_{-\infty}^{\infty} \phi(x)dx - \int_{a}^{\infty} \phi(x)dx = 1 - \int_{-\infty}^{-a} \phi(x)dx = 1 - \Phi(-a)$
- Valori tabellati di $\Phi(a)$ per l'intervallo $a \in [0, \infty)$ (per $a \in (-\infty, 0)$ usare simmetria):

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

Thm (CLT) $\{X_1 \dots X_n\}$ sequenze di VA iid $\begin{cases} \mathbb{E}(X_i) = \mu \in \mathbb{R} \\ Var(X_i) = \sigma^2 < \infty \end{cases}$

definiamo $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$

$$\frac{\sqrt{n}\,(\overline{X}_n - \mu)}{\sigma} \longrightarrow \mathcal{N}(0,1)$$

Def: $X \sim \mathcal{N}(\mu, \sigma^2)$ significa

$$\mathbb{P}(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-(x-\mu)^2/2\sigma^2} dx$$

$$\mathbb{P}\left(\frac{\sqrt{n}\,(\overline{X}_n - \mu)}{\sigma} < z \right) \xrightarrow{n \to \infty} \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} dx$$

$$= \Phi(z)$$

$$\Longrightarrow \mathbb{P}(a \leq \overline{X}_n \leq b) = \mathbb{P}\left(\alpha_n \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq \beta_n \right)$$

$$= \Phi(\beta_n) - \Phi(\alpha_n)$$

$$\begin{cases} \alpha_n = \dfrac{a - \mu}{\sigma/\sqrt{n}} \\[2ex] \beta_n = \dfrac{b - \mu}{\sigma/\sqrt{n}} \end{cases}$$

1. Un esperimento consiste nel ripetere 60 volte l'estrazione (con reinserimento) di una biglia da un'urna contenente 50 biglie, 15 delle quali sono bianche.

   (a) Calcolare approssimativamente, usando il teorema del limite centrale (TLC), la probabilità che il numero di biglie bianche estratte sia maggiore di 10 e minore o uguale di 30.

   (b) Supponendo ora che il numero totale di palline nel contenitore sia 500 (sempre con 15 biglie bianche), si calcoli approssimativamente la probabilità che il numero di biglie bianche estratte sia minore o uguale a 2 usando un'approssimazione più appropriata di quella data dal TLC.

a) $p = \mathbb{P}(\text{pescare biglia bianca}) = \frac{15}{50} = \frac{3}{10} = 0.3$

$n = 60$

$S_n = \#\ \text{biglie bianche estratte} \sim \text{Bin}(n,p)$

$\mathbb{E}(S_n) = n \cdot p = 60 \cdot 0.3 = 18$

$\text{Var}(S_n) = n \cdot p \cdot (1-p) = 18 \cdot 0.7 = 12{,}2$

$\Longrightarrow SD(S_n) = \sqrt{\text{Var}(S_n)} \simeq 3.5$

$$\mathbb{P}(10 \leq S_n \leq 30) = \mathbb{P}\left(\frac{10-18}{3.5} \leq \frac{S_n - \mathbb{E}(S_n)}{SD(S_n)} \leq \frac{30-18}{3.5}\right)$$

$$\simeq \mathbb{P}(-2.3 \leq Z \leq 3.5) = \Phi(3.5) - \Phi(-2.3)$$

$$= \Phi(3.5) - (1 - \Phi(2.3)) \approx \Phi(2{,}3) \approx 0.9893$$

b) Siccome $p = \frac{15}{500} = 0.03 \ll 1$ meglio usare Poisson.

con $\mu = n \cdot p = 1.8$

$$\Longrightarrow \mathbb{P}(S_n \leq 2) \approx e^{-\mu}\frac{\mu^0}{0!} + e^{-\mu}\frac{\mu^1}{1!} + e^{-\mu}\frac{\mu^2}{2!}$$

$$\simeq 0.73.$$

2. Sia $X \sim \text{NegBin}(100, 1/3)$.

(a) Maggiorare $P(X > 500)$ usando la disuguaglianza di Markov.

(b) Maggiorare $P(X > 500)$ usando la disuguaglianza di Chebyshev.

(c) Ricordiamo che X può essere scritta come la somma di 100 variabili aleatorie IID rcon distribuzione Geom(1/3). Usando questa rappresentazione ed il TLC, trovare un valore approssimativo per $P(X > 500)$.

a) $\mathbb{P}(X > 500) \overset{M}{\leq} \dfrac{\mathbb{E}(X)}{500} = \dfrac{100 \cdot \frac{1}{1/3}}{500} \leq \dfrac{3}{5}$

b) $\mathbb{P}(X > 500) \leq \mathbb{P}\left(X - \mathbb{E}(X) \geq 500 - 300\right)$

$\text{Var}(X) = 100 \dfrac{2/3}{(1/3)^2} = 600$

$\leq \mathbb{P}(|X - \mathbb{E}(X)| \geq 200) \leq \dfrac{SD(X)^2}{200^2} = \dfrac{3}{2} 10^{-2}$

c) $X = S_n = \sum_{i=1}^{100} T_i \qquad T_i \overset{iid}{\sim} \text{Geom}\left(\frac{1}{3}\right)$

$\mathbb{P}(X > 500) = \mathbb{P}\left(\dfrac{S_n - \mathbb{E}(S_n)}{SD(S_n)} \leq \dfrac{500 - \mathbb{E}(S_n)}{SD(S_n)}\right)$

$\approx \mathbb{P}\left(Z \leq \dfrac{20}{\sqrt{6}}\right) \approx \Phi(8.16) \approx 1$

3. Se non ancora fatto, svolgere l'esercizio 12 del foglio precedente.

a) $\mathbb{E}(e^{X_1}) = \sum_{k=1}^{\infty} e^{k}(1-p)^{k-1} \cdot p = p \cdot e \cdot \sum_{k=1}^{\infty} [e(1-p)]^{k-1} =$

$= p \cdot e \cdot \sum_{k=0}^{\infty} [e(1-p)]^{k} = \frac{p \cdot e}{1-e(1-p)}$ se $e \cdot (1-p) < 1$

$\Longleftrightarrow p > \frac{e-1}{e}$

$Var(e^{X_1}) = \mathbb{E}(e^{2X_1}) - \mathbb{E}(e^{X_1})^2$

$\mathbb{E}(e^{2X_1}) = \sum_{k=1}^{\infty} e^{2k}(1-p)^{k-1} p = pe^2 \sum_{k=1}^{\infty} [e^2(1-p)]^{k-1} = pe^2 \frac{1}{1-e^2(1-p)}$

se $e^2(1-p) < 1 \Longleftrightarrow p > \frac{e^2-1}{e^2} > \frac{e-1}{e}$

b) $p = 0.9 > \frac{e^2-1}{e^2} \Longrightarrow \mathbb{E}(e^{X_i}), Var(e^{X_i})$ exist.

$\mu = \mathbb{E}\left(\sum_{i=1}^{100} e^{X_i}\right) = 100\, \mathbb{E}(e^{X_1}) = \frac{100 \cdot p \cdot e}{1-e(1-p)} = 335,97$

$Var\left(\sum_{i=1}^{100} e^{X_i}\right) = 100\, Var(e^{X_1}) = \frac{100\, pe^2}{1-e^2(1-p)} - 100\left(\frac{pe}{1-e(1-p)}\right)^2$

$= 1418.25$

$\sigma = \sqrt{Var\left(\sum_{i=1}^{100} e^{X_i}\right)} = 37.66$

$\mathbb{P}\left(\sum_{i=1}^{100} e^{X_i} < 300\right) = \mathbb{P}\left(\frac{\sum_{i=1}^{100} e^{X_i} - \mu}{\sigma} < -\frac{35.97}{37.66}\right)$

$$\overset{CLT}{\approx} \Phi(-0.955)$$

$$\approx 0.171$$

Revisione generale:

Spazi di probabilità: $(\Omega, \Sigma, \mathbb{P})$ → spazio degli esiti (Ω), σ-algebra (Σ), misura di probabilità ($\mathbb{P}$)

Assiomi di probabilità:
- $\mathbb{P}(\Omega) = 1$
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A) \quad \forall A \in \Sigma$
- $\mathbb{P}(\bigcup_j A_j) = \sum \mathbb{P}(A_j) \quad \forall A_1, A_2, \dots \quad A_i \cap A_j = \phi$

⇒ Probabilità uniforme: $\mathbb{P}(\omega) = \dfrac{1}{|\Omega|}$, $\mathbb{P}(A) = \dfrac{|A|}{|\Omega|}$ $A \in \Sigma$

⇒ Combinatoria:

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

- Permutazioni: $|\{(\omega_1 \dots \omega_n) \in \Omega^n : \omega_i \neq \omega_j\}| = n!$
- Disposizioni 1: $|\{(\omega_1, \dots, \omega_k) \in \Omega^k\}| = n^k$
- Disposizioni 2: $|\{(\omega_1, \dots, \omega_k) \in \Omega^k : \omega_i \neq \omega_j\}| = \dfrac{n!}{(n-k)!}$
- Combinazioni: $|\{\{\omega_1, \dots, \omega_k\} \in \Omega^k : \omega_i \neq \omega_j\}| = \binom{n}{k}$

Probabilità condizionale: $\mathbb{P}(A|B) := \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ se $\mathbb{P}(B) > 0$

Teorema di Bayes: $\mathbb{P}(B|A) = \dfrac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)}$

Disintegrazione: $\mathbb{P}(A) = \sum \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)$ se $\{B_i\}$ partizione di A.

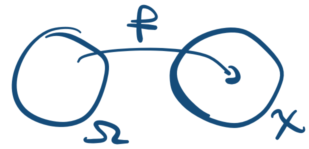
Indipendenza: $\{A_1, A_2, \dots\}$ indip. se $\forall J \subset \{1, 2, \dots\}$ finito

$$\mathbb{P}\Big(\bigcap_{j\in J} A_j\Big) = \prod_{j\in J}\mathbb{P}(A_j)$$

Variabili aleatorie

Def: Una variabile aleatoria è una funzione (misurabile) $X : \Omega \to \mathcal{X}$

Scriviamo il suo insieme delle immagini $\text{range}(X)$



Def: La densità discreta (o fdm) di una variabile aleatoria è una funzione $p_X : \text{range}(X) \to \mathbb{R}_{\geq 0}$ tale che

$$p_X(x) = \mathbb{P}(X = x) \qquad \forall x \in \text{range}(X)$$

Rem: Dalla fdm possiamo ricavare la distribuzione di X:

$$\mathbb{P}(X \in B) = \sum_{x\in B} p_X(x)$$

Def: Valore atteso $\mathbb{E}(X) = \sum_{x \in \text{range}(X)} x\, p_X(x)$ se $\sum |x|\, p_X(x) < \infty$

Varianza $\text{Var}(X) = \sum_{x\in \text{ran}(X)} (x - \mathbb{E}(X))^2 p_X(x) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \geq 0$

↳ se esiste.

$$SD(X) = \sqrt{\text{Var}(X)}$$

Covarianza $\text{Cov}(X,Y) = \sum_{\substack{x\in \text{ran}(X)\\ y \in \text{ran}(Y)}} (x - \mathbb{E}(X))(y - \mathbb{E}(Y)) =$

$$= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Proprietà: • (linearità) $\mathbb{E}(aX+b) = a\mathbb{E}(X)+b$

• (linearità') $\mathbb{E}(X+Y) = \mathbb{E}(X)+\mathbb{E}(Y)$

• se $X \perp\!\!\!\perp Y$: $\mathbb{E}(X\cdot Y) = \mathbb{E}(X)\cdot\mathbb{E}(Y)$

• se $X \perp\!\!\!\perp Y$: $Cov(X,Y) = 0$

• $Cov(aX+Z, Y) = a\,Cov(X,Y) + Cov(Z,Y)$

• se $X \perp\!\!\!\perp Y$: $Var(X+Y) = Var(X) + Var(Y)$

Esempi notevoli:

• <u>Variabili di Bernoulli</u>: $I \sim Ber(p)$ $\quad p \in (0,1)$ se

$range(I) = \{0,1\}$, $\mathbb{P}(I=1) = p$, $\mathbb{P}(I=0) = 1-p$

$\mathbb{E}(I) = \mathbb{P}(I=1) = p$ $\qquad Var(I) = p(1-p)$

• <u>Variabili Binomiali</u>: $X \sim Bin(n,p)$ $\quad n \in \mathbb{N}$, $p \in (0,1)$

$X = \sum_{j=1}^{n} I_j$ $\qquad I_j \overset{iid}{\sim} Ber(p)$ $\qquad \mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$

$\mathbb{E}(X) = n\cdot p$ $\qquad Var(X) = np(1-p)$

- Variabili Geometriche $T \sim Geom(p)$ $p \in (0,1)$

$$\mathbb{P}(T=k) = p(1-p)^{k-1}$$

$$\mathbb{E}(T) = \frac{1}{p} \qquad Var(T) = \frac{1-p}{p^2}$$

- Variabili negative binomiali $S \sim NegBin(m, p)$

$$\mathbb{P}(S=k) = \binom{k-1}{m-1} p^m (1-p)^{k-m} \qquad S = \sum_{i=1}^{m} T_i$$

$$\mathbb{E}(S) = \frac{m}{p} \qquad Var(S) = m\,\frac{1-p}{p^2} \qquad T_i \overset{iid}{\sim} Geom(p)$$

- Variabili di Poisson $X \sim Pois(\lambda)$ $\lambda > 0$

$$\mathbb{P}(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

$$\mathbb{E}(X) = \lambda \quad , \quad Var(X) = \lambda$$

$$X \sim Pois(\lambda_1)\ ,\ Y \sim Pois(\lambda_2) \quad X \perp\!\!\!\perp Y \implies X+Y \sim Pois(\lambda_1+\lambda_2)$$

- Condizionale: $\mathbb{P}(X=k \mid Y=\ell) = \dfrac{\mathbb{P}(X=k, Y=\ell)}{\mathbb{P}(Y=\ell)} \implies \begin{cases} \mathbb{E}(X \mid Y=\ell) \\ Var(X \mid Y=\ell) \end{cases}$

- Thm (Markov): $Y \geq 0 \rightarrow \mathbb{P}(Y \geq a) \leq \dfrac{\mathbb{E}(Y)}{a}$

- Thm (Chebyshev): $\mu = \mathbb{E}(X)$, $\sigma = SD(X) \rightarrow \mathbb{P}(|X-\mu| \geq \alpha) \leq \dfrac{\sigma^2}{\alpha^2}$

- <u>LLN</u>: X_i iid con $\mathrm{Var}(X_i) < \infty$, $\bar{X}_n = \frac{1}{n}\sum^n X_i$

$$\Longrightarrow \mathbb{P}(|\bar{X}_n - \mathbb{E}(X)| > \varepsilon) \xrightarrow{n\to\infty} 0 .$$

Foglio di esercizi 7

Discussione soluzioni: 02.05.2022

* Rappresenta gli esercizi prioritari

1. Sia X una v.a. reale con funzione di ripartizione

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & \text{per } x < -1; \\ \frac{1}{3}x + \frac{2}{3}, & \text{per } -1 \leq x < 1; \\ 1, & \text{per } x \geq 1. \end{cases}$$

   (a) Verificare che F_X sia una funzione di ripartizione.

   (b) X è discreta? continua? assolutamente continua?

   (c) Calcolare le probabilità $P(X \geq 0)$, $P(X = 0)$ e $P(X = -1)$.

2. Sia X una v.a. reale con funzione di ripartizione

$$F_X(x) = \sqrt{x}1_{0<x<1} + 1_{x\geq 1}.$$

   Verificare che X è assolutamente continua e calcolarne la densità.

3. (a) Sia X una v.a. reale discreta e sia $g : \mathbb{R} \to \mathbb{R}$ una funzione boreliana. Si dimostri che $g(X)$ è una v.a. discreta.

   (b) Mostrare che, se X è una variabile aleatoria continua, la variabile aleatoria $W = \max(0, \min(X, 1))$ può essere continua, discreta o né continua né discreta, a seconda della legge di X.

4. * Si consideri X variabile aleatoria di densità:

$$f_X(x) = \begin{cases} \lambda e^{-(x-\lambda)}, & x > \lambda, \\ 0, & x \leq \lambda. \end{cases}$$

   (a) Determinare per quali λ questa è una densità di probabilità e calcolare la funzione di ripartizione corrispondente.

   (b) Posta $Y = e^X$, si determini la distribuzione di Y.

   (c) Si calcoli $P(Y < 2)$.

   (d) Si calcolino $\mathbb{E}(Y)$, $\mathbb{E}(Y^2)$ e $\mathrm{Var}(Y)$.

   (e) Si calcolino $P(X > 3)$ e $P(X^3 > 27)$.

   (f) Si calcoli la media di $3X^2 - 1$.

   (g) Stimare la probabilità che X assuma valori distanti dalla media più di 2 unità, utilizzando la disuguaglianza di Chebychev.

5. Il raggio R di un certo tipo di particella inquinante, espresso in micron, è una variabile aleatoria con densità di probabilità

$$f(x) = \begin{cases} c\,x\,e^{-x^2}, & \text{per } x > 0, \\ 0, & \text{per } x \leq 0. \end{cases}$$

   (a) Determinare il valore della costante reale c.

   (b) Calcolare la funzione di ripartizione di R.

(c) Calcolare la mediana e la media di R.

(d) Calcolare la probabilità che una particella abbia un raggio superiore a 2 micron.

(e) Calcolare la legge della variabile aleatoria X che vale 1 se $R < 2$ e 0 altrimenti.

(f) Supponendo che le particelle inquinanti siano delle sfere, calcolare la probabilità che una particella abbia un volume superiore ad 1 micron cubo.

6. Sia X una v.a. reale, definita su uno spazio $(\Omega, \mathcal{A}, P)$, con $X \geq 0$ q.c.; sia F_X la sua funzione di ripartizione. Dimostrare che

$$E[X] = \int_0^\infty P\{X > t\}dt = \int_0^\infty (1 - F_X(t))dt.$$

Sia ora X una v.a. reale (non necessariamente ≥ 0 q.c.). Dedurre che X è integrabile se e solo se

$$\int_0^\infty P\{X > t\}dt < \infty \quad \text{e} \quad \int_{-\infty}^0 P\{X < t\}dt < \infty$$

e in questo caso vale

$$E[X] = \int_0^\infty P\{X > t\}dt - \int_{-\infty}^0 P\{X < t\}dt.$$

7. * Data $X \sim \text{Exp}(\lambda)$, $\lambda > 0$, se ne consideri il suo arrotondamento per eccesso Y, ovvero

$$Y = \sum_{k=1}^{+\infty} k\, I_{(k-1,k]}(X) = \lceil X \rceil.$$

Se quindi X rappresenta un tempo d'attesa, Y rappresenta il corrispondente tempo d'attesa per un osservatore stroboscopico. Si mostri che Y è una variabile aleatoria e se ne calcoli la distribuzione.

Sia poi $Z = Y - X$. Si calcoli la funzione di ripartizione di Z.
Suggerimento: si considerino tutti i possibili valori di Y.

8. * Un vettore aleatorio (X, Y) ha la seguente densità di probabilità

$$f_{(X,Y)}(x, y) = \frac{4xy}{\theta^2} e^{-x^2/\theta}\ 1_{(0,\infty)\times(0,x)}(x, y).$$

(a) Le variabili aleatorie X e Y sono indipendenti?

(b) Calcolare le densità marginali di (X, Y).

(c) Per $\theta = 1$, calcolare $P\{X > 1, Y > 1\}$.

(d) Per $\theta = 1$, calcolare $E[X/Y]$.

9. * Siano $X_1, \dots, X_n$ variabili aleatorie indipendenti con densità $f_1, \dots, f_n$ e funzioni di ripartizione $F_1, \dots, F_n$. Calcolare, la funzione di ripartizione e la densità di $U = \max\{X_1, \dots, X_n\}$, $W = \min\{X_1, \dots, X_n\}$. Considerare il caso $f_1 = \dots = f_n = f$, ed in particolare se f è

(a) la densità di una variabile aleatoria uniforme su $(0, 1)$

(b) la densità di una variabile aleatoria esponenziale con parametro $\lambda > 0$.

10. Per quali valori dei parametri le seguenti sono funzioni di ripartizione? per quali valori le corrispondenti v.a. sono continue? assolutamente continue? Per i casi di v.a. assolutamente continue, si determini la densità.

(a) $F(x) = \lambda(\arctan x + \pi/2)$;

(b) $F(x) = ae^{\lambda x} 1_{(-\infty,0)}(x) + (b - ce^{-\lambda x}) 1_{[0,+\infty)}(x)$, con $\lambda > 0$.

11. * Siano X, Y v.a. reali indipendenti con $X \sim U([0,1])$ e Y assolutamente continua con densità $f_Y(y) = 2y1_{(0,1)}(y)$. Calcolare:

    (a) $P\{X > 1/2, Y > 1/2\}$;

    (b) $P\{X^2 > Y\}$;

    (c) $P\{X > 1/2, X + Y < 1\}$;

    (d) $E\left[\frac{X^2}{Y}\right]$.

12. Sia X una v.a. reale continua (cioè tale che $P\{X = x\} = 0$ per ogni $x \in \mathbb{R}$) e contentrata su $(0, +\infty)$ (cioè tale che $P\{X \in (0, +\infty)\} = 1$). Si dice che X gode della proprietà di assenza di memoria se

$$P(X > s + t|X > s) = P(X > t), \quad \forall s, t > 0.$$

    Dimostrare che una v.a. X continua e concentrata su $(0, +\infty)$ gode della proprietà di assenza di memoria se e solo se $X \sim \text{Exp}(\lambda)$ (variabile esponenziale) per qualche $\lambda > 0$.
    Suggerimento: calcolare prima la funzione di ripartizione della v.a. esponenziale.

13. Una persona compra un lotto di 5 lampadine led. Il lotto può arrivare da due aziende (a) e (b): arriva da (a) con probabilità 2/3 e da (b) con probabilità 1/3.
    Il tempo di vita (misurato in anni) di una lampadina prodotta da (a), rispettivamente da (b), segue una distribuzione esponenziale di parametro 1/4, rispettivamente 1/6.
    Siano X_i, $i = 1, \ldots 5$, i tempi di vita delle 5 lampadine.

    (a) Modellizzare il problema.

    (b) Calcolare la legge di X_1.

    (c) Le v.a. X_i sono indipendenti?

    (d) Calcolare la probabilità che almeno 2 lampadine abbiano tempo di vita maggiore di 6 anni.

    (e) Se almeno 2 lampadine hanno tempo di vita maggiore di 6 anni, calcolare la probabilità che le lampadine vengano dall'azienda (a).

14. (Disuguaglianza di Jensen) Sia ϕ una funzione convessa[1]. Si consideri una variabile aleatoria X tale che $P(X \in (a, b)) = 1$. Si assuma inoltre che $E(X)$ e $E(\phi(X))$ esistono, cioé che $E(|X|) < \infty$ e $\mathbb{E}(|\phi(X)|) < \infty$.

    (a) Mostrare che per ogni $c \in (a, b)$ si può trovare una funzione lineare $l : \mathbb{R} \to \mathbb{R}$ tale che

$$\phi(x) \geq l(x) \qquad \text{per ogni } x \in (a, b)$$

    e che $\phi(c) = l(c)$.
    Suggerimento: Si può assumere che la funzione ϕ ammetta derivate da sinistra ϕ'_- e da destra ϕ'_+ per ogni $x \in (a, b)$. Spiegare perché $\phi_-(x) \leq \phi_+(x)$ per ogni $x \in (a, b)$ e poi dimostrare che per ogni $s \in [\phi_-(c), \phi_+(c)]$ si può costruire una funzione l con le proprietà desiderate.

    (b) Mostrare che sotto le assunzioni su X di cui sopra vale che

$$\mathbb{E}(\phi(X)) \geq \phi(\mathbb{E}(X))$$

    Suggerimento: usare il risultato di a) scegliendo attentamente il valore di c.

[1] Ricordiamo che una funzione ϕ è convessa sull'intervallo (a, b) se per ogni $x, y \in (a, b)$ e $\lambda \in [0, 1]$ vale che

$$\phi(\lambda x + (1 - \lambda)y) \leq \lambda\phi(x) + (1 - \lambda)\phi(y)$$

1. Sia X una v.a. reale con funzione di ripartizione

$$F_X(x) = P(X \le x) = \begin{cases} 0, & \text{per } x < -1; \\ \frac{1}{3}x + \frac{2}{3}, & \text{per } -1 \le x < 1; \\ 1, & \text{per } x \ge 1. \end{cases}$$

(a) Verificare che F_X sia una funzione di ripartizione.

(b) X è discreta? continua? assolutamente continua?

(c) Calcolare le probabilità $P(X \ge 0)$, $P(X = 0)$ e $P(X = -1)$.

Richiamo: funzione di ripartizione $F_X(x) = \int_{-\infty}^{x} f_X(z)\,dz$

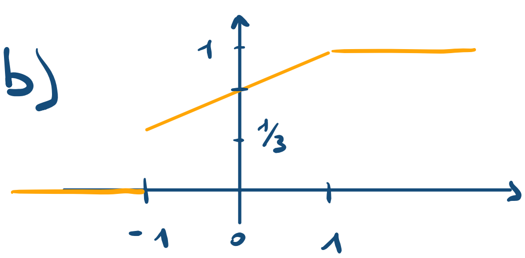
- non decrescente, non negativa
- continua a destra
- $\lim_{x \to \infty} F_X(x) = 1$, $\lim_{x \to -\infty} F_X(x) = 0$

a) • Chiaramente $F_X(x) \ge 0$

• $F_X(x) \ge F_X(x')$ se $x \ge x'$

• $\lim_{z \downarrow x} F_X(z) = F_X(x)$ (gli unici punti dove non si ha continuità sono $x = 1$, $x = -1$, dove la condizione è facilmente verificata)

• $\lim_{z \to -\infty} F_X(z) = 0$, $\lim_{z \to \infty} F_X(z) = 1$

b) 

• la variabile non è discreta:

$P(X \in [-1,1)) = \frac{2}{3}$ ma

$P(X = x) = F_X(x) - \lim_{z \uparrow x} F_X(z) = 0 \quad \forall x \in (-1,1)$

• non è continua: $P(X = -1) = F_X(-1) - \lim_{z \uparrow -1} F_X(z) = \frac{1}{3} \neq 0$

$\Longrightarrow$ non è assol. continua

c) $P(X = -1) = \frac{1}{3}$, $P(X = 0) = 0$, $P(X \ge 0) = 1 - \lim_{z \uparrow 0} F_X(z) = 1 - \frac{2}{3} = \frac{1}{3}$

2. Sia X una v.a. reale con funzione di ripartizione

$$F_X(x) = \sqrt{x}1_{0<x<1} + 1_{x\geq 1}.$$

Verificare che X è assolutamente continua e calcolarne la densità.

Consideriamo l'ansatz per la derivata:

$$f_X(x) = \begin{cases} \frac{1}{2\sqrt{x}} & x \in (0,1) \\ 0 & \text{altrimenti} \end{cases}$$

notiamo che $\displaystyle\int_{-\infty}^{x} f_X(t)\,dz = \begin{cases} 0 & x \in (-\infty, 0) \\ \sqrt{x} & x \in [0,1) \\ 1 & x \in [1,\infty) \end{cases}$

che corrisponde a $F_X(x)$. F_X è quindi assolutamente continua con densità $f_X(x)$.

3. (a) Sia X una v.a. reale discreta e sia $g : \mathbb{R} \to \mathbb{R}$ una funzione boreliana. Si dimostri che $g(X)$ è una v.a. discreta.

(b) Mostrare che, se X è una variabile aleatoria continua, la variabile aleatoria $W = \max(0, \min(X, 1))$ può essere continua, discreta o né continua né discreta, a seconda della legge di X.

a) Sia R_{p_X} il range di X, allora per ogni $y \in g(R_{p_X})$ abbiamo $y = g(x)$ per almeno un $x \in R_{p_X}$ e siccome X è discreta $\mathbb{P}(X=x) > 0$, quindi

$$\mathbb{P}(g(X)=y) \geq \mathbb{P}(X=x) > 0.$$

b) continua: $f_X(x) = \begin{cases} 1 & x \in (0,1) \\ 0 & \text{altrimenti} \end{cases}$

$$\Rightarrow f_W(x) = F_W'(x) = F_X'(x) = f_X(x) \Rightarrow \text{continua}$$

discreta: $f_X(x) = \begin{cases} 1 & x \in (1,2) \\ 0 & \text{altrimenti} \end{cases}$

$$\Rightarrow F_W(x) = \begin{cases} 0 & \text{se } x < 1 \\ 1 & \text{se } x \geq 1 \end{cases}$$

né discreta né continua: $f_X(x) = \begin{cases} \frac{1}{2} & x \in (0,2) \\ 0 & \text{altrimenti} \end{cases}$

$$\Rightarrow F_W(x) = \begin{cases} 0 & \text{se } x < 0 \\ \frac{1}{2}x & \text{se } x \in [0,1) \\ 1 & \text{se } x \geq 1 \end{cases}$$

4. * Si consideri X variabile aleatoria di densità:

$$f_X(x) = \begin{cases} \lambda e^{-(x-\lambda)}, & x > \lambda, \\ 0, & x \leq \lambda. \end{cases}$$

(a) Determinare per quali λ questa è una densità di probabilità e calcolare la funzione di ripartizione corrispondente.

(b) Posta $Y = e^X$, si determini la distribuzione di Y.

(c) Si calcoli $P(Y < 2)$.

(d) Si calcolino $\mathbb{E}(Y)$, $\mathbb{E}(Y^2)$ e $\text{Var}(Y)$.

(e) Si calcolino $P(X > 3)$ e $P(X^3 > 27)$.

(f) Si calcoli la media di $3X^2 - 1$.

(g) Stimare la probabilità che X assuma valori distanti dalla media più di 2 unità, utilizzando la disuguaglianza di Chebychev.

a) Sempre nonnegativa.

$$1 \overset{!}{=} \int_{-\infty}^{\infty} f_X(x)\,dx = \int_{\lambda}^{\infty} \lambda e^{-(x-\lambda)}\,dx = \lambda e^{\lambda} \int_{\lambda}^{\infty} e^{-x}\,dx = \lambda e^{\lambda} e^{-\lambda} = \lambda$$

$\Rightarrow \lambda = 1$

$$\Rightarrow F_X(x) = \begin{cases} 0 & x < 1 \\ 1 - e^{-(x-1)} & x \geq 1 \end{cases}$$

b) $F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(e^X \leq y) = \mathbb{P}(X \leq \log(y))$

$$= \begin{cases} 0 & \log y < 1 \\ 1 - e^{-(\log(y) - 1)} & \log y \geq 1 \end{cases} = \begin{cases} 0 & y < e \\ 1 - \frac{e}{y} & y \geq e \end{cases}$$

$$\Rightarrow f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{e}{y^2} I_{[e,\infty)}(y)$$

c) $$\mathbb{P}(Y < 2) = \lim_{y \uparrow 2} F_Y(y) = 0$$

d) Siccome $\int_e^{\infty} |y| f_Y(y)\, dy = \infty$ $\mathbb{E}(Y), \mathbb{E}(Y^2), Var(Y)$

non sono definite

e) $\mathbb{P}(X \geq 3) = \int_3^{\infty} e^{(1-x)}\, dx = e^{-2}$

$\mathbb{P}(X^3 \geq 27) = \mathbb{P}(X \geq 3) = e^{-2}$

f) $\mathbb{E}(3X^2 - 1) = 3\mathbb{E}(X^2) - 1 = 3\int_1^{\infty} x^2 e^{1-x}\, dx - 1 = 3 \cdot 5 - 1 = 14$

g) $\mathbb{E}(X) = \int_1^{\infty} x e^{1-x}\, dx = 2$ $\quad Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 5 - 4 = 1$

$\Longrightarrow \mathbb{P}(|X - \mathbb{E}(X)| \geq 2) = \frac{\sigma^2}{4} = \frac{1}{4}$

5. Il raggio R di un certo tipo di particella inquinante, espresso in micron, è una variabile aleatoria con densità di probabilità

$$f(x) = \begin{cases} c\,x\,e^{-x^2}, & \text{per } x > 0, \\ 0, & \text{per } x \leq 0. \end{cases}$$

(a) Determinare il valore della costante reale c.

(b) Calcolare la funzione di ripartizione di R.

(c) Calcolare la mediana e la media di R.

(d) Calcolare la probabilità che una particella abbia un raggio superiore a 2 micron.

(e) Calcolare la legge della variabile aleatoria X che vale 1 se $R < 2$ e 0 altrimenti.

(f) Supponendo che le particelle inquinanti siano delle sfere, calcolare la probabilità che una particella abbia un volume superiore ad 1 micron cubo.

a) $1 \overset{!}{=} c \int_0^\infty x\, e^{-x^2} dx = \frac{c}{2} \implies c = 2$

b) $F_R(x) = \begin{cases} 1 - e^{-x^2} & x \geq 0 \\ 0 & x < 0 \end{cases}$

c) Mediana m : $\mathbb{P}(X \leq m) = \mathbb{P}(X \geq m) = \frac{1}{2}$ se

$$1 - e^{-x^2} = \frac{1}{2} \iff x = \sqrt{\log 2}$$

Media μ : $\mathbb{E}(R) = \int_0^\infty x\, f_X(x)\, dx = \frac{\sqrt{\pi}}{2}$

d) $\mathbb{P}(R > 2) = 1 - \mathbb{P}(R \leq 2) = 1 - F_R(2) = e^{-4}$

e) $p_X(x) = \begin{cases} e^{-4} & \text{se } x = 0 \\ 1 - e^{-4} & \text{se } x = 1 \end{cases}$

f) $P\left(\frac{4}{3}\pi R^3 > 1\right) = P\left(R > \sqrt[3]{3/4\pi}\right) = \exp\left(-\left(\frac{3}{4\pi}\right)^{2/3}\right)$

6. Sia X una v.a. reale, definita su uno spazio $(\Omega, \mathcal{A}, P)$, con $X \geq 0$ q.c.; sia F_X la sua funzione di ripartizione. Dimostrare che

$$E[X] = \int_0^\infty P\{X > t\}dt = \int_0^\infty (1 - F_X(t))dt.$$

Sia ora X una v.a. reale (non necessariamente ≥ 0 q.c.). Dedurre che X è integrabile se e solo se

$$\int_0^\infty P\{X > t\}dt < \infty \quad \text{e} \quad \int_{-\infty}^0 P\{X < t\}dt < \infty$$

e in questo caso vale

$$E[X] = \int_0^\infty P\{X > t\}dt - \int_{-\infty}^0 P\{X < t\}dt.$$

a) $$\mathbb{E}(X) = \int_0^\infty x f_X(x)\,dx = \int_0^\infty x F_X'(x)\,dx = \int_0^\infty x \left(F_X(x) - 1\right)'\,dx$$

$$\overset{PI}{=} -\int_0^\infty x' \left(F_X(x) - 1\right)dx = \int_0^\infty (1 - F_X(x))\,dx = \int_0^\infty \mathbb{P}(X > x)\,dx.$$

b) $$\int_{-\infty}^\infty |x| f_X(x) = \int_{-\infty}^0 (-x) f_X(x)\,dx + \int_0^\infty x f_X(x)\,dx = \mathbb{E}(-X_-) + \mathbb{E}(X_+)$$

$$= \int_0^\infty \mathbb{P}(-X_- > x)\,dx + \int_0^\infty \mathbb{P}(X_+ > x)\,dx$$

$$= \int_0^\infty \mathbb{P}(X_- < -x)\,dx + \int_0^\infty \mathbb{P}(X_+ > x)\,dx$$

$$= \underbrace{\int_{-\infty}^0 \mathbb{P}(X < x)\,dx}_{\geq 0} + \underbrace{\int_0^\infty \mathbb{P}(X > x)\,dx}_{\geq 0}$$

$\longrightarrow$ entrambi devono essere finiti perchè il risultato lo sia

$$\longrightarrow \mathbb{E}(X) = \int_{-\infty}^0 x f_X(x)\,dx + \int_0^\infty x f_X(x)\,dx = -\int_{-\infty}^0 (-x) f_X(x)\,dx + \int_0^\infty x f_X(x)\,dx$$

$$= -\int_{-\infty}^0 \mathbb{P}(X < x)\,dx + \int_0^\infty \mathbb{P}(X > x)\,dx$$

7. * Data $X \sim \mathrm{Exp}(\lambda)$, $\lambda > 0$, se ne consideri il suo arrotondamento per eccesso Y, ovvero

$$Y = \sum_{k=1}^{+\infty} k\, I_{(k-1,k]}(X) = \lceil X \rceil.$$

Se quindi X rappresenta un tempo d'attesa, Y rappresenta il corrispondente tempo d'attesa per un osservatore stroboscopico. Si mostri che Y è una variabile aleatoria e se ne calcoli la distribuzione.

Sia poi $Z = Y - X$. Si calcoli la funzione di ripartizione di Z.
Suggerimento: si considerino tutti i possibili valori di Y.

a) range$(Y) = \mathbb{N} / \{0\} = \{1, 2, 3, \dots\}$

$$\mathbb{P}(Y = k) = \int_{k-1}^{k} \lambda e^{-\lambda x}\, dx = \frac{\lambda}{\lambda}\left(e^{-\lambda(k-1)} - e^{-\lambda k}\right) = e^{-\lambda(k-1)}(1 - e^{-\lambda})$$

$$= (1-q)^{k-1} q \qquad \text{dove} \qquad q = (1 - e^{-\lambda})$$

$$Y \sim \mathrm{Geom}(1 - e^{-\lambda})$$

b) range$(Z) = (0, 1)$.

$$f_Z(z) = \sum_{k=1}^{\infty} \lambda e^{-\lambda(z + (k-1))} = \lambda e^{-z\lambda}\left[\sum_{j=0}^{\infty} e^{-\lambda j}\right] = \frac{\lambda}{1 - e^{-\lambda}} e^{-\lambda z}$$

$z \in (0,1)$

8. * Un vettore aleatorio (X, Y) ha la seguente densità di probabilità

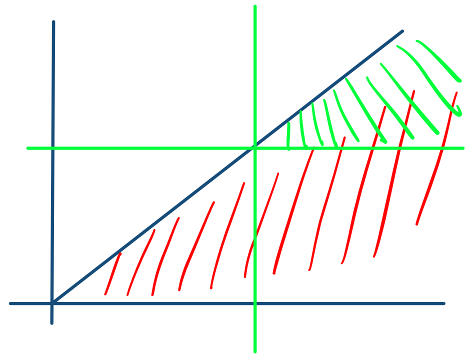
$$f_{(X,Y)}(x,y) = \frac{4xy}{\theta^2} e^{-x^2/\theta} \, 1_{(0,\infty)\times(0,x)}(x,y).$$

(a) Le variabili aleatorie X e Y sono indipendenti?

(b) Calcolare le densità marginali di (X, Y).

(c) Per $\theta = 1$, calcolare $P\{X > 1, Y > 1\}$.

(d) Per $\theta = 1$, calcolare $E[X/Y]$.



a) Nel caso di variabili indipendenti abbiamo

$$f_{(XY)}(x,y) = f_X(x) \cdot f_Y(y) \qquad \forall\, x \in \text{range}(X),\ y \in \text{range}(Y)$$

In questo caso chiaramente $\text{range}(X) = \text{range}(Y) = (0, \infty)$

ma

$$f_{XY}(3,5) = 0 \neq f_X(3) \cdot f_Y(5)$$

$\Longrightarrow$ non sono indipendenti

b) $$f_X(x) = \int_0^x f_{(XY)}(x,y)\, dy = \frac{2x^3 e^{-x^2/\theta}}{\theta^2} \qquad x > 0$$

$$f_Y(y) = \int_y^\infty f_{(XY)}(x,y)\, dx = \frac{2y e^{-y^2/\theta}}{\theta} \qquad y > 0$$

c) $$\mathbb{P}(X > 1, Y > 1) = \int_1^\infty \int_1^x 4xy e^{-x^2} \, dy\, dx = e^{-1}$$

d) $$\mathbb{E}\left(\frac{X}{Y}\right) = \int_0^\infty \int_0^x \frac{x}{y} 4xy e^{-x^2} \, dy\, dx = 2$$

9. * Siano $X_1, \dots, X_n$ variabili aleatorie indipendenti con densità $f_1, \dots, f_n$ e funzioni di ripartizione $F_1, \dots, F_n$. Calcolare, la funzione di ripartizione e la densità di $U = \max\{X_1, \dots, X_n\}$, $W = \min\{X_1, \dots, X_n\}$. Considerare il caso $f_1 = \cdots = f_n = f$, ed in particolare se f è

(a) la densità di una variabile aleatoria uniforme su $(0, 1)$

(b) la densità di una variabile aleatoria esponenziale con parametro $\lambda > 0$.

a) $F_U(u) = \mathbb{P}(\max(X_1 \dots X_n) \leq u) =$

$= \mathbb{P}(X_1 \leq u, X_2 \leq u \dots X_n \leq u) =$

$= \mathbb{P}(X_1 \leq u) \cdot \dots \cdot \mathbb{P}(X_n \leq u) = \prod_{i=1}^{n} F_i(u)$

$\longrightarrow f_U(u) = F_U'(u) = \sum_j F_1(u) \dots F_{j-1}(u) \cdot f_j(u) F_{j+1}(u) \dots F_n(u)$

b) $F_W(w) = \mathbb{P}(\min(X_1 \dots X_n) \leq w) = 1 - \mathbb{P}(\min(X_1 \dots X_n) > w)$

$= 1 - \prod_{i=1}^{n} \mathbb{P}(X_i > w) = 1 - \prod_{i=1}^{n} (1 - \mathbb{P}(X_i \leq w))$

$= 1 - \prod_{i=1}^{n} (1 - F_i(w))$

$f_W(w) = F_W'(w) = \sum_{i=1}^{n} (1 - F_1(w)) \dots (1 - F_{i-1}(w)) \cdot f_i(w) (1 - F_{i+1}(w)) \dots (1 - F_n(w))$

a') $X \sim \mathrm{Unif}((0,1)) \longrightarrow f_X(x) = \begin{cases} 1 & \text{if } x \in (0,1) \\ 0 & \text{else} \end{cases} \longrightarrow F_X(x) = \begin{cases} 0 & x < 1 \\ x & x \in [0,1) \\ 1 & x \geq 1 \end{cases}$

$\Rightarrow$ per $u, w \in (0,1)$ $\begin{cases} f_U(u) = n u^{n-1} \\ f_W(w) = n(1-w)^{n-1} \end{cases}$

b') $X \sim \mathrm{Exp}(\lambda) \Rightarrow f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{else} \end{cases} \Longrightarrow F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & \text{else} \end{cases}$

$\Rightarrow$ per $u, w \geq 0$ $\begin{cases} f_U(u) = n(1 - e^{-\lambda u})^{n-1} \lambda e^{-\lambda u} \\ f_W(w) = n(e^{-\lambda w})^{n-1} \lambda e^{-\lambda w} = n\lambda e^{-\lambda n w} \sim \mathrm{Exp}(n\lambda) \end{cases}$

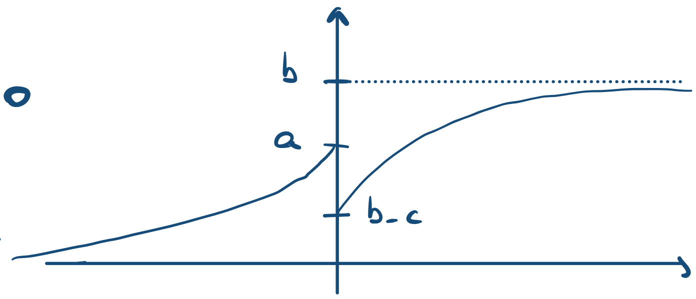
10. Per quali valori dei parametri le seguenti sono funzioni di ripartizione? per quali valori le corrispondenti v.a. sono continue? assolutamente continue? Per i casi di v.a. assolutamente continue, si determini la densità.

(a) $F(x) = \lambda(\arctan x + \pi/2)$;

(b) $F(x) = ae^{\lambda x}1_{(-\infty,0)}(x) + (b - ce^{-\lambda x})1_{[0,+\infty)}(x)$, con $\lambda > 0$.

a) $\lambda = \frac{1}{\pi}$

b) • $F(x) \geq 0$ se $b \geq c$, $b \geq 0$



• Comportamento crescente:

$F'(x) = a\lambda e^{\lambda x} 1_{(-\infty,0)}(x) + c\lambda e^{-\lambda x} 1_{[0,\infty)}(x) \geq 0$

$\lim_{x \uparrow 0} F(x) = a \leq b - c = F(0)$ se $a \leq b - c$ se $\begin{cases} a \geq 0 \\ c \geq 0 \end{cases}$

• $\lim_{x \to -\infty} F(x) = 0$ $\forall a, b, c$

$\lim_{x \to \infty} F(x) = 1$ $b = 1$, $\forall a, c$

$\Longrightarrow$ $b = 1$, $0 \leq a \leq 1 - c$ $0 \leq c \leq 1$

11. * Siano X, Y v.a. reali indipendenti con $X \sim U([0,1])$ e Y assolutamente continua con densità $f_Y(y) = 2y1_{(0,1)}(y)$. Calcolare:

(a) $P\{X > 1/2, Y > 1/2\}$;

(b) $P\{X^2 > Y\}$;

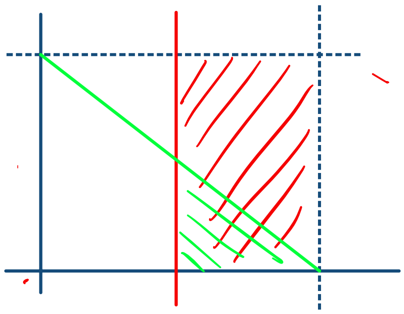
(c) $P\{X > 1/2, X + Y < 1\}$;

(d) $E\left[\frac{X^2}{Y}\right]$.

a) $\mathbb{P}\left(X > \frac{1}{2}, Y > \frac{1}{2}\right) = \mathbb{P}\left(X > \frac{1}{2}\right) \cdot \mathbb{P}\left(Y > \frac{1}{2}\right) = \frac{1}{2} \cdot \int_{1/2}^{1} 2y\, dy = \frac{1}{2} \cdot \frac{3}{4} = \frac{3}{8}$

b) $\mathbb{P}(X^2 > Y) = \int_0^1 \int_0^{x^2} f_{XY}(x,y)\, dy\, dx = \int_0^1 x^4\, dx = \frac{1}{5}$

c) $\mathbb{P}\left(\underline{X > \frac{1}{2}}, \underline{X + Y < 1}\right) = \int_{1/2}^{1} \int_0^{1-x} f_{XY}(x,y)\, dy\, dx = \frac{1}{24}$



d) $\mathbb{E}\left(\frac{X^2}{Y}\right) \overset{X \perp\!\!\!\perp Y}{=} \mathbb{E}(X^2) \cdot \mathbb{E}\left(\frac{1}{Y}\right) = \int_0^1 x^2 dx \cdot \int_0^1 \frac{1}{y} \cdot 2y\, dy = \frac{1}{3} \cdot 2 = \frac{2}{3}$

12. Sia X una v.a. reale continua (cioè tale che $P\{X = x\} = 0$ per ogni $x \in \mathbb{R}$) e contentrata su $(0, +\infty)$ (cioè tale che $P\{X \in (0, +\infty)\} = 1$). Si dice che X gode della proprietà di assenza di memoria se

$$P(X > s + t | X > s) = P(X > t), \quad \forall s, t > 0.$$

Dimostrare che una v.a. X continua e concentrata su $(0, +\infty)$ gode della proprietà di assenza di memoria se e solo se $X \sim \text{Exp}(\lambda)$ (variabile esponenziale) per qualche $\lambda > 0$.
Suggerimento: calcolare prima la funzione di ripartizione della v.a. esponenziale.

$$\mathbb{P}(X > u) = \int_u^\infty \lambda e^{-\lambda t} dt = e^{-\lambda u}$$

"$\Leftarrow$" $\quad \mathbb{P}(X > s+t \mid X > s) = \dfrac{\mathbb{P}(X > s+t)}{\mathbb{P}(X > s)} = \dfrac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathbb{P}(X > t)$

"$\Rightarrow$" $\quad P(X > s+t) = \mathbb{P}(X > s) \cdot \mathbb{P}(X > t)$

$\Leftarrow\!\!\Rightarrow \; F(s+t) = F(s) \cdot F(t)$

$\Rightarrow$ • per ogni intero n : $F(nt) = F(t)^n$

• per ogni intero m : $F(t/m) = F(t)^{1/m}$

$\Rightarrow$ per ogni razionale $F(t \frac{n}{m}) = F(t)^{n/m}$

$\Rightarrow$ prendendo i numeri irrazionali come limiti di razionali

$$\forall x \in \mathbb{R}_+ : \quad F(xt) = F(t)^x$$

$\Rightarrow$ fissiamo $t = 1$: $\quad F(x) = F(1)^x = e^{\ln F(1) \cdot x}$

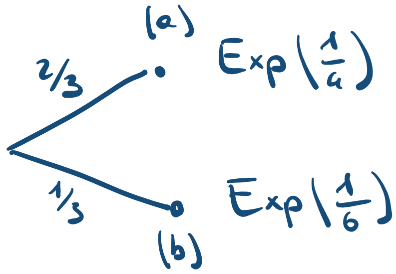
$\Rightarrow \quad F(x) = e^{-\lambda x} \qquad \lambda := -\ln(F(1))$.

$\Rightarrow \quad f_X(x) = \dfrac{d}{dx}(1 - F(x)) = \lambda e^{-\lambda x}$

13. Una persona compra un lotto di 5 lampadine led. Il lotto può arrivare da due aziende (a) e (b): arriva da (a) con probabilità 2/3 e da (b) con probabilità 1/3.
Il tempo di vita (misurato in anni) di una lampadina prodotta da (a), rispettivamente da (b), segue una distribuzione esponenziale di parametro 1/4, rispettivamente 1/6.
Siano X_i, $i = 1, \dots 5$, i tempi di vita delle 5 lampadine.

   (a) Modellizzare il problema.
   (b) Calcolare la legge di X_1.
   (c) Le v.a. X_i sono indipendenti?
   (d) Calcolare la probabilità che almeno 2 lampadine abbiano tempo di vita maggiore di 6 anni.
   (e) Se almeno 2 lampadine hanno tempo di vita maggiore di 6 anni, calcolare la probabilità che le lampadine vengano dall'azienda (a).

a)



$$X_i = I\,Y_i + (1-I)\,Z_i$$

$$Z_i \overset{iid}{\sim} \mathrm{Exp}\left(\tfrac{1}{6}\right) \quad Y_i \overset{iid}{\sim} \mathrm{Exp}\left(\tfrac{1}{4}\right)$$

$$I \sim \mathrm{Be}\left(\tfrac{2}{3}\right)$$

b)
$$\mathbb{P}(X_1 > x) = \mathbb{P}(X_1 > x \mid I = 1)\,\mathbb{P}(I = 1) + \mathbb{P}(X_1 > x \mid I = 0)\,\mathbb{P}(I = 0)$$
$$= \mathbb{P}(Y_1 > x)\,\mathbb{P}(I = 1) + \mathbb{P}(Z_1 > x)\,\mathbb{P}(I = 0)$$
$$= e^{-x/4}\,\frac{2}{3} + \frac{1}{3}\,e^{-x/6}$$

$$\rightarrow f_{X_1}(x) = F_{X_1}'(x) = \mathbb{1}_{(0,\infty)}(x)\,\frac{1}{18}\left(3e^{-x/4} + e^{-x/6}\right)$$

c) Non lo sono:

$$\mathbb{P}(X_1 > x_1, X_2 > x_2) = \frac{2}{3} \cdot e^{-x_1/4}\, e^{-x_2/4} + \frac{1}{3}\, e^{-x_1/6}\, e^{-x_2/6}$$
$$\neq \left(\frac{2}{3} e^{-x_1/4} + \frac{1}{3} e^{-x_1/6}\right)\left(\frac{2}{3} e^{-x_2/4} + \frac{1}{3} e^{-x_2/6}\right)$$

d)
$$\mathbb{P}(N_6 \geq 2) = 1 - \mathbb{P}(N_6 = 0) - \mathbb{P}(N_6 = 1)$$

$$\mathbb{P}(Z_i \geq 6) = e^{-6/6} = e^{-1}$$

$$\mathbb{P}(Y_i \geq 6) = e^{-6/4} = e^{-3/2}$$

$$\Longrightarrow \mathbb{P}(N_6 = 0) = \mathbb{P}(N_6 = 0 \mid I = 0)\,\mathbb{P}(I = 0) + \mathbb{P}(N_6 = 0 \mid I = 1)\,\mathbb{P}(I = 1)$$

$$= \frac{1}{3}\left(1 - e^{-1}\right)^5 + \frac{2}{3}\left(1 - e^{-3/2}\right)^5$$

$$\mathbb{P}(N_6 = 1) = \dots = \frac{1}{3}\binom{5}{1}\left(1 - e^{-1}\right)^4 e^{-1} + \frac{2}{3}\binom{5}{1}\left(1 - e^{-2/3}\right)^4 e^{-2/3}$$

e) $$\mathbb{P}(I = 1 \mid N_6 \geq 2) = \frac{\mathbb{P}(N_6 \geq 2 \mid I = 1)\cdot\mathbb{P}(I = 1)}{\mathbb{P}(N_6 \geq 2)} = \dots$$

14. (Disuguaglianza di Jensen) Sia ϕ una funzione convessa[1]. Si consideri una variabile aleatoria X tale che $P(X \in (a,b)) = 1$. Si assuma inoltre che $E(X)$ e $E(\phi(X))$ esistono, cioé che $E(|X|) < \infty$ e $\mathbb{E}(|\phi(X)|) < \infty$.

(a) Mostrare che per ogni $c \in (a,b)$ si può trovare una funzione lineare $l : \mathbb{R} \to \mathbb{R}$ tale che

$$\phi(x) \geq l(x) \qquad \text{per ogni } x \in (a,b)$$

e che $\phi(c) = l(c)$.
Suggerimento: Si può assumere che la funzione ϕ ammetta derivate da sinistra ϕ'_- e da destra ϕ'_+ per ogni $x \in (a,b)$. Spiegare perché $\phi_-(x) \leq \phi_+(x)$ per ogni $x \in (a,b)$ e poi dimostrare che per ogni $s \in [\phi_-(c), \phi_+(c)]$ si può costruire una funzione l con le proprietà desiderate.

(b) Mostrare che sotto le assunzioni su X di cui sopra vale che

$$\mathbb{E}(\phi(X)) \geq \phi(\mathbb{E}(X))$$

Suggerimento: usare il risultato di a) scegliendo attentamente il valore di c.

a) Sia $x \in (a,b)$, $b - x > t \geq \varepsilon > 0$

$$\Longrightarrow \varphi(x+\varepsilon) = \varphi\left(\frac{\varepsilon}{t}(x+t) + \left(1-\frac{\varepsilon}{t}\right)x\right) \leq \frac{\varepsilon}{t}\varphi(x+t) + \left(1-\frac{\varepsilon}{t}\right)\varphi(x)$$

$$\longrightarrow \frac{\varphi(x+\varepsilon)-\varphi(x)}{\varepsilon} \leq \frac{\varphi(x+t)-\varphi(x)}{t}$$

$$\Longrightarrow \frac{\varphi(x+\varepsilon)-\varphi(x)}{\varepsilon} \text{ è crescente in } \varepsilon.$$

Inoltre se $z = \lambda x + (1-\lambda y)$ $\quad \dfrac{f(z)-f(x)}{z-x} \leq \dfrac{f(y)-f(x)}{y-x}$

$$\Longrightarrow \underbrace{\frac{\varphi(x+\varepsilon)-\varphi(x)}{\varepsilon}}_{\to \varphi'_+(x)} \geq \underbrace{\frac{-\varphi(x-\varepsilon)+\varphi(x)}{\varepsilon}}_{\to \varphi'_-(x)} \Longrightarrow \varphi'_+(x) \geq \varphi'_-(x)$$

$$\begin{cases} \text{for } x < c & \dfrac{\varphi(c)-\varphi(x)}{c-x} \leq \varphi'_-(x) \\ \text{for } x > c & \dfrac{\varphi(x)-\varphi(c)}{x-c} \geq \varphi'_+(x) \end{cases} \quad \xrightarrow{s \in [\varphi_-, \varphi_+]} \quad \begin{cases} \dfrac{\varphi(c)-\varphi(x)}{c-x} \leq s \\ \dfrac{\varphi(x)-\varphi(c)}{x-c} \geq s \end{cases}$$

$$\Longrightarrow \varphi(x) - \varphi(c) \geq s(x-c) \Longrightarrow \varphi(x) \geq sx + (\varphi(c) - sc)$$

b) Sia $c = \mathbb{E}(X) \in (a,b)$

$$\Longrightarrow \mathbb{E}(\varphi(X)) \geq \mathbb{E}(\ell(X)) = \ell(\mathbb{E}(X)) = \varphi(\mathbb{E}(X))$$

Foglio di esercizi 8

Discussione soluzioni: 09.05.2022

* Rappresenta gli esercizi prioritari

1. * Data $X \sim N(\mu, \sigma^2)$, si dimostri che $aX + b \sim N(a\mu + b, a^2\sigma^2)$ se $a \neq 0$.
2. * Data $Z \sim \mathcal{N}(0,1)$, si calcoli la distribuzione di Z^2.
3. * Data $X \sim U\big((0,1)\big)$, si stabilisca per quali α la variabile aleatoria $Y = (1-X)^{-\alpha} \in L^1$.
4. Siano $X_1, \dots, X_n$ variabili aleatorie iid con $X_i \sim Exp(\lambda)$. Si calcoli la legge di $Z_n = \sum_{i=1}^n X_i$. Si tratta di una legge conosciuta?
5. Una variabile aleatoria X è detta di Cauchy se ha densità
$$p(x) = \frac{1}{\pi(1+x^2)}$$
   (a) Verificare che p è una densità di probabilità e scrivere esplicitamente la funzione di ripartizione. X ammette media? E varianza?
   (b) Calcolare la legge di $Y = X^2$. Y ha media?
   (c) Dopo aver verificato che la variabile aleatoria $Z = 1/X$ è ben posta, calcolare la legge di Z. Z ha media? E varianza?
6. Siano X e Y variabili aleatorie indipendenti tali che $X \sim Exp(\lambda)$ e $Y \sim Exp(\mu)$. Calcolare la densità e la funzione di ripartizione di $Z = X + Y$ e $W = \alpha X$, $\alpha \in \mathbb{R}$. Si tratta di leggi note?
7. * Siano (X, Y) le coordinate aleatorie di un punto nel piano, con X e Y due variabili aleatorie Gaussiane standard e indipendenti. Siano (R, θ) le coordinate polari del medesimo punto. Dopo aver determinato la distribuzione di R^2 e di θ, ottenuto la densità di R e aver verificato che R e θ sono indipendenti,
   (a) calcolare il valore atteso di $\frac{X^2}{R^2}$
   (b) calcolare il valore atteso del quoziente tra le variabili aleatorie $\min\{|X|, |Y|\}$ e $\max\{|X|, |Y|\}$.
8. Fissato $\mu \in \mathbb{R}$, sia
$$p(x) = cx^\mu \mathbb{1}_{\{x \in (0,1)\}}.$$
   (a) Determinare μ affinché p sia integrabile su $\mathbb{R}$.
   (b) Per tali valori di μ, trovare c affinché p sia una densità di probabilità.
   (c) Sia X una variabile aleatoria con densità data da p con $\mu = 1$. Calcolare la densità della variabile aleatoria $X^2 + 1$.
   (d) Siano X e Y due variabili aleatorie indipendenti, X con densità p con $\mu = 1$ e Y esponenziale di parametro 1. Calcolare la legge di $X + Y$.
9. * Sia (X, Y) una variabile aleatoria su $\mathbb{R}^2$ con densità continua
$$f_{X,Y}(x, y) = \frac{3}{x^3 y^2} \mathbb{1}_{\{x>1, y>x\}}$$
   (a) Calcolare $P[Y < 2X]$.
   (b) Calcolare la legge congiunta di $U = \log X$ e $V = X/Y$ e le relative densità marginali.

(c) Calcolare la legge di $Z = X + Y$.

10. Si consideri la densità di probabilità data da

$$f_X(x) = \frac{\theta}{x^2} \mathbb{1}_{[\theta,\infty)}(x),$$

con $\theta > 0$. Si proponga un opportuno stimatore per θ basato su un campione casuale $X_1, X_2, \ldots, X_n$ estratto da una popolazione di densità f_X .

(a) Stabilire se lo stimatore proposto è corretto.

(b) Stabilire se la successione di stimatori proposta, al variare della numerosità del campione è consistente.

(c) Utilizzare lo stimatore trovato per determinare intervalli di fiducia al livello $1-\alpha$ per il parametro θ.

11. * Sia dato un campione $X_1, \ldots, X_n$ di legge uniforme su $(0, \theta)$, $\theta > 0$. Siano

$$\hat{\theta}_n = \max(X_1, \ldots, X_n), \qquad \bar{\theta}_n = \frac{2}{n} \sum_{i=1}^{n} X_i$$

(a) $\hat{\theta}_n$ è una stima corretta di θ? è asintoticamente corretta?

(b) Calcolare il rischio quadratico (medio) della stima $\hat{\theta}_n$ di θ $R(\theta, \hat{\theta}_n) = \mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2]$.

(c) è vero che $\lim_{n\to\infty} R(\theta, \hat{\theta}_n) = 0$? La successione di stime $(\hat{\theta}_n)_n$ è consistente?

(d) Dimostrare che $\hat{\theta}_n$ coincide con la stima di massima verosimiglianza per θ.

(e) La stima $\bar{\theta}_n$ di θ è corretta? è asintoticamente corretta?

(f) Calcolare il rischio quadratico (medio) della stima $\bar{\theta}_n$ di θ $R(\theta, \bar{\theta}_n) = \mathbb{E}_\theta[(\bar{\theta}_n - \theta)^2]$

(g) è vero che $\lim_{n\to\infty} R(\theta, \bar{\theta}_n) = 0$? Tale stima $\bar{\theta}_n$ è consistente?

(h) $\hat{\theta}_n$ è preferibile rispetto a $\bar{\theta}_n$?

12. Si consideri un campione $X_1, \ldots, X_n$ di legge uniforme su $[a - b, a + b]$, $b > 0$. Trovare lo stimatore di massima verosimiglianza per (a, b), studiarne la correttezza, la consistenza e calcolarne il rischio quadratico (medio).

Def: Stimatore di massima verosimiglianza per un parametro λ: dato un campione iid $\underbrace{\{X_1 ... X_n\}}_{\underline{X}}$ con distribuzione $X_i \sim f_\lambda(x)$

$$\hat{\lambda}(\underline{X}) = \arg\max_{\lambda} \mathcal{L}_\lambda(\underline{X})$$

con $\mathcal{L}$ funzione di verosimiglianza

$$\mathcal{L}_\lambda(\underline{X}) = f_\lambda(\underline{X}) \overset{\text{iid}}{=} \prod_{j=1}^{n} f_\lambda(x_j)$$

Def: $\hat{\lambda}$ è corretto se $\mathbb{E}_\lambda(\hat{\lambda}) - \lambda = 0$

$\hat{\lambda}$ è consistente se $\lim_{n \to \infty} \mathbb{P}(|\hat{\lambda}(\underline{X}_n) - \lambda| > \varepsilon) = 0$

Def: Errore quadratico medio

$$\mathbb{E}_\lambda((\hat{\lambda}(\underline{X}) - \lambda)^2)$$

$$\mathbb{E}((\lambda - \hat{\lambda})^2) = \mathbb{E}(((\lambda - \mathbb{E}(\hat{\lambda})) - (\hat{\lambda} - \mathbb{E}(\hat{\lambda})))^2) =$$

$$= \mathbb{E}(\underbrace{(\lambda - \mathbb{E}(\hat{\lambda}))^2}_{\in \mathbb{R}} + (\hat{\lambda} - \mathbb{E}(\hat{\lambda}))^2 - 2\underbrace{(\lambda - \mathbb{E}(\hat{\lambda}))}_{\in \mathbb{R}}(\hat{\lambda} - \mathbb{E}(\hat{\lambda})))$$

$$= \underbrace{(\lambda - \mathbb{E}(\hat{\lambda}))^2}_{Bias^2} + \underbrace{\mathbb{E}((\hat{\lambda} - \mathbb{E}(\hat{\lambda}))^2)}_{Var(\hat{\lambda})} - \cancelto{0}{2(\lambda - \mathbb{E}(\hat{\lambda}))\mathbb{E}\underbrace{(\hat{\lambda} - \mathbb{E}(\hat{\lambda}))}_{\mathbb{E}(\hat{\lambda}) - \mathbb{E}(\hat{\lambda})}}$$

1. * Data $X \sim N(\mu, \sigma^2)$, si dimostri che $aX + b \sim N(a\mu + b, a^2\sigma^2)$ se $a \neq 0$.
2. * Data $Z \sim \mathcal{N}(0, 1)$, si calcoli la distribuzione di Z^2.
3. * Data $X \sim U((0, 1))$, si stabilisca per quali α la variabile aleatoria $Y = (1 - X)^{-\alpha} \in L^1$.
4. Siano $X_1, \ldots, X_n$ variabili aleatorie iid con $X_i \sim Exp(\lambda)$. Si calcoli la legge di $Z_n = \sum_{i=1}^{n} X_i$. Si tratta di una legge conosciuta?

1) Sappiamo che per ogni $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}_+$ se $Z \sim N(0,1)$ allora $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$

$$\Longrightarrow Y = aX + b = a(\mu + \sigma Z) + b = \underbrace{(a\mu + b)}_{\mu'} + \underbrace{a\sigma}_{\sigma'} Z$$

$$\longrightarrow aX + b \sim N(a\mu + b, (a\sigma)^2)$$

2) $F_{Z^2}(z) = \mathbb{P}(Z^2 \leq z) = \mathbb{P}(-\sqrt{z} \leq Z \leq \sqrt{z}) = \Phi(\sqrt{z}) - \Phi(-\sqrt{z})$

$$= 2\Phi(\sqrt{z}) - 1$$

$$\Longrightarrow f_{Z^2}(z) = \frac{d}{dz} F_{Z^2}(z) = 2\frac{d}{dz} \int_{-\infty}^{\sqrt{z}} \phi(y)\,dy = 2\,\phi(\sqrt{z}) \cdot \frac{1}{2\sqrt{z}}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-z/2} \cdot \frac{1}{\sqrt{z}}$$

3) $f_Y(y)\,dy = f_X(x(y)) \left|\frac{dx}{dy}\right| dy = f_X(x(y)) \left|\left(\frac{dy}{dx}\right)^{-1}\right| dy$

$y = (1-x)^{-\alpha} \Longrightarrow 1 - x = y^{-1/\alpha} \Longrightarrow x(y) = 1 - y^{-1/\alpha}$, $\frac{d}{dx} y(x) = \alpha(1-x)^{-(\alpha+1)}$

$$\text{range}(Y) = \begin{cases} [0,1] & \alpha > 0 \\ \{1\} & \alpha = 0 \\ [1, \infty) & \alpha < 0 \end{cases}$$

$-\frac{\alpha+1}{\alpha}$

$$\Longrightarrow f_Y(y) = \begin{cases} \mathbb{1}_{[0,1]}(y)\frac{1}{\alpha}(1-x)^{+(\alpha+1)} = \mathbb{1}_{[0,1]}(y)\frac{1}{\alpha}y^{-(1+\frac{1}{\alpha})} & \alpha > 0 \\ \delta(y-1) & \alpha = 0 \\ \mathbb{1}_{[1,\infty)}(y)\frac{1}{|\alpha|}y^{-(1+\frac{1}{\alpha})} & \alpha < 0 \end{cases}$$

$$\int_{-\infty}^{\infty} f_Y(y)\,dy = \begin{cases} \int_0^1 \cancel{y}\frac{1}{\alpha}y^{-(\cancel{1}+\frac{1}{\alpha})} < \infty & \text{se } \alpha > 1 \\ \int_{-\infty}^{\infty} y\,\delta(y-1)\,dy = 1 & \text{se } \alpha = 0 \\ \int_1^{\infty} \cancel{y}\frac{1}{|\alpha|}y^{-(\cancel{1}+\frac{1}{\alpha})}\,dy = \infty & \forall\ \alpha < 0 \end{cases}$$

$Y \in L^1$ solo se $\alpha \in \{0\} \cup (1,\infty)$

4) Dimostriamo per induzione che $Y_k = \sum_{j=1}^{k} X_j \sim \Gamma(k,\lambda)$

$\underline{k=1}$: $X_1 \sim \mathrm{Geom}(\lambda) = \Gamma(1,\lambda)$

$\underline{k \to k+1}$: $Y_{k+1} = Y_k + X_{k+1}$

$$f_{Y_{k+1}}(t) = \int_{-\infty}^{\infty} \lambda e^{-\lambda(t-x)}\mathbb{1}_{[0,\infty)}(t-x)\frac{\lambda^k}{(k-1)!}x^{k-1}e^{-\lambda x}\mathbb{1}_{[0,\infty)}(x)\,dx$$

$$= \int_{-\infty}^{\infty} \lambda e^{-\lambda(t-\cancel{x})}\mathbb{1}_{(-\infty,t]}(x)\frac{\lambda^k}{(k-1)!}x^{k-1}\cancel{e^{-\lambda x}}\mathbb{1}_{[0,\infty)}(x)\,dx$$

$$= e^{-\lambda t}\lambda^{k+1}\cdot\frac{1}{(k-1)!}\int_0^t x^{k-1}\,dx = \lambda^{k+1}\frac{1}{k!}t^k e^{-\lambda t}$$

$\Longrightarrow \square$

7. * Siano (X, Y) le coordinate aleatorie di un punto nel piano, con X e Y due variabili aleatorie Gaussiane standard e indipendenti. Siano (R, θ) le coordinate polari del medesimo punto. Dopo aver determinato la distribuzione di R^2 e di θ, ottenuto la densità di R e aver verificato che R e θ sono indipendenti,

(a) calcolare il valore atteso di $\frac{X^2}{R^2}$

(b) calcolare il valore atteso del quoziente tra le variabili aleatorie $\min\{|X|, |Y|\}$ e $\max\{|X|, |Y|\}$.

$$f_{XY}(x,y) = f_X(x) \cdot f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$$

$$= \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$$

$$\frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} dx\,dy = \underbrace{\frac{1}{2\pi} e^{-\frac{r^2}{2}} r\,dr\,d\theta}_{f_{R,\Theta}(r,\theta) = f_R(r) \cdot f_\Theta(\theta)}$$

$$f_R(r) \to e^{-\frac{r^2}{2}} \cdot r \qquad f_\Theta(\theta) \to \frac{1}{2\pi} \mathbb{1}_{[0,2\pi)}$$

a) $$\mathbb{E}\left(\frac{X^2}{R^2}\right) = \mathbb{E}\left(\frac{(R\cos(\Theta))^2}{R^2}\right) =$$

$$= \mathbb{E}(\cos(\Theta)^2) = \int_0^\infty e^{-\frac{r^2}{2}} r\,dr \; \frac{1}{2\pi}\int_0^{2\pi} \cos(\theta)^2 d\theta$$

$X = R\cos\Theta$

$Y = R\sin\Theta$

$$= 1 \cdot \frac{\frac{1}{2}(2\pi)}{2\pi} = \frac{\pi}{2\pi} = \frac{1}{2}$$

b) $$\mathbb{E}\left(\frac{\min |X|,|Y|}{\max |X|,|Y|}\right) = \mathbb{E}\left(\frac{\cancel{R} \min |\cos\Theta|, |\sin\Theta|}{\cancel{R} \max |\cos\Theta|, |\sin\Theta|}\right)$$

$$= \int_0^\infty e^{-\frac{r^2}{2}} r\,dr \cdot \frac{1}{2\pi} \int_0^{2\pi} \frac{\min|\cos\theta|, |\sin\theta|}{\max|\cos\theta|, |\sin\theta|} d\theta = \frac{4}{2\pi} \int_0^{\pi/2} \frac{\min}{\max} d\theta$$

$$= \frac{4}{2\pi}\left(\int_0^{\pi/4} \frac{\sin(\vartheta)}{\cos(\vartheta)}\, d\vartheta + \int_{\pi/4}^{\pi/2} \frac{\cos(\vartheta)}{\sin(\vartheta)}\, d\vartheta\right)$$

$$= \frac{4}{2\pi}\left(\int_0^{\pi/4} \tan(\vartheta)\, d\vartheta + \int_{\pi/4}^{\pi/2} \frac{\sin(\frac{\pi}{2}-\vartheta)}{\cos(\frac{\pi}{2}-\vartheta)}\, d\vartheta\right)$$

$$= \frac{4}{2\pi}\left(\left(-\log\cos(\vartheta)\right)\Big|_0^{\pi/4} + \int_0^{\pi/4} \tan(\vartheta)\, d\vartheta\right)$$

$$= \frac{4}{2\pi}\left(-\log\cos\frac{\pi}{4}\cdot 2\right) = \frac{8}{2\pi}\log\sqrt{2}$$

$$= \frac{2}{\pi}\log 2$$

10. Si consideri la densità di probabilità data da

$$f_X(x) = \frac{\theta}{x^2}\mathbb{1}_{[\theta,\infty)}(x),$$

con $\theta > 0$. Si proponga un opportuno stimatore per θ basato su un campione casuale $X_1, X_2, \ldots, X_n$ estratto da una popolazione di densità f_X .

(a) Stabilire se lo stimatore proposto è corretto.

(b) Stabilire se la successione di stimatori proposta, al variare della numerosità del campione è consistente.

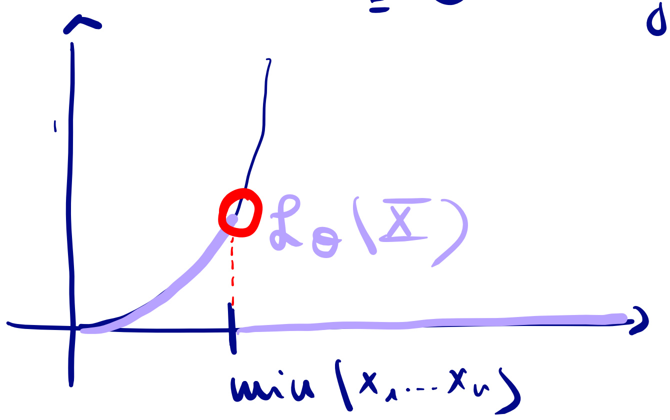
(c) Utilizzare lo stimatore trovato per determinare intervalli di fiducia al livello $1-\alpha$ per il parametro θ.

z) $\mathcal{L}_\theta(\overline{X}) = f_\theta(\overline{X}) = \prod_{j=1}^{n} f_\theta(x_i)$

$$= \prod_{j=1}^{n} \frac{\theta}{x_j^2} \mathbb{1}_{x_j \geq \theta} = \theta^n \left(\prod_{j=1}^{n} \frac{1}{x_j^2}\right)\left(\prod_{j=1} \mathbb{1}_{x_j \geq \theta}\right)$$

$$= \theta^n \prod_{j=1}^{n} \frac{1}{x_j^2} \cdot \mathbb{1}_{\cap\{x_j \geq \theta\}}$$

$$= \theta^n \prod x_j^{-2} \cdot \mathbb{1}_{\theta \leq \min(x_1 \ldots x_n)}$$



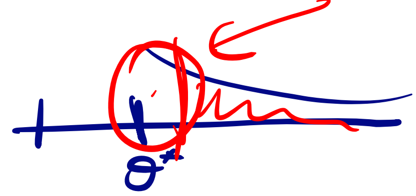
$\Longrightarrow \hat{\theta}_n(\overline{X}) = \min(x_1 \ldots x_i)$

a) $\mathbb{E}(\min(x_1)) = \mathbb{E}(x_1) = \int_{-\infty}^{\infty} \cancel{x}\, \theta \frac{1}{x^{\cancel{2}}} \mathbb{1}_{x \geq \theta}$

$$= \theta \int_{\theta}^{\infty} \frac{1}{x}\, dx = \theta(\log \infty - \log(\theta)) = \infty$$

$\longrightarrow$ non è corretto.

b) Sia $X_i \overset{iid}{\sim} f_{\theta^*}(x)$



$$\mathbb{P}\left(|\min(X_i) - \theta^*| > \varepsilon\right) = \mathbb{P}\left(\min(X_i) - \theta^* > \varepsilon\right)$$

$$= \mathbb{P}\left(\min X_i > \theta^* + \varepsilon\right) = \mathbb{P}\left(\bigcap \{X_i > \theta^* + \varepsilon\}\right)$$

$$= \prod_{i=1}^{n} \mathbb{P}\left(X_i \geq \theta^* + \varepsilon\right) = \left(\mathbb{P}\left(X_i \geq \theta^* + \varepsilon\right)\right)^n$$

$$= \left(\int_{\theta^*+\varepsilon}^{\infty} \frac{\theta^*}{x^2}\, dx\right)^n = \left(\frac{1}{1+\frac{\varepsilon}{\theta^*}}\right)^n \xrightarrow{n \to \infty} 0$$

11. * Sia dato un campione $X_1, \dots, X_n$ di legge uniforme su $(0, \theta)$, $\theta > 0$. Siano

$$\hat{\theta}_n = \max(X_1, \dots, X_n), \qquad \bar{\theta}_n = \frac{2}{n}\sum_{i=1}^{n} X_i$$

(a) $\hat{\theta}_n$ è una stima corretta di θ? è asintoticamente corretta?

(b) Calcolare il rischio quadratico (medio) della stima $\hat{\theta}_n$ di θ $R(\theta, \hat{\theta}_n) = \mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2]$.

(c) è vero che $\lim_{n\to\infty} R(\theta, \hat{\theta}_n) = 0$? La successione di stime $(\hat{\theta}_n)_n$ è consistente?

(d) Dimostrare che $\hat{\theta}_n$ coincide con la stima di massima verosimiglianza per θ.

(e) La stima $\bar{\theta}_n$ di θ è corretta? è asintoticamente corretta?

(f) Calcolare il rischio quadratico (medio) della stima $\bar{\theta}_n$ di θ $R(\theta, \bar{\theta}_n) = \mathbb{E}_\theta[(\bar{\theta}_n - \theta)^2]$

(g) è vero che $\lim_{n\to\infty} R(\theta, \bar{\theta}_n) = 0$? Tale stima $\bar{\theta}_n$ è consistente?

(h) $\hat{\theta}_n$ è preferibile rispetto a $\bar{\theta}_n$?

a) Sappiamo che per $U = \max(X_1, \dots X_n)$, $X_i \sim unif((0,\theta))$

$$f_U(u|\theta) = n\left(\frac{u}{\theta}\right)^{n-1}\frac{1}{\theta}\mathbb{1}_{[0,\theta]}(u) \qquad \frac{u}{\theta} = t$$

$$\Longrightarrow \mathbb{E}_\theta(\hat{\theta}_n) = \mathbb{E}_\theta(U) = \int_0^\theta n\left(\frac{u}{\theta}\right)^n du = n\theta\int_0^1 t^n dt$$

$$= \frac{n}{n-1}\theta \neq \theta \Longrightarrow \text{stimatore non è corretto}$$

$$\lim_{n\to\infty}\frac{n}{n-1}\theta = \theta \Longrightarrow \text{stimatore asintoticamente corretto.}$$

b) $$\mathbb{E}_\theta((\theta - \hat{\theta}_n)^2) = \mathbb{E}_\theta((\theta - U)^2) = \theta^2 - 2\theta\,\mathbb{E}_\theta(U) + \mathbb{E}_\theta(U^2)$$

$$= \theta^2 - 2\theta\frac{n}{n-1}\theta + \int_0^\theta n\left(\frac{u}{\theta}\right)^{n-1}\frac{1}{\theta}u^2 du =$$

$$= \theta^2\left(1 - \frac{2n}{n-1}\right) + \theta^2\int_0^1 n\,t^{n+1}dt$$

$$= \theta^2\left(1 - \frac{2n}{n-1}\right) + \theta^2\frac{n}{n} = 2\theta^2\frac{1}{n-1} = \frac{2\theta^2}{n-1}$$

c) $\lim_{n\to\infty} \frac{2\vartheta^2}{n-1} = 0 \implies$ la successione di stime è consistente per Markov:

$$\mathbb{P}(|\hat{\vartheta}_n - \vartheta| > \varepsilon) = \mathbb{P}((\hat{\vartheta}_n - \vartheta)^2 > \varepsilon^2) \leq \frac{\mathbb{E}((\hat{\vartheta}_n - \vartheta)^2)}{\varepsilon^2} = \frac{\frac{2\vartheta^2}{n-1}}{\varepsilon^2}$$

$$\implies \lim_{n\to\infty} \mathbb{P}(|\hat{\vartheta}_n - \vartheta| > \varepsilon) = 0$$

d) vedi note corso.

e) $\mathbb{E}_\vartheta(\bar{\vartheta}_n) = \mathbb{E}_\vartheta\left(\frac{2}{n}\sum_{j=1}^n X_j\right) = \frac{2}{n}\sum_{j=1}^n \mathbb{E}_\vartheta(X_j) = \frac{2}{n}\sum_{j=1}^n \frac{\vartheta}{2} = \vartheta$

$\implies$ stimatore corretto e dunque asintoticamente corretto

f) $\mathbb{E}_\vartheta((\vartheta - \bar{\vartheta}_n)^2) = \vartheta^2 - 2\vartheta\,\mathbb{E}_\vartheta(\bar{\vartheta}_n) + \mathbb{E}_\vartheta(\bar{\vartheta}_n^2) =$

$$= \frac{4}{n^2}\mathbb{E}_\vartheta\left(\sum_j X_j^2 + \sum_{i\neq j} X_j X_i\right) - \vartheta^2$$

$$= \frac{4}{n^2}\left(n\,\mathbb{E}_\vartheta(X_1^2) + (n^2 - n)\,\mathbb{E}_\vartheta(X_1)^2\right) - \vartheta^2$$

$$= \frac{4}{n^2}\left(n\,\mathrm{Var}_\vartheta(X_1^2) + n^2\left(\frac{\vartheta}{2}\right)^2\right) - \vartheta^2$$

$$= \frac{1}{n}\cdot\frac{4}{12}\vartheta^2 = \frac{\vartheta^2}{3n}$$

g) Come sopra $\lim_{n \to \infty} \mathbb{E}((\vartheta - \bar{\vartheta}_n)^2) = 0$

$\implies$ consistente.

h) Notiamo che $R_n(\vartheta, \bar{\vartheta}_n) \leq R_n(\vartheta, \hat{\vartheta}_n)$

quindi $\hat{\vartheta}_n$ non è preferibile rispetto a $\bar{\vartheta}_n$

9. * Sia (X,Y) una variabile aleatoria su $\mathbb{R}^2$ con densità continua

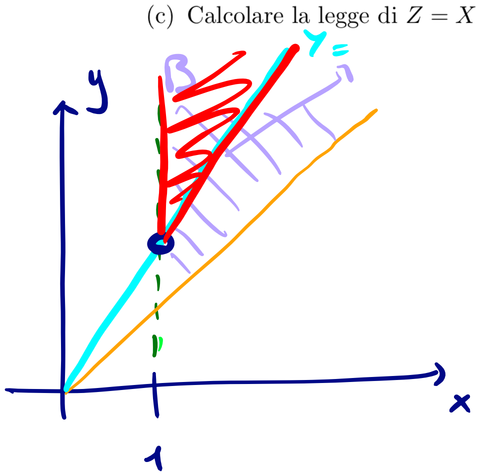
$$f_{X,Y}(x,y) = \frac{3}{x^3y^2}\mathbb{1}_{\{x>1,y>x\}}$$

(a) Calcolare $P[Y<2X]$.

(b) Calcolare la legge congiunta di $U=\log X$ e $V = X/Y$ e le relative densità marginali.

(c) Calcolare la legge di $Z = X+Y$.

a)



$$\mathbb{P}(Y<2X) = 1-\mathbb{P}(Y\geq 2X)$$

$$=1-\int_1^\infty\int^\infty \frac{3}{x^7y^2}\,dy\,dx =$$

$$=1-\int_1^\infty \frac{\cancel{3}}{x^3}\underbrace{\int_{2x}^\infty \frac{1}{y^2}\,dy}_{\frac{1}{\cancel{3}}\frac{1}{(2x)^3}}\,dx = 1-\int_1^\infty \frac{1}{8\cdot x^6}\,dx = \frac{1}{8\cdot 7}\underbrace{\left[\frac{1}{x^7}\right]_1^\infty}_{1}$$

$$=1-\frac{1}{56} = \frac{55}{56}$$

Euristica: $f_{UV}(u,v)\,du\,dv = f_{XY}(x(u,v),y(u,v))\,|\det J|\,du\,dv = f_{XY}(x,y)\,dx\,dy$

$$J = \partial_{(u,v)}\begin{pmatrix}x(u,v)\\ y(u,v)\end{pmatrix}$$

b) $(x(u,v),y(u,v)) = \left(\exp(u),\frac{x}{v}\right) = \left(\exp(u),\frac{\exp(u)}{v}\right)$

$$f_{U,V}(u,v) = f_{X,Y}\left(\exp(u),\frac{\exp(u)}{v}\right)\left|\det\begin{pmatrix}\exp(u) & 0\\ \frac{\exp(u)}{v} & -\frac{1}{v^2}\exp(u)\end{pmatrix}\right|$$

$$=\frac{3}{(e^u)^3\cancel{\left(\frac{1}{v}e^u\right)^2}}\mathbb{1}\left(e^u>1,\cancel{\frac{e^u}{v}}>\cancel{e^u}\right)\cancel{e^u\cdot\frac{1}{v^2}e^u}$$

$$=3e^{-3u}\cdot\mathbb{1}_{(u>0,\ v\in(0,1))}$$

$$f_U(u) = \int_{-\infty}^\infty f_{UV}(u,v)\,dv = 3e^{-3u}\mathbb{1}_{u>0}\int_{-\infty}^\infty \mathbb{1}_{(0,1)}(v)\,dv = 3e^{-3u}\mathbb{1}_{u>0}$$

$$f_V(v) = \int_{-\infty}^{\infty} f_{UV}(u,v)\,du = \mathbb{1}_{v\in(0,1)} \int_0^{\infty} 3e^{-3u} = \mathbb{1}_{v\in(0,1)}$$

c) $$f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(x, z-x)\,dx = \int_{-\infty}^{\infty} \frac{3}{x^3(z-x)^2}\,\mathbb{1}_{x>1,\ \underbrace{z-x>x}_{x<\frac{z}{2}}}\,dx$$

$$= \int_1^{z/2} \frac{3}{x^3(z-x)^2}\,dx = 3\left(z - \frac{2}{z-1} - \frac{8}{z} + 4 + 6\log\left(1-\frac{z}{2}\right)\right)$$

6. Siano X e Y variabili aleatorie indipendenti tali che $X \sim Exp(\lambda)$ e $Y \sim Exp(\mu)$. Calcolare la densità e la funzione di ripartizione di $Z = X + Y$ e $W = \alpha X$, $\alpha \in \mathbb{R}$. Si tratta di leggi note?

a) $$f_{X+Y}(z) = \int_{-\infty}^{\infty} \lambda e^{-\lambda w} \mathbb{1}_{w \geq 0} \cdot \mu e^{-\mu(z-w)} \mathbb{1}_{\underbrace{z-w \geq 0}_{w \leq z}} \, dw$$

$$= \begin{cases} 0 & \text{se } z \leq 0 \\ \lambda\mu \int_0^z e^{-\lambda w} \cdot e^{-\mu \cdot z} e^{\mu w} dw = \lambda\mu \, e^{-\mu z} \int_0^z e^{-(\lambda-\mu)w} dw = \dfrac{\lambda\mu}{\mu-\lambda}\left(e^{-z\lambda} - e^{-z\mu}\right) & z \geq 0 \end{cases}$$

b) $W = \alpha X \quad g(x) = \alpha x \quad\quad g^{-1}(w) = \frac{1}{\alpha} \cdot w \quad\quad$ assume $\alpha \geq 0$

$$f_W(w) = f_X(g^{-1}(w)) \cdot \left|\frac{d}{dw} g^{-1}(w)\right| = \lambda \cdot e^{-\lambda \cdot \frac{1}{\alpha} w} \cdot \left|\frac{1}{\alpha}\right| = \begin{cases} \dfrac{\lambda}{|\alpha|} e^{-\frac{\lambda}{\alpha} w} & w \geq 0 \\ 0 & w < 0 \end{cases}$$

5. Una variabile aleatoria X è detta di Cauchy se ha densità

$$p(x) = \frac{1}{\pi(1+x^2)}$$

(a) Verificare che p è una densità di probabilità e scrivere esplicitamente la funzione di ripartizione. X ammette media? E varianza?

(b) Calcolare la legge di $Y = X^2$. Y ha media?

(c) Dopo aver verificato che la variabile aleatoria $Z = 1/X$ è ben posta, calcolare la legge di Z. Z ha media? E varianza?

a) $\int_{-\infty}^{\infty} \frac{1}{\pi} \cdot \frac{1}{1+x^2}\, dx = \frac{1}{\pi} \left(\arctan(x) \right)_{-\infty}^{\infty} = \frac{1}{\pi}\left(\frac{\pi}{2} + \frac{\pi}{2} \right) = 1$ ✓

⚠ per definire valore atteso $\int_{-\infty}^{\infty} |x|\, p(x)\, dx < \infty$

$\int_{-\infty}^{\infty} |x| \frac{1}{\pi(1+x^2)}\, dx = \frac{2}{\pi} \int_0^{\infty} \frac{x}{(1+x^2)}\, dx \overset{y=1+x^2,\ dy = 2x\,dx}{=} \frac{1}{\pi} \int_1^{\infty} \frac{1}{y}\, dy \cancel{<\infty}$

$\Longrightarrow \mathbb{E}(X)$ non def. $\mathrm{Var}(X)$ non def.

b) $g(x) = x^2$ $\quad g^{-1}(y) = \pm\sqrt{y}$ $\quad g_1^{-1}(y) = -\sqrt{y}$ $\quad g_2^{-1}(y) = \sqrt{y}$

$f_Y(y) = f_X(\sqrt{y}) \cdot \left|\frac{d}{dy}\sqrt{y}\right| + f_X(-\sqrt{y}) \cdot \left|\frac{d}{dy}(-\sqrt{y})\right| =$

$= \frac{1}{\pi}\left(\frac{1}{1+\sqrt{y}^2} \cdot \frac{1}{2\sqrt{y}} + \frac{1}{1+(-\sqrt{y})^2} \frac{1}{2\sqrt{y}} \right) = \frac{1}{\pi} \frac{1}{1+y} \cdot \frac{1}{\sqrt{y}}$

$f_Y(y) = \begin{cases} \frac{1}{\pi} \frac{1}{1+y} \cdot \frac{1}{\sqrt{y}} & y > 0 \\ 0 & \text{altrimenti} \end{cases}$

$\int_0^{\infty} y \cdot \frac{1}{\pi} \frac{1}{1+y} \frac{1}{\sqrt{y}}\, dy = \infty$ $\quad \not\exists$

c) $z = \frac{1}{x} \iff x = \frac{1}{z} \longrightarrow f_Z(z) = f_X\left(\frac{1}{z}\right) \cdot \left| \frac{d}{dz} \frac{1}{z} \right|$ $\quad z \neq 0$

$= \frac{1}{\pi} \frac{1}{1+\frac{1}{z^2}} \cdot \left|\frac{1}{z^2}\right| = \frac{1}{1+z^2} \frac{1}{\pi}$

$\Longrightarrow z \sim$ Cauchy $\Rightarrow \not\exists\ \mathbb{E}(z)\ \mathrm{Var}(z)$

8. Fissato $\mu \in \mathbb{R}$, sia
$$p(x) = cx^{\mu}\mathbb{1}_{\{x\in(0,1)\}}.$$

(a) Determinare μ affinché p sia integrabile su $\mathbb{R}$.

(b) Per tali valori di μ, trovare c affinché p sia una densità di probabilità.

(c) Sia X una variabile aleatoria con densità data da p con $\mu = 1$. Calcolare la densità della variabile aleatoria $X^2 + 1$.

(d) Siano X e Y due variabili aleatorie indipendenti, X con densità p con $\mu = 1$ e Y esponenziale di parametro 1. Calcolare la legge di $X + Y$.

Richiamo: funzione di densità $f_X(x)$ $[p(x)]$ soddisfa
$$\mathbb{P}(X\in(a,b)) = \int_a^b f_X(x)\,dx \quad \forall a<b \in \mathbb{R}$$
$$\Longrightarrow f_X(x) \geq 0 \quad \forall x \qquad \Longrightarrow \int_{-\infty}^{\infty} f_X(x)\,dx = 1$$

a) $$\int_{-\infty}^{\infty} p(x)\,dx = \int_{-\infty}^{\infty} c\cdot x^{\mu}\mathbb{1}_{(0,1)}(x)\,dx = c\int_0^1 x^{\mu}\,dx = \begin{cases}\left[c\,\dfrac{x^{\mu+1}}{\mu+1}\right]_0^1 & \mu \neq -1\\ \left[c\log(x)\right]_0^1 & \mu = -1\end{cases}$$
$$\Longrightarrow \mu > -1$$

b) $$1 \stackrel{!}{=} \int_{-\infty}^{\infty} p(x)\,dx = \left[c\cdot\frac{x^{\mu+1}}{\mu+1}\right]_0^1 = c\cdot\frac{1}{1+\mu} \Longrightarrow c = 1+\mu$$

Richiamo: $Y = g(X)$

1) se g è invertibile $\Longrightarrow f_Y(y) = f_X(g^{-1}(y))\left|\dfrac{dg^{-1}(y)}{dy}\right|$

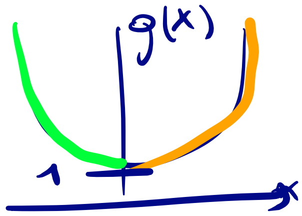
2) se g non è invertibile (su $\mathbb{R}$)

$\Longrightarrow$ scomponiamo g in $\{g_i\}^m$ $g_i : A_i \to \mathbb{R}$

$\bigcup_{i=1}^m A_i = \mathbb{R}$ tali che g_i sia invertibile.

$$\Longrightarrow f_Y(y) = \sum_{i=1}^m f_X(g_i^{-1}(y))\left|\frac{dg^{-1}(y)}{dy}\right|$$

Ex :



$g_1(x): \mathbb{R}_- \to \mathbb{R}$

$g_2(x): \mathbb{R}_+ \to \mathbb{R}$

$f_Y(x) = f_X(g_1^{-1}) \dots + f_X(g_2^{-1}) \dots$

c) $\mu = 1 \longrightarrow p(x) = 2 \cdot x \cdot \mathbb{1}_{(0,1)}(x).$

$$g(x) = x^2 + 1 \qquad g^{-1}(y) = \pm\sqrt{y-1} \longrightarrow \begin{cases} g_1^{-1}(y) = -\sqrt{y-1} \\ g_2^{-1}(y) = \sqrt{y-1} \end{cases}$$

$$f_Y(y) = f_X(-\sqrt{y-1}) \cdot \left|\frac{1}{2\sqrt{y-1}}\right| + f_X(\sqrt{y-1}) \cdot \left|\frac{1}{2\sqrt{y-1}}\right|$$

$$= 0 + \frac{\cancel{2\sqrt{y-1}} \cdot \mathbb{1}_{(0,1)}(\sqrt{y-1})}{\cancel{2\sqrt{y-1}}} = \mathbb{1}_{(1,2)}(y)$$

d) $f_X(x) = 2x\, \mathbb{1}_{(0,1)}(x)$

$f_Y(y) = e^{-y}\, \mathbb{1}_{y>0}$

Richiamo: $f_{X+Y}(z) = f_X * f_Y(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(z-w)\, dw$

$X ⫫ Y$

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} 2w\, \mathbb{1}_{(0,1)}(w) \cdot e^{-(z-w)} \cdot \mathbb{1}_{z-w>0}\, dw$$

$$= \int_0^1 2w e^{-(z-w)}\, \mathbb{1}_{z>w}(w)\, dw = 2e^{-z} \int_0^1 w e^w\, \mathbb{1}_{w<z}(w)\, dw$$

$$= \begin{cases} 2e^{-z} \int_0^z w e^w\, dw = 2e^{-z}\left(e^w(w-1)\right)\Big|_0^z & z < 1 \\ 2e^{-z}\left(e^w(w-1)\right)\Big|_0^1 & z \geq 1 \end{cases}$$

12. Si consideri un campione $X_1, \dots, X_n$ di legge uniforme su $[a-b, a+b]$, $b > 0$. Trovare lo stimatore di massima verosimiglianza per (a, b), studiarne la correttezza, la consistenza e calcolarne il rischio quadratico (medio).

$\underline{X} = \{X_1 \dots X_n\}$ $\qquad X_i \sim Unif(a-b, a+b)$

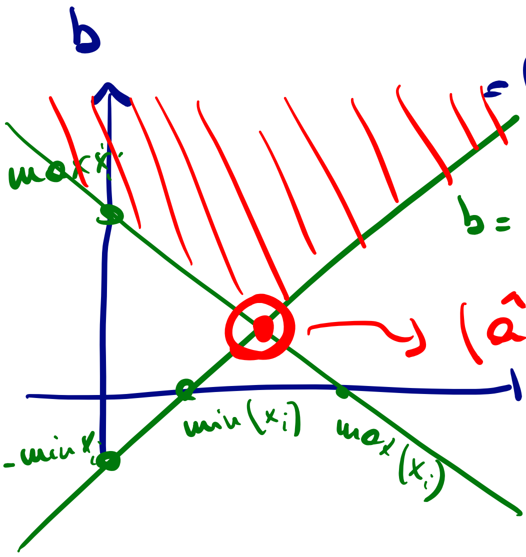
$$\mathcal{L}_{(a,b)}(\underline{X}) \overset{indep}{=} \prod_{i=1}^{n} \frac{1}{2b} \cdot \mathbb{1}_{x_i \in (a-b, a+b)}$$

$$= \prod_{i=1}^{n} \frac{1}{2b} \mathbb{1}_{x_i > a-b} \cdot \mathbb{1}_{x_i < a+b}$$

$$= \left(\frac{1}{2b}\right)^n \cdot \prod_{i} \mathbb{1}_{x_i > a-b} \cdot \prod_{i=1}^{n} \mathbb{1}_{x_i < a+b}$$

$$= \left(\frac{1}{2b}\right)^n \cdot \mathbb{1}_{\cap\{x_i > a-b\}} \, \mathbb{1}_{\cap\{x_i < a+b\}}$$

$$= \left(\frac{1}{2b}\right)^n \mathbb{1}_{\underbrace{\min_i x_i > a-b}_{b > a - \min(x_i)}} \cdot \mathbb{1}_{\underbrace{\max_i(x_i) < a+b}_{b > \max(x_i) - a}}$$



$b = a - \min(x_i)$

$\to (\hat{a}, \hat{b})$

$b = -a + \max(x_i)$

$$\Longrightarrow \begin{cases} \hat{a} = \dfrac{\min(x_i) + \max(x_i)}{2} \\[2ex] \hat{b} = \dfrac{\max(x_i) - \min(x_i)}{2} \end{cases}$$

$X_1 \sim unif(-2, 2)$

$$\Longrightarrow \begin{cases} \mathbb{E}\left(\dfrac{\max(X_1) - \min(X_1)}{2}\right) = \mathbb{E}\left(\dfrac{X_1 - X_1}{2}\right) = 0 \neq 2 = b \\[2ex] \mathbb{E}\left(\dfrac{\max(X_1) + \min(X_1)}{2}\right) = \mathbb{E}\left(\dfrac{X_1 + X_1}{2}\right) = \mathbb{E}(X_1) = 0. \end{cases}$$

$\Longrightarrow$ b non è corretto. $\qquad$ a è corretto.

• Errore quadratico medio.

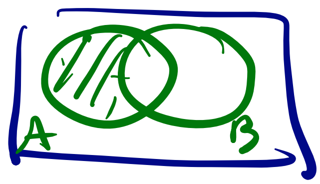
$$\mathbb{E}\left(\left(b - \frac{\min(x_i) - \max(x_i)}{2}\right)^2\right) = b^2 + \mathbb{E}\left(\left(\frac{\min(x_i) - \max(x_i)}{2}\right)^2\right) - 2b\,\mathbb{E}\left(\frac{\min(x_i) - \max(x_i)}{2}\right)$$

densità congiunta di $\min(x_i), \max(x_i)$

$X_i \sim \mathrm{Unif}(a-b, a+b)$.

$$\mathbb{P}(\min X_i \le s, \max X_i \le t) = F_{\min,\max}(s,t)$$

$$= \mathbb{P}\Big(\underbrace{\cap\{X_i \le t\}}_{A} \setminus \underbrace{\cap\{X_i \ge s\}}_{B}\Big)$$



$$= \underbrace{\mathbb{P}\Big(\bigcap_{i=1}^{n}\{X_i \le t\}\Big)}_{A} - \underbrace{\mathbb{P}\Big(\bigcap_{i=1}^{n}\{s \le X_i \le t\}\Big)}_{B \cap A}$$

$$= \prod \mathbb{P}(X_i \le t) - \prod \mathbb{P}(s \le X_i \le t)$$

$$= \underbrace{\mathbb{P}(X_i \le t)^n}_{\left(\frac{t-(a-b)}{2b}\right)^n} - \underbrace{\mathbb{P}(s \le X_i \le t)^n}_{\left(\frac{t-s}{2b}\right)^n}$$

$$\Rightarrow f_{\min,\max}(s,t) = \partial_s \partial_t F_{\min,\max}(s,t) = \frac{n(n-1)}{(2b)^2} \cdot \left(\frac{t-s}{2b}\right)^{n-2}$$

$$\mathbb{E}\left((b-\hat{b})^2\right) = \frac{(n+1)(n+2) + n(n+2) - 2(n-1)(n+2)}{(n+1)(n+2)}\, b^2$$

$$= \frac{3+4}{(n+1)(n+2)}\, b^2 \xrightarrow{n \to \infty} 0$$

Foglio di esercizi 9

Discussione soluzioni: 16.05.2023

* Rappresenta gli esercizi prioritari

1. Sia dato $\{X_1, \dots, X_n\}$ un campione di legge (μ, σ^2). Per stimare μ si vuole usare

$$T_n = \frac{1}{2}\left(X_1 + \frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

   (a) Lo stimatore T_n è corretto?
   (b) T_n è consistente?

2. * Sia dato $\{X_1, \dots, X_n\}$ un campione di legge Unif$((-\theta, 4\theta))$. Calcolare lo stimatore di massima verosimiglianza di θ e discuterne la correttezza e la consistenza.

3. (continua dal foglio precedente) Si consideri la densità di probabilità data da

$$f_X(x) = \frac{\theta}{x^2}\mathbb{1}_{[\theta,\infty)}(x),$$

   con $\theta > 0$. Si proponga un opportuno stimatore per θ basato su un campione casuale $X_1, X_2, \dots, X_n$ estratto da una popolazione di densità f_X .

   (a) Stabilire se lo stimatore proposto è corretto.
   (b) Stabilire se la successione di stimatori proposta, al variare della numerosità del campione è consistente.
   (c) Utilizzare lo stimatore trovato per determinare intervalli di fiducia al livello $1-\alpha$ per il parametro θ.

4. * Sia dato un campione $X_1, \dots, X_n$ di legge Gaussiana di parametri $(\mu, \sigma^2) \in \mathbb{R}(0, \infty)$.

   (a) Determinare lo stimatore di massima verosimiglianza di μ con $\sigma^2 = 1$.
   (b) Sempre per $\sigma^2 = 1$, utilizzare il metodo della quantità pivotale per trovare un intervallo di fiducia al livello 0.95 per μ.
   (c) Determinare lo stimatore di massima verosimiglianza di σ^2 con μ noto.
   (d) Determinare lo stimatore di massima verosimiglianza per (μ, σ^2).

5. Una impresa di componenti elettronici decide di stimare il numero medio di piccole imperfezioni su un sistema assemblato. A tale scopo analizza 16 sistemi e rileva che il numero medio di imperfezioni ammonta a 32 e che la varianza campionaria risulta 8. Supponendo che le imperfezioni abbiano distribuzione normale, determinare l'intervallo di fiducia al 95% e al 99% per il numero medio di piccole imperfezioni su sistemi assemblati.

6. * Dato $n \geq 1$, sia $(X_1, X_2, \dots, X_n)$ un campione estratto da una popolazione di densità

$$f_\theta(x) = \theta \exp(x - \theta \exp(x)), \qquad \text{per } \theta > 0\,.$$

   (a) Determinare lo stimatore Θ_n di massima verosimiglianza di θ
   (b) Stabilire la correttezza e la consistenza della successione $(\Theta_n)_{n\geq 1}$
   (c) Dato $\alpha \in (0, 1)$, determinare una regione di fiducia al livello $1-\alpha$ per θ

7. Dato $n \geq 1$, sia $(X_1, X_2, \ldots, X_n)$ un campione estratto da una popolazione di densità
$$f_\theta(x) = \frac{1}{12\theta^5} x^9 e^{-x^2/\theta} \mathbb{1}_{(0,\infty)}(x),$$
dove $\mathbb{1}_{(0,\infty)}(x)$ rappresenta la funzione indicatrice sull'insieme $(0, \infty)$.
   (a) Determinare lo stimatore Θ_n di massima verosimiglianza di θ.
   (b) Stabilire la correttezza della successione $(\Theta_n)_{n\geq 1}$.
   (c) Dato $\alpha \in (0, 1)$, determinare una regione di fiducia al livello $1 - \alpha$ per θ.

8. Sia dato un campione $X_1, \ldots, X_n$ con comune funzione di ripartizione (sconosciuta) $F : \mathbb{R} \to [0, 1]$. Sia $u \in \mathbb{R}$ e definiamo, per $i = 1, \ldots, n$,
$$Y_i(u) = \mathbb{1}_{X_i \leq u}$$
Il campione $Y_1(u), \ldots, Y_n(u)$ ha legge Bernoulli di parametro $F(u) \in [0, 1]$. Sia
$$\hat{F}_n(u) := \frac{1}{n} \sum_{i=1}^{n} Y_i(u)$$
   (a) $\hat{F}_n(u)$ è una stima non distorta di $F(u)$? Asintoticamente distorta?
   (b) Calcolare il rischio quadratico (medio) $R(F(u), \hat{F}_n(u)) = \mathbb{E}_{F(u)}[(\hat{F}_n(u) - F(u))^2]$.
   (c) La successione $\{\hat{F}_n(u)\}_n$ è consistente?
   (d) Trovare un intervallo di fiducia al livello $1 - \alpha$, $0 < \alpha \ll 1$, per $F(u)$.

1. Sia dato $\{X_1, \ldots, X_n\}$ un campione di legge $\mathcal{N}(\mu, \sigma^2)$. Per stimare μ si vuole usare

$$T_n = \frac{1}{2}\left(X_1 + \frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

(a) Lo stimatore T_n è corretto?

(b) T_n è consistente?

a) Calcoliamo:

$$\mathbb{E}_{\mu,\sigma}(T_n) = \mathbb{E}_{\mu,\sigma}\left(\frac{1}{2}\left(X_1 + \frac{1}{n}\sum_{j=1}^{n} X_j\right)\right)$$

$$= \frac{1}{2}\left(\mathbb{E}_{\mu,\sigma}(X_1) + \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}(X_j)\right)$$

$$= \frac{1}{2}\left(\mu + \frac{1}{n}\cdot n\mu\right) = \mu \qquad \forall n \in \mathbb{N}$$

$\Rightarrow$ lo stimatore è corretto

b) $$\mathbb{P}\left(\left|\frac{1}{2}\left(X_1 + \frac{1}{n}\sum_{j=1}^{n} X_j\right) - \mu\right| > \varepsilon\right)$$

$$= \mathbb{P}\left(\left|\frac{1}{2}\left(\frac{n+1}{n}X_1 + \frac{1}{n}\sum_{j=2}^{n} X_j\right) - \mu\right| > \varepsilon\right)$$

$$\geq \mathbb{P}\left(\left|\frac{n+1}{n}X_1 - \mu\right| > 2\varepsilon, \left|\frac{1}{n}\sum_{j=2}^{n} X_j - \mu\right| < \varepsilon\right)$$

$$\geq \mathbb{P}\left(\frac{n+1}{n}X_1 - \mu > 2\varepsilon\right)\mathbb{P}\left(\left|\frac{1}{n}\sum_{j=2}^{n} X_j - \mu\right| < \varepsilon\right)$$

$$\geq \underbrace{\mathbb{P}\left(X_1 > (\mu + 2\varepsilon)\frac{n}{n+1}\right)}_{\geq \mathbb{P}(X_1 > \mu + 2\varepsilon) = 1 - \Phi\left(\frac{2\varepsilon}{\sigma}\right)}\underbrace{\mathbb{P}\left(\left|\frac{1}{n}\sum_{j=1}^{n} X_j - \mu\right| < \frac{\varepsilon}{2}\right)}_{\geq \frac{1}{2}}\underbrace{\mathbb{P}\left(\left|\frac{X_1}{n}\right| < \frac{\varepsilon}{2}\right)}_{\geq \frac{1}{2}}$$

per $n \gg 1$

$$\Rightarrow \mathbb{P}\left(\left|\frac{1}{2}\left(X_1 + \frac{1}{n}\sum_{j=1}^{n} X_j\right) - \mu\right| > \varepsilon\right) \geq \left(1 - \Phi\left(\frac{2\varepsilon}{\sigma}\right)\right)\frac{1}{4} \underset{n\to\infty}{\not\to} 0$$

$\Rightarrow$ non consistente.

2. * Sia dato $\{X_1, \dots, X_n\}$ un campione di legge Unif$((-\theta, 4\theta))$. Calcolare lo stimatore di massima verosimiglianza di θ e discuterne la correttezza e la consistenza.

Scriviamo la densità $f_\theta(x) = \frac{1}{5\theta} \mathbb{1}_{(-\theta,4\theta)}(x)$

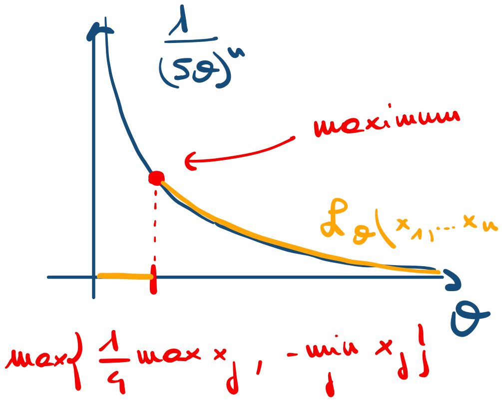
$$= \frac{1}{5\theta} \mathbb{1}_{(-\theta,\infty)}(x) \cdot \mathbb{1}_{(-\infty,4\theta)}(x)$$

$$\Longrightarrow \mathcal{L}_\theta(x_1, \dots, x_n) = \prod_{j=1}^{n} f_\theta(x_j) =$$

$$= \prod_{j=1}^{n} \frac{1}{5\theta} \mathbb{1}_{(-\infty,4\theta)}(x_j)\, \mathbb{1}_{(-\theta,\infty)}(x_j)$$

$$= \frac{1}{(5\theta)^n} \prod_{j=1}^{n} \mathbb{1}_{(-\infty,4\theta)}(x_j) \prod_{e=1}^{n} \mathbb{1}_{(-\theta,\infty)}(x_e)$$

$$= \frac{1}{(5\theta)^n} \mathbb{1}_{(\max_j x_j,\infty)}(4\theta)\, \mathbb{1}_{(-\min_j x_j,\infty)}(\theta)$$



$$= \frac{1}{(5\theta)^n} \underbrace{\mathbb{1}_{(\frac{1}{4}\max x_j,\infty)}(\theta)\, \mathbb{1}_{(-\min x_j,\infty)}(\theta)}_{\mathbb{1}_{(\max\{\frac{1}{4}\max x_j, -\min_j x_j\},\infty)}(\theta)}$$

$$\Longrightarrow \Theta_n = \max\{\underbrace{\tfrac{1}{4}\max x_j}_{\Theta_n'}, \underbrace{-\min x_j}_{\Theta_n''}\}$$

Corretto: Discutiamo la distribuzione di Θ_n:

$$F_{x_j}(x) = \frac{x+\theta}{5\theta} \mathbb{1}_{(-\theta,4\theta)}(x) + \mathbb{1}_{[4\theta,\infty)}(x)$$

$$\Longrightarrow F_{\max x_j} = \left(\frac{x+\theta}{5\theta}\right)^n \mathbb{1}_{(-\theta,4\theta)}(x) + \mathbb{1}_{[4\theta,\infty)}(x)$$

$$\Longrightarrow F_{\frac{1}{4}\max x_j}(x) = \left(\frac{4x+\vartheta}{5\vartheta}\right)^n \mathbb{1}_{\left(-\frac{\vartheta}{4},\vartheta\right)}(x) + \mathbb{1}_{[\vartheta,\infty)}(x)$$

$$\Longrightarrow F_{-\min x_j}(x) = P(-\min X_j \leq x) = P(\min X_j \geq -x)$$

$$= P(X_1 \geq -x)^n = (1 - F_{x_j}(-x))^n =$$

$$= \left(1 - \frac{-x+\vartheta}{5\vartheta}\right)^n \mathbb{1}_{(-\vartheta,4\vartheta)}(-x) + \mathbb{1}_{(-\infty,-\vartheta]}(-x)$$

$$= \left(\frac{4\vartheta+x}{5\vartheta}\right)^n \mathbb{1}_{(-4\vartheta,\vartheta)}(x) + \mathbb{1}_{[\vartheta,\infty)}(x)$$

$$\longrightarrow F_{\Theta_n}(x) = \left(\frac{4\vartheta+x}{5\vartheta}\right)^n \left(\frac{4x+\vartheta}{5\vartheta}\right)^n \mathbb{1}_{\left(-\frac{\vartheta}{4},\vartheta\right)}(x) + \mathbb{1}_{[\vartheta,\infty)}(x)$$

Notiamo che calcolare la densità associata ad F_{Θ_n} è calcoloso:

$$\Longrightarrow f_{\Theta_n}(x) = \mathbb{1}_{\left(-\frac{\vartheta}{4},\vartheta\right)}(x)\, n \left(\frac{4\vartheta+x}{5\vartheta}\right)^{n-1} \left(\frac{4x+\vartheta}{5\vartheta}\right)^{n-1} \left(\frac{1}{5\vartheta}\left(\frac{4x+\vartheta}{5\vartheta}\right) + \frac{4}{5\vartheta}\left(\frac{4\vartheta+x}{5\vartheta}\right)\right)$$

$$= \mathbb{1}_{\left(-\frac{\vartheta}{4},\vartheta\right)}(x)\, n \left(\frac{4\vartheta+x}{5\vartheta}\right)^{n-1} \left(\frac{4x+\vartheta}{5\vartheta}\right)^{n-1} \left(\frac{8x+17\vartheta}{5\vartheta}\right)$$

quindi ricorriamo a un "trucco":

$$\mathbb{E}_\theta(\Theta_n) = \int_0^\theta x\, f_{\Theta_n}(x)\, dx = \int_0^{\theta/2} x\, f_{\Theta_n}(x)\, dx + \int_{\theta/2}^{\theta} x\, f_{\Theta_n}(x)\, dx$$

$$\leq \frac{\theta}{2} F_{\Theta_n}\left(\frac{\theta}{2}\right) + \theta\left(F_{\Theta_n}(\theta) - F_{\Theta_n}\left(\frac{\theta}{2}\right)\right)$$

$$\begin{cases} F_{\Theta_n}\left(\frac{\theta}{2}\right) > 0 \\ F_{\Theta_n}(\theta) = 1 \end{cases} \quad < \theta\, F_{\Theta_n}(\theta) = \theta \implies \text{non corretto.}$$

<u>Consistenza</u>: $\forall \varepsilon > 0$

$$\mathbb{P}(|\Theta_n - \theta| \geq \varepsilon) = \mathbb{P}(\Theta_n \leq \theta - \varepsilon) = F_{\Theta_n}(\theta - \varepsilon)$$

$$= \left(1 - \frac{\varepsilon}{5\theta}\right)^n \left(1 - \frac{4\varepsilon}{5\theta}\right)^n \xrightarrow[n \to \infty]{} 0$$

$\implies$ consistente.

3. (continua dal foglio precedente) Si consideri la densità di probabilità data da

$$f_X(x) = \frac{\theta}{x^2} \mathbb{1}_{[\theta,\infty)}(x),$$

con $\theta > 0$. Si proponga un opportuno stimatore per θ basato su un campione casuale $X_1, X_2, \ldots, X_n$ estratto da una popolazione di densità f_X .

(a) Stabilire se lo stimatore proposto è corretto.

(b) Stabilire se la successione di stimatori proposta, al variare della numerosità del campione è consistente.

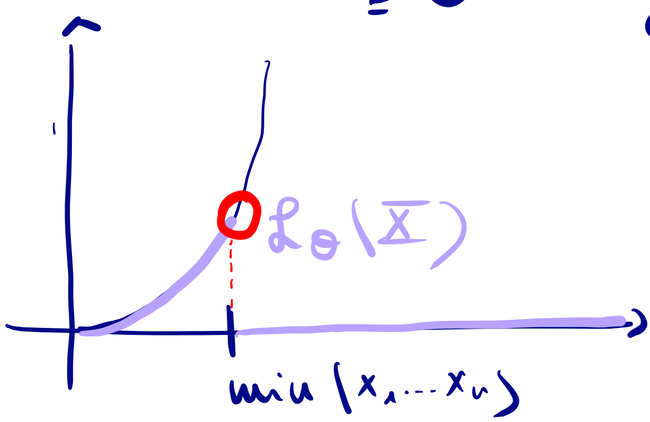
(c) Utilizzare lo stimatore trovato per determinare intervalli di fiducia al livello $1-\alpha$ per il parametro θ.

z) $\mathcal{L}_\theta(\underline{X}) = f_\theta(\underline{X}) = \prod_{j=1} f_\theta(x_i)$

$$= \prod_{j=1}^{n} \frac{\theta}{x_j^2} \mathbb{1}_{x_j \geq \theta} = \theta^n \left(\prod_{j=1}^{n} \frac{1}{x_j^2}\right)\left(\prod_{j=1}^{n} \mathbb{1}_{x_j \geq \theta}\right)$$

$$= \theta^n \prod_{j=1}^{n} \frac{1}{x_j^2} \cdot \mathbb{1}_{\cap\{x_j \geq \theta\}}$$

$$= \theta^n \prod x_j^{-2} \cdot \mathbb{1}_{\theta \leq \min(x_1 \cdots x_n)}$$



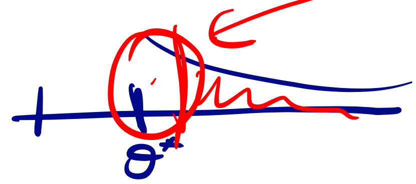
$$\Longrightarrow \hat{\theta}_n(\underline{X}) = \min(x_1 .. x_i)$$

a) $\mathbb{E}(\min(X_1)) = \mathbb{E}(X_1) = \int_{-\infty}^{\infty} \cancel{x}\,\theta \frac{1}{x^{\cancel{2}}} \mathbb{1}_{x \geq \theta}$

$$= \theta \int_{\theta}^{\infty} \frac{1}{x} dx = \theta(\log\infty - \log(\theta)) = \infty$$

$\longrightarrow$ non è corretto.

b) Sia $X_i \overset{iid}{\sim} f_{\theta^*}(x)$



$$\mathbb{P}(|\min(X_i) - \theta^*| > \varepsilon) = \mathbb{P}(\min(X_i) - \theta^* > \varepsilon)$$

$$= \mathbb{P}(\min X_i > \theta^* + \varepsilon) = \mathbb{P}(\cap \{X_i > \theta^* + \varepsilon\})$$

$$= \prod_{i=1}^{n} \mathbb{P}(X_i \geq \theta^* + \varepsilon) = (\mathbb{P}(X_i \geq \theta^* + \varepsilon))^n$$

$$= \left(\int_{\theta^*+\varepsilon}^{\infty} \frac{\theta^*}{x^2}\,dx\right)^n = \left(\frac{1}{1+\frac{\varepsilon}{\theta^*}}\right)^n \xrightarrow{n \to \infty} 0$$

c) Si consideri la distribuzione della variabile pivotale

$$\bar{\theta}_n = \frac{\min X_1 \dots X_n}{\theta} = \min \frac{X_1}{\theta} \dots \frac{X_n}{\theta}$$

con funzione di ripartizione

$$F_{\bar{\theta}_n}(y) = \mathbb{P}(\bar{\theta}_n < y) = 1 - \mathbb{P}(\bar{\theta}_n > y) =$$

$$= 1 - \prod_{i=1}^{n} \mathbb{P}\left(\frac{X_i}{\theta} > y\right) = 1 - \left(\int_{y}^{\infty} \frac{1}{x^2}\,dx\right)^n$$

$$= 1 - \frac{1}{y^n} \qquad y \geq 1$$

$$\hat{\theta}_n = \min X_1 \dots X_n$$

$$\Longrightarrow \quad 1 - \alpha = \mathbb{P}(\bar{\theta}_n < y_\alpha) = \mathbb{P}\left(\frac{\hat{\theta}_n}{\theta} < y_\alpha\right) = \mathbb{P}\left(\theta > \frac{\hat{\theta}_n}{y_\alpha}\right)$$

dove $1 - \frac{1}{y_\alpha^n} = 1 - \alpha \longleftrightarrow y_\alpha = \frac{1}{\sqrt[n]{\alpha}}$

4. * Sia dato un campione $X_1, \ldots, X_n$ di legge Gaussiana di parametri $(\mu, \sigma^2) \in \mathbb{R}(0, \infty)$.

(a) Determinare lo stimatore di massima verosimiglianza di μ con $\sigma^2 = 1$.

(b) Sempre per $\sigma^2 = 1$, utilizzare il metodo della quantità pivotale per trovare un intervallo di fiducia al livello 0.95 per μ.

(c) Determinare lo stimatore di massima verosimiglianza di σ^2 con μ noto.

(d) Determinare lo stimatore di massima verosimiglianza per (μ, σ^2).

$$f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\Longrightarrow \mathcal{L}_{\mu,\sigma}(\underline{X}) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(X_j-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}^n} \cdot \frac{1}{\sigma^n} e^{-\frac{\sum_j (X_j-\mu)^2}{2\sigma^2}}$$

$$\ell_{\mu,\sigma}(\underline{X}) = \log \mathcal{L}_{\mu,\sigma}(\underline{X}) = -\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{\sum_j (X_j-\mu)^2}{2\sigma^2}$$

a) $$\partial_\mu \ell_{\mu,\sigma}(\underline{X}) = 0 \iff -\partial_\mu \frac{\sum_j (X_j-\mu)^2}{2\sigma^2} = 0$$

$$\iff \cancel{-}\frac{\sum_j \cancel{2}(X_j-\mu)\cdot\cancel{(-1)}}{\cancel{2}\sigma^2} = 0$$

$$\iff \sum_j X_j - n\mu = 0$$

$$\iff \hat{\mu} = \frac{1}{n}\sum_j X_j$$

b) $$\frac{1}{n}\sum_{j=1}^{n} X_j \sim N\left(\mu, \frac{1}{n}\right) \Longrightarrow Z_n = \frac{\frac{1}{n}\sum_{j=1}^{n} X_j - \mu}{1/\sqrt{n}} \sim N(0,1)$$

$\Longrightarrow$ Sia (A_-^α, A_+^α) un intervallo di confidenza per una distribuzione $N(0,1)$ al livello $1-\alpha$, allora

$$1-\alpha = \mathbb{P}\left(Z_n \in (A_-^\alpha, A_+^\alpha)\right) = \mathbb{P}\left(\frac{\overline{X}_n - \mu}{1/\sqrt{n}} \in (A_-^\alpha, A_+^\alpha)\right)$$

$$= \mathbb{P}\left(\mu \in \left(\overline{X}_n - \frac{1}{\sqrt{n}} A_+^\alpha, \overline{X}_n + \frac{1}{\sqrt{n}} A_-^\alpha\right)\right)$$

c) $\partial_\sigma \ell_{\mu,\sigma}(\underline{X}) = 0 \iff -n\frac{1}{\sigma} + \frac{\sum_j (X_j - \mu)^2}{\sigma^3} = 0$

$$\iff \sigma^2 = \frac{1}{n}\sum_{j=1}^n (X_j - \mu)^2$$

d) $\begin{pmatrix} \partial_\mu \\ \partial_\sigma \end{pmatrix} \ell_{\mu,\sigma}(\underline{X}) = 0 \iff \begin{cases} \sum_j X_j - n\mu = 0 \\ -n\frac{1}{\sigma} + \frac{\sum_j (X_j-\mu)^2}{\sigma^3} = 0 \end{cases}$

$$\iff \begin{cases} \hat{\mu} = \frac{1}{n}\sum_j X_j \\ \hat{\sigma}^2 = \frac{1}{n}\sum_j (X_j - \hat{\mu})^2 \end{cases}$$

5. Una impresa di componenti elettronici decide di stimare il numero medio di piccole imperfezioni su un sistema assemblato. A tale scopo analizza 16 sistemi e rileva che il numero medio di imperfezioni ammonta a 32 e che la varianza campionaria risulta 8. Supponendo che le imperfezioni abbiano distribuzione normale, determinare l'intervallo di fiducia al 95% e al 99% per il numero medio di piccole imperfezioni su sistemi assemblati.

$n = 16 \quad , \quad \overline{X}_{16} = 32 \quad , \quad S^2 = 8$

$$\Longrightarrow \quad \frac{\overline{X}_{16} - \mu}{S/\sqrt{16}} = T_{16} \sim t(15)$$

$\Longrightarrow$ Intervallo di fiducia bilatero

$$0.95 = P(a \le T_{16} \le b) = P\left(a \le \frac{\overline{X}_{16} - \mu}{S/4} \le b\right)$$

$$= P\left(\frac{\overline{X}_{16} - b}{S/4} \le \mu \le \frac{\overline{X}_{16} - a}{S/4}\right)$$

- $a_{5\%} = b_{5\%} = 2.131$

$$\Longrightarrow \quad IC_{95\%} = \left(\frac{32 - 2.131}{\sqrt{8}/4}, \frac{32 + 2.131}{\sqrt{8}/4}\right)$$

- $a_{2\%} = b_{2\%} = 2.602$

$$\Longrightarrow \quad IC_{95\%} = \left(\frac{32 - 2.602}{\sqrt{8}/4}, \frac{32 + 2.602}{\sqrt{8}/4}\right)$$

One-sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two-sided	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291
One-sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two-sided	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%

6. * Dato $n \geq 1$, sia $(X_1, X_2, \ldots, X_n)$ un campione estratto da una popolazione di densità

$$f_\theta(x) = \theta \exp(x - \theta \exp(x)), \qquad \text{per } \theta > 0\,.$$

(a) Determinare lo stimatore Θ_n di massima verosimiglianza di θ

(b) Stabilire la correttezza e la consistenza della successione $(\Theta_n)_{n\geq 1}$

(c) Dato $\alpha \in (0,1)$, determinare una regione di fiducia al livello $1-\alpha$ per θ

a) $$\mathcal{L}_\vartheta(x_1, \ldots x_n) = \prod_{j=1}^n f_\vartheta(x_j) = \vartheta^n \exp\Big(\sum_{j=1}^n x_j - \vartheta \sum_{j=1}^n \exp(x_j)\Big)$$

$$= \exp\Big(n \log\vartheta + \sum_{j=1}^n x_j - \vartheta \sum_{j=1}^n \exp(x_j)\Big) \qquad \vartheta \in (0,\infty)$$

$$\frac{\partial}{\partial\vartheta} \log(\mathcal{L}_\vartheta(x_1, \ldots x_n)) = \frac{n}{\vartheta} - \sum_{j=1}^n \exp(x_j) \overset{!}{=} 0$$

$$\Longrightarrow \quad \Theta_n = \Big(\frac{1}{n} \sum_{j=1}^n \exp(X_j)\Big)^{-1}$$

maximum: $\partial_\vartheta^2 \log \mathcal{L}_\vartheta < 0$

b) Per $n=1$ abbiamo

$$\mathbb{E}_\vartheta(\Theta_1) = \mathbb{E}_\vartheta(e^{-X_1}) = \int_{-\infty}^{\infty} \cancel{e^{-x}}\, \vartheta\, e^{\cancel{x} - \vartheta \exp(x)}\, dx =$$

$$= \vartheta \int_{-\infty}^{\infty} e^{-\vartheta \exp(x)}\, dx = \infty \quad \Longrightarrow \text{ non corretto.}$$
per $n=1$

Per $n \geq 1$: Notiamo che $Y = \exp(X)$ ha densité

$x = \log y$

$$f_Y(y)\, dy = f_X(x(y)) \cdot \Big|\frac{dx}{dy}\Big|\, dy = \vartheta \cdot \exp(\log(y) - \vartheta y) \cdot \frac{1}{y}$$

$$= \vartheta \exp(-\vartheta y)$$

$$\Longrightarrow Y_i = \exp(X_i) \sim \mathcal{E}xp(\vartheta)$$

$$\Longrightarrow \mathbb{E}_\vartheta\left(\left(\frac{1}{n}\sum_{j=1}^{n} e^{X_j}\right)^{-1}\right) = n\,\tilde{\mathbb{E}}_\vartheta\left(\frac{1}{\sum_{j=1}^{n} Y_j}\right) =$$

$$\sum_{j=1}^{n} Y_j = Z \sim \Gamma(\vartheta, n)$$

$$= n\,\hat{\tilde{\mathbb{E}}}_\vartheta\left(\frac{1}{Z}\right)$$

$$= n \int_0^\infty \frac{1}{z}\cdot\frac{1}{(n-1)!}\vartheta^n z^{n-1} e^{-\vartheta z}\,dz$$

$$= n\,\frac{1}{n-1}\,\vartheta \int_0^\infty \vartheta^{n-1} z^{n-2} e^{-\vartheta z}\,dz$$

$$= \frac{n}{n-1}\vartheta$$

$\Longrightarrow$ non corretto.

Consistenza: Consideriamo lo stimatore $\Psi_n = \frac{1}{\Theta_n}$

$$\mathbb{E}_\vartheta(\Psi_n) = \mathbb{E}_\vartheta\left(\frac{1}{n}\sum_{j=1}^{n} e^{X_j}\right) = \mathbb{E}_\vartheta(e^{X_1}) = \int \vartheta e^{2x}\exp(-\vartheta e^{x})\,dx = \frac{1}{\vartheta}$$

$$\overset{LLN}{\Longrightarrow} \mathbb{P}\left(\left|\Psi_n - \frac{1}{\vartheta}\right| \le \delta\right) = \mathbb{P}\left(\left|\frac{1}{n}\sum_{j=1}^{n} e^{X_j} - \mathbb{E}(e^{X_j})\right| \le \delta\right) \xrightarrow{n\to\infty} 1 \quad \forall \delta > 0$$

Ora, per ogni $\vartheta > 0$ e per ogni $\varepsilon > 0$ esiste $\delta_{\vartheta,\varepsilon}$ tale che

$$\left|x - \frac{1}{\vartheta}\right| \leq \delta_{\theta,\varepsilon} \implies \left|\frac{1}{x} - \vartheta\right| \leq \varepsilon$$

$$\implies P(|\Theta_n - \vartheta| \leq \varepsilon) = P\left(\left|\frac{1}{\bar{\Psi}_n} - \vartheta\right| \leq \varepsilon\right)$$

$$\geq \mathbb{P}\left(\left|\bar{\Psi}_n - \frac{1}{\vartheta}\right| \leq \varepsilon\right) \xrightarrow{n\to\infty} 1 \implies \text{consistente}$$

c) Notiamo che $Y = \exp(X)$ ha densità

$$f_Y(y)\,dy = f_X(x(y)) \cdot \left|\frac{dx}{dy}\right| dy \overset{x = \log y}{=} \vartheta \cdot \exp(\log(y) - \vartheta y) \cdot \frac{1}{y}$$

$$= \vartheta \exp(-\vartheta y)$$

$$\implies Y_i = \exp(X_i) \sim \mathcal{E}xp(\vartheta)$$

$$\implies Z_i = 2\vartheta \exp(X_i) \sim \mathcal{E}xp\left(\tfrac{1}{2}\right)$$

$$\implies \underbrace{2n \cdot \vartheta \cdot \hat{\vartheta}_n}_{\text{variabile pivotale.}} = \sum_{j=1}^n \vartheta \exp(X_j) = \sum_{j=1}^n Z_j \sim \Gamma\left(n, \tfrac{1}{2}\right) = \chi^2(2n)$$

$\implies$ Sia (A_-^α, A_+^α) un intervallo di confidenza al livello $1-\alpha$ per $\chi^2(2n)$, allora

$$(1-\alpha) = \mathbb{P}(2n\vartheta\,\Theta_n \in (A_-^\alpha, A_+^\alpha)) = \mathbb{P}\Big(\vartheta \in \underbrace{\left(\frac{A_-^\alpha}{n\Theta_n 2}, \frac{A_+^\alpha}{n\Theta_n 2}\right)}_{\text{IC di } \vartheta \text{ al liv. } (1-\alpha)}\Big)$$

7. Dato $n \geq 1$, sia $(X_1, X_2, \dots, X_n)$ un campione estratto da una popolazione di densità

$$f_\theta(x) = \frac{1}{12\theta^5} x^9 e^{-x^2/\theta} \mathbb{1}_{(0,\infty)}(x),$$

dove $\mathbb{1}_{(0,\infty)}(x)$ rappresenta la funzione indicatrice sull'insieme $(0,\infty)$.

(a) Determinare lo stimatore Θ_n di massima verosimiglianza di θ.

(b) Stabilire la correttezza della successione $(\Theta_n)_{n\geq 1}$.

(c) Dato $\alpha \in (0,1)$, determinare una regione di fiducia al livello $1-\alpha$ per θ.

a) Calcoliamo:

$$\mathcal{L}_\theta(x_1,\dots,x_n) = \prod_{j=1}^n f_\theta(x_j) = \frac{1}{12^n} \theta^{-5n} e^{-\frac{1}{\theta}\sum_{j=1}^n x_j^2} \prod_{j=1}^n x_j^9$$

$$\Longrightarrow \log \mathcal{L}_\theta(x_1,\dots x_n) =$$

$$= -n\log 12 - 5n\log\theta + 9\sum_{j=1}^n \log x_j - \frac{1}{\theta}\sum_{j=1}^n x_j^2$$

$$\Longrightarrow \partial_\theta \log \mathcal{L}_\theta(x_1,\dots,x_n) = -\frac{5n}{\theta} + \frac{1}{\theta^2}\sum_{j=1}^n x_j^2 \overset{!}{=} 0$$

$$\Longleftrightarrow \quad \Theta_n = \frac{1}{5n}\sum_{j=1}^n X_j^2$$

maximum: $\partial_\theta^2 \log \mathcal{L}_{\Theta_n}(\dots) > 0$

b) $$\mathbb{E}_\theta(\Theta_n) = \mathbb{E}_\theta\left(\frac{1}{5n}\sum_{j=1}^n X_j^2\right) = \frac{1}{5n}\sum_{j=1}^n \mathbb{E}_\theta(X_j^2) =$$

$$= \frac{1}{5}\int_0^\infty x^2 \frac{1}{12\theta^5} x^9 e^{-x^2/\theta} dx = \theta \underbrace{\int_0^\infty \frac{1}{5!} y^5 e^{-y} dy}_{\text{integrale sulla densità di una } \Gamma(6,1)} = \theta$$

$y = \frac{x^2}{\theta}$

$\Longrightarrow$ lo stimatore è corretto

c) Sulla base della sostituzione fatta al punto precedente definiamo: $Y_j = \frac{X_j^2}{\theta}$

$$f_Y(y)\,dy = \frac{1}{4!} y^4 e^{-y}\,dy \implies Y_j = \frac{X_j^2}{\theta} \sim \Gamma(5,1)$$

$$\implies Z_j = 2\frac{X_j^2}{\theta} \sim \Gamma\left(5,\tfrac{1}{2}\right) = \chi^2(10)$$

$$\implies \sum_{j=1}^n Z_j = \frac{2}{\theta}\sum_{j=1}^n X_j^2 = \frac{2}{\theta}5n\frac{1}{5n}\sum_{j=1}^n X_j^2$$

$$= \frac{10n}{\theta}\Theta_n \sim \chi^2(10n)$$

$\implies$ Sia (A_-^α, A_+^α) un intervallo di confidenza al livello $1-\alpha$ per la distribuzione $\chi^2(10n)$ allora

$$1-\alpha = \mathbb{P}\left(\sum_{j=1}^n Z_j \in (A_-^\alpha, A_+^\alpha)\right) =$$

$$= \mathbb{P}\left(\frac{10n}{\theta}\Theta_n \in (A_-^\alpha, A_+^\alpha)\right)$$

$$= \mathbb{P}\left(\theta \in \underbrace{\left(\frac{10n\Theta_n}{A_+^\alpha}, \frac{10n\Theta_n}{A_-^\alpha}\right)}_{\text{IC al livello } 1-\alpha \text{ per } \theta.}\right)$$

8. Sia dato un campione $X_1, \ldots, X_n$ con comune funzione di ripartizione (sconosciuta) $F : \mathbb{R} \to [0,1]$. Sia $u \in \mathbb{R}$ e definiamo, per $i = 1, \ldots, n$,

$$Y_i(u) = \mathbb{1}_{X_i \leq u}$$

Il campione $Y_1(u), \ldots, Y_n(u)$ ha legge Bernoulli di parametro $F(u) \in [0,1]$. Sia

$$\hat{F}_n(u) := \frac{1}{n} \sum_{i=1}^{n} Y_i(u)$$

(a) $\hat{F}_n(u)$ è una stima non distorta di $F(u)$? Asintoticamente distorta?

(b) Calcolare il rischio quadratico (medio) $R(F(u), \hat{F}_n(u)) = \mathbb{E}_{F(u)}[(\hat{F}_n(u) - F(u))^2]$.

(c) La successione $\{\hat{F}_n(u)\}_n$ è consistente?

(d) Trovare un intervallo di fiducia al livello $1-\alpha$, $0 < \alpha \ll 1$, per $F(u)$.

a) $\mathbb{E}_F(\hat{F}_n(u)) = \mathbb{E}_F\left(\frac{1}{n}\sum_{i=1}^{n} Y_i(u)\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_F(\mathbb{1}_{X_i \leq u})$

$= \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}_F(X_i \leq u) = \frac{1}{n} \cdot n \cdot F(u) = F(u)$

$\Longrightarrow$ lo stimatore è corretto (non distorto) e quindi asintoticamente corretto

b) $\mathbb{E}_F\left(\left(\frac{1}{n}\sum_{i=1}^{n} Y_i(u) - F(u)\right)^2\right) =$

$= \mathbb{E}_F\left(\left(\frac{1}{n}\sum_{i=1}^{n} Y_i(u)\right)^2\right) - 2F(u)\,\mathbb{E}_F\left(\frac{1}{n}\sum_{j=1}^{n} Y_j(u)\right) + F(u)^2$

$= \frac{1}{n^2}\sum_{i,j=1}^{n} \mathbb{E}_F(Y_i(u) Y_j(u)) - F(u)^2$

$= \underbrace{\frac{1}{n^2}\sum_{i,j=1}^{n} \mathbb{E}_F(Y_i(u))\mathbb{E}_F(Y_j(u)) - F(u)^2}_{0} + \frac{1}{n^2}\sum_{j=1}^{n} \underbrace{\mathbb{E}(Y_j(u)^2) - \mathbb{E}(Y_j(u))^2}_{Var(Y_j(u)) = F(u)(1-F(u))}$

$= \frac{1}{n} F(u)(1-F(u))$

c) Siccome $R(F(v), \hat{F}_n(v)) \xrightarrow{n\to\infty} 0$

$$P_F(|\hat{F}_n(v) - F(v)|^2 > \varepsilon^2) \overset{(M)}{\leq} \frac{\mathbb{E}_F((\hat{F}_n(v) - F(v))^2)}{\varepsilon^2} = \frac{R(F(v), \hat{F}_n(v))}{\varepsilon^2} \xrightarrow[n\to\infty]{} 0$$

$\longrightarrow$ $\hat{F}_n(v)$ è consistente

d) Notiamo che $n\hat{F}_n(v)$ ha distribuzione binomiale con parametri $(n, F(v))$

Calcoliamo un intervallo di confidenza approssimativo

Per il CLT: $\hat{F}_n(v) \underset{n\gg 1}{\sim} N\left(F(v), \frac{F(v)(1-F(v))}{n}\right)$

$$\Longrightarrow \frac{\hat{F}_n(v) - F(v)}{\sqrt{F(v)(1-F(v))}}\sqrt{n} \sim N(0,1) \qquad \sigma_F = \sqrt{F(v)(1-F(v))}$$

Sia (z_-^α, z_+^α) un intervallo di conf. al livello $1-\alpha$ per $Z \sim N(0,1)$, allora

$$1-\alpha = \mathbb{P}\left(\frac{\hat{F}_n(v) - F(v)}{\sigma_F}\sqrt{n} \in (z_-^\alpha, z_+^\alpha)\right)$$

dipende da $F(v)$!

per $F \in (0,1)$

$$= \mathbb{P}\left(F(v) \in \left(\hat{F}_n(v) - \frac{1}{\sqrt{n}} z_+^\alpha \sigma_F, \hat{F}_n(v) + \frac{1}{\sqrt{n}} z_-^\alpha \sigma_F\right)\right)$$

$\sigma_F = \sqrt{F(1-F)} \leq \sqrt{\frac{1}{2}(1-\frac{1}{2})} = \frac{1}{2}$

$$\leq \mathbb{P}\Bigg(F(v) \in \underbrace{\left(\hat{F}_n(v) - \frac{1}{\sqrt{n}}\frac{1}{2} z_+^\alpha, \hat{F}_n(v) + \frac{1}{\sqrt{n}}\frac{1}{2} z_-^\alpha\right)}_{\text{IC al livello } 1-\alpha \text{ per } F(v)}\Bigg)$$

Alternativamente, notiamo che

$$\left(\frac{n}{2} - a_\alpha^{(n)}, \frac{n}{2} + a_\alpha^{(n)}\right) \text{ è l'IC al livello } 1-\alpha$$

di $\mathrm{Bin}(n, \frac{1}{2})$, allora

$$\left(np - a_\alpha^{(n)}, np + a_\alpha^{(n)}\right) \text{ contiene l'IC al}$$

livello $1-\alpha$ di $\mathrm{Bin}(n,p)$ $\forall p > 0$

Quindi otteniamo

$$1-\alpha = \mathbb{P}\left(|\hat{F}_n(v) - F(v)| \leq a_\alpha^{(n)}\right)$$

$$= \mathbb{P}\left(F(v) \in \left(\hat{F}_n(v) - a_\alpha^{(n)}, \hat{F}_n(v) + a_\alpha^{(n)}\right)\right)$$

Foglio di esercizi 10

Discussione soluzioni: 24.05.2022

* rappresenta esercizi prioritari. Si consiglia di provare gli esercizi sui test statistici dopo la lezione di Martedì 23.05, durante la quale sarà fatto un richiamo su questo argomento.

1. * Sia dato un campione $X_1, \ldots, X_n$ con densità esponenziale di parametro $\lambda > 0$.
   (a) Trovare la stima di massima verosimiglianza $\hat{\lambda}_n$ del parametro λ e $\hat{e}ta_n$ dei $h(\lambda) = \lambda^{-1}$.
   (b) Queste stime sono corrette? asintoticamente corrette?
   (c) La successione di stime $(\hat{\lambda}_n)_n$ è consistente?
   (d) Esaminare il test unilatero di ipotesi nulla $H_\theta : \lambda \leq 1$ e ipotesi alternativa $\mathcal{H}_1 : \lambda > 1$ al livello α.
   (e) Esaminare il test di ipotesi nulla $H_\theta : \lambda = 2$ e ipotesi alternativa $\mathcal{H}_1 : \lambda \neq 2$ al livello α.

2. * Sia dato un campione $X_1, \ldots, X_n$ con gegge Normale con parametri $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{\geq 0}$.
   (a) Trovare un intervallo di confidenza a livello $1 - \alpha$ per il parametro μ assumendo di non conoscere σ^2.
   (b) Trovare un intervallo di confidenza a livello $1 - \alpha$ per il parametro σ^2 assumendo di conoscere μ.
   (c) Esaminare il test unilatero di ipotesi nulla $H_\theta : \mu \leq \mu_0$ e ipotesi alternativa $\mathcal{H}_1 : \mu > \mu_0$ al livello α assumendo di non conoscere σ^2.
   (d) Esaminare il test di ipotesi nulla $H_\theta : \mu = \mu_0$ e ipotesi alternativa $\mathcal{H}_1 : \mu \neq \mu_0$ al livello α assumendo di non conoscere σ^2.
   (e) Esaminare il test di ipotesi nulla $H_\theta : \sigma^1 \leq \sigma_0^2$ e ipotesi alternativa $\mathcal{H}_1 : \sigma^1 > \sigma_0^2$ al livello α assumendo di conoscere μ.

3. (Continua dal foglio precedente) Dato $n \geq 1$, sia $(X_1, X_2, \ldots, X_n)$ un campione estratto da una popolazione di densità
$$f_\theta(x) = \frac{1}{12\theta^5} x^9 e^{-x^2/\theta} \mathbb{1}_{(0,\infty)}(x),$$
   dove $\mathbb{1}_{(0,\infty)}(x)$ rappresenta la funzione indicatrice sull'insieme $(0, \infty)$.
   (a) Determinare lo stimatore Θ_n di massima verosimiglianza di θ.
   (b) Formulare un test per l'ipotesi nulla $\mathcal{H}_0 : \theta \leq \theta_0$, ed ipotesi alternativa $\mathcal{H}_1 : \theta > \theta_0$ al livello di significatività α.

4. (Continua dal foglio precedente) Dato $n \geq 1$, sia $(X_1, X_2, \ldots, X_n)$ un campione estratto da una popolazione di densità
$$f_\theta(x) = \theta \exp(x - \theta \exp(x)), \qquad \text{per } \theta > 0\,.$$
   (a) Determinare lo stimatore Θ_n di massima verosimiglianza di θ.
   (b) Formulare un test per l'ipotesi nulla $\mathcal{H}_0 : \theta \leq \theta_0$, ed ipotesi alternativa $\mathcal{H}_1 : \theta > \theta_0$ al livello di significatività α.

5. Si consideri, per $a \in \mathbb{R}$ e $\theta > 0$ un campione $X_1, X_2, ..., X_n$ estratto da una popolazione di densità
$$f_{a,\theta}(x) = \frac{1}{\theta} e^{\frac{1}{\theta}(x-a)} \mathbb{1}_{[a,\infty)}(x).$$
   (a) Determinare gli stimatori di massima verosimiglianza (A_n, Θ_n) dei parametri (a, θ).

(b) Per a costante, formulare un test unilatero con ipotesi

$$\mathcal{H}_0 : \theta \leq \theta_0, \qquad \mathcal{H}_1 : \theta > \theta_0$$

e livello $\alpha \in (0,1)$.

(c) Stabilire se $(A_n)_{n\geq 2}$ è consistente.

(d) Stabilire se $(\Theta_n)_{n\geq 2}$ è consistente.

6. * Sia dato un campione $X_1, \ldots, X_n$ con gegge di Poisson con parametro $\lambda > 0$.

(a) Trovare la stima di massima verosimiglianza $\hat{\lambda}_n$ del parametro λ.

(b) Questa stima è corretta? asintoticamente corretta?

(c) La successione di stime $(\hat{\lambda}_n)_n$ è consistente?

(d) Trovare un intervallo di confidenza a livello $1-\alpha$ per il parametro λ.

(e) Esaminare il test unilatero di ipotesi nulla $H_\theta : \lambda \leq 1$ e ipotesi alternativa $\mathcal{H}_1 : \lambda > 1$ al livello α.

(f) Esaminare il test di ipotesi nulla $H_\theta : \lambda = 2$ e ipotesi alternativa $\mathcal{H}_1 : \lambda \neq 2$ al livello α.

<u>Richiamo</u>: $T \sim \exp(\lambda) \implies \alpha T \sim \exp\left(\frac{\lambda}{\alpha}\right)$

$g(t) = \alpha t \qquad f_{\alpha T}(s) = f_T(s) \cdot \frac{1}{\alpha} = \frac{\lambda}{\alpha} e^{-\lambda \frac{s}{\alpha}} \sim \exp\left(\frac{\lambda}{\alpha}\right)$

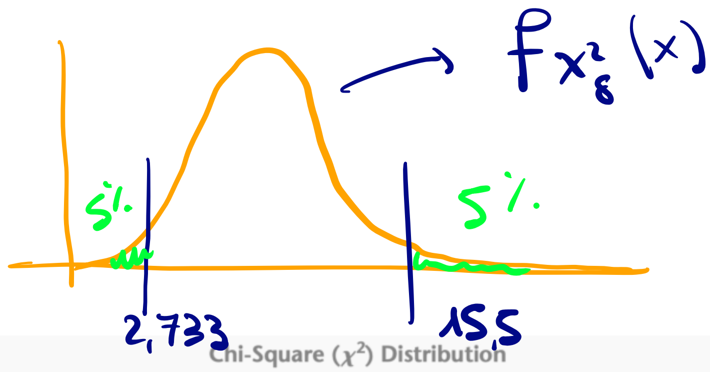
$\sum_{i=1}^{n} T_i = S_n \sim \text{Gamma}(n, \lambda)$

$f_{S_n}(t) = \frac{\lambda^{n-1}}{(n-1)!} t^{n-1} \cdot \lambda e^{-\lambda t} \qquad t \geq 0$

$\alpha S_n = \alpha \sum_{i=1}^{n} T_i = \sum_{i=1}^{n} \alpha T_i \sim \text{Gamma}\left(n, \frac{\lambda}{\alpha}\right)$

$\text{Gamma}(\nu, 2) = \chi^2_{2\nu}$

Se $X_i \overset{iid}{\sim} N(0,1) \implies \sum_{i=1}^{n} X_i^2 \sim \chi^2_n$



Chi-Square (χ^2) Distribution

Area to the Right of Critical Value

Degrees of Freedom	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

1. * Sia dato un campione $X_1, \ldots, X_n$ con densità esponenziale di parametro $\lambda > 0$.

   (a) Trovare la stima di massima verosimiglianza $\hat{\lambda}_n$ del parametro λ e $\hat{eta}_n$ dei $h(\lambda) = \lambda^{-1}$.

   (b) Queste stime sono corrette? asintoticamente corrette?

   (c) La successione di stime $(\hat{\lambda}_n)_n$ è consistente?

   (d) Esaminare il test unilatero di ipotesi nulla $H_\theta : \lambda \leq 1$ e ipotesi alternativa $\mathcal{H}_1 : \lambda > 1$ al livello α.

   (e) Esaminare il test di ipotesi nulla $H_\theta : \lambda = 2$ e ipotesi alternativa $\mathcal{H}_1 : \lambda \neq 2$ al livello α.

a) $$\mathcal{L}_\lambda(x_1, \ldots x_n) = \prod_{j=1} f_\lambda(x_i) = \prod_{j=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{j=1} x_j}$$

$$\ell_\lambda(x_1, \ldots x_n) = + n \log \lambda - \lambda \sum_i x_i$$

$$\partial_\lambda \ell_\lambda(x_1, \ldots x_n) = + \frac{n}{\lambda} - \sum_i x_i \stackrel{!}{=} 0 \iff \hat{\lambda} = \frac{1}{\frac{1}{n} \sum x_i}$$

Analogamente: $$\mathcal{L}_\eta(x_1, \ldots x_n) = \eta^{-n} e^{-\eta^{-1} \sum_{j=1}^{n} x_j}$$

$$\partial_\eta \ell_\eta(x_1, \ldots x_n) = -\frac{n}{\eta} + \frac{1}{\eta^2} \sum_{j=1}^{n} x_j \iff \hat{\eta} = \frac{1}{n} \sum_{j=1}^{n} x_j$$

b) $$\mathbb{E}(\hat{\eta}) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{E}(x_j) = \frac{1}{n} \cdot n \cdot \eta = \eta \implies \text{corretto}$$

$$\mathbb{E}(\hat{\lambda}) = \mathbb{E}\left(\frac{1}{\frac{1}{n} \sum X_i}\right) \overset{n=1}{=} \mathbb{E}\left(\frac{1}{X_1}\right) = \int_0^\infty \frac{1}{x} \lambda e^{-\lambda x} dx$$

$$= \infty \neq \lambda$$

$$\mathbb{E}(\hat{\lambda}_n) = \mathbb{E}\left(\frac{1}{\frac{1}{n} \sum_i X_i}\right) = n\, \mathbb{E}\left(\frac{1}{\sum_i X_i}\right) = n\mathbb{E}'\left(\frac{1}{Y}\right)$$

$Y \sim \text{Gamma}(n, \lambda)$ $$= n \int_0^\infty \frac{1}{\cancel{y}} \cdot \frac{y^{n-2} \lambda^n}{(n-1)!} e^{-\lambda y} dy$$

$$= \frac{n\lambda}{n-1} \int_0^\infty \frac{y^{n-2} \lambda^{n-1}}{(n-2)!} e^{-\lambda y} dy =$$

$$= \frac{n\lambda}{n-1} \xrightarrow{n\to\infty} \lambda$$

$f_Z \quad Z \sim Gamma(n-1,\lambda)$

$\hat\lambda_n$ non è corretto ma è asintoticamente corretto.

c) La consistenza di $\hat\lambda_n$ segue dal teorema visto in classe:

$f_\lambda(x) = \lambda \mathbb{1}_{(0,\infty)}(x)\, e^{\lambda(-x)}$ è un modello esponenziale con

$c_\theta \; g(x) \quad e^{\theta T(x)}$

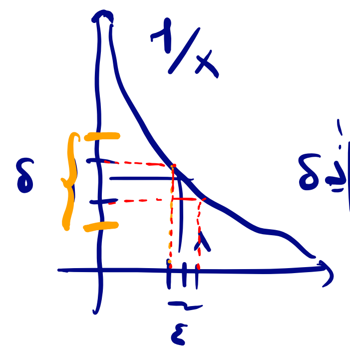
$x \mapsto \mathbb{1}_{(0,\infty)}(x)\, x^2 e^{-\lambda x}$ integrabile

$\Longrightarrow$ consistente

Alternativamente:

$$P(|\hat\lambda_n - \lambda| > \varepsilon) = P\left(\left|\frac{1}{\frac{1}{n}\sum_i X_i} - \lambda\right| > \varepsilon\right)$$

$$\le P\left(\left|\frac{1}{n}\sum_i X_i - \frac{1}{\lambda}\right| > \delta\right) \longrightarrow 0$$



$$\delta \doteq \left|\varepsilon \left(\frac{1}{x}\right)'\Big|_\lambda\right| = \varepsilon \frac{1}{\lambda^2}$$

$\longrightarrow \hat\lambda$ è consistente.

d) $H_0 : \lambda \le 1$

Consideriamo lo stimatore $\frac{1}{\hat\lambda} = \frac{1}{n}\sum_i X_i \sim Gamma(n, \lambda n)$

$$\left[\sum_i X_i \sim Gamma(n,\lambda) \Longrightarrow \alpha \sum_i X_i \Big|_{\alpha=\frac{1}{n}} \sim Gamma\left(n, \frac{\lambda}{\alpha}\right)\right]_{\alpha=\frac{1}{n}}$$

"Standardizziamo" la VA $\frac{1}{n}\sum_i X_i$

Richiamo: $Gamma(n,2) = \chi^2_{2n}$

Vediamo che $\frac{\lambda n}{2}\left(\frac{1}{n}\sum_i X_i\right) \sim Gamma\left(n, \frac{\lambda n}{\frac{\lambda n}{2}}\right)$

$$= \chi^2_{2n}$$

Quantile $q_{2n,\alpha}$ al livello α

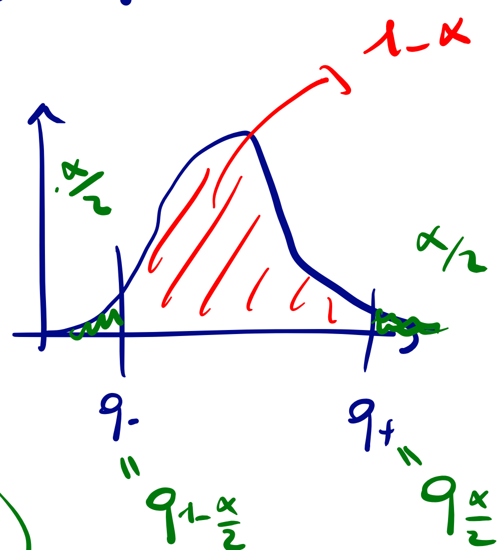
Superiamo il test se

$$\frac{\lambda n}{2} \cdot \left(\frac{1}{n} \sum_i X_i\right) < q_{2n,\alpha}$$

Concretamente: superiamo il test se

$$\frac{n}{2} \left(\frac{1}{n} \sum_i X_i\right) \leq q_{2n,}$$

e) $H_0 : \lambda = 2$



Superiamo il test se

$$\frac{n \overset{\lambda}{2}}{2} \cdot \left(\frac{1}{n} \sum_i X_i\right) \in \left(q_{1-\frac{\alpha}{2}}, q_{\frac{\alpha}{2}}\right)$$

In pratica $1-\alpha = 95\%$

$q_{2.5\%}$ si trova sulla tabella per χ^2_{2n}

e vale per $n = 5$

$q_{2.5\%, 2n=10} = 3.247$

$q_{97.5\%, 2n=10} = 20.4$

2. * Sia dato un campione $X_1, \ldots, X_n$ con gegge Normale con parametri $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{\geq 0}$.

(a) Trovare un intervallo di confidenza a livello $1-\alpha$ per il parametro μ assumendo di non conoscere σ^2.

(b) Trovare un intervallo di confidenza a livello $1-\alpha$ per il parametro σ^2 assumendo di conoscere μ.

(c) Esaminare il test unilatero di ipotesi nulla $H_\theta : \mu \leq \mu_0$ e ipotesi alternativa $\mathcal{H}_1 : \mu > \mu_0$ al livello α assumendo di non conoscere σ^2.

(d) Esaminare il test di ipotesi nulla $H_\theta : \mu = \mu_0$ e ipotesi alternativa $\mathcal{H}_1 : \mu \neq \mu_0$ al livello α assumendo di non conoscere σ^2.

(e) Esaminare il test di ipotesi nulla $H_\theta : \sigma^2 \leq \sigma_0^2$ e ipotesi alternativa $\mathcal{H}_1 : \sigma^2 > \sigma_0^2$ al livello α assumendo di conoscere μ.

a) Sia $S^2 = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \bar{X})^2$, $\quad \bar{X} = \frac{1}{n} \sum_{j=1}^{n} X_j$

allora $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$

$$\Longrightarrow 1-\alpha = \mathbb{P}\left(T \in (t^-_{\upsilon,\alpha}, t^+_{\upsilon,\alpha})\right) = \mathbb{P}\left(\frac{\bar{X}-\mu}{S/\sqrt{n}} \in (t^-_{\upsilon,\alpha}, t^+_{\upsilon,\alpha})\right)$$

$$= \mathbb{P}\Big(\mu \in \underbrace{\left(\bar{X} - \frac{S}{\sqrt{n}} t^+_{\upsilon,\alpha}, \bar{X} + \frac{S}{\sqrt{n}} t^-_{\upsilon,\alpha}\right)}_{\text{IC al livello } 1-\alpha}\Big)$$

b) $\sum_{j=1}^{n} \left(\frac{X_j - \mu}{\sigma}\right)^2 \sim \chi^2(n)$

Sia (χ^-_n, χ^+_n) un intervallo di confid. al livello $1-\alpha$ per la variabile Z_n con distribuzione $\chi^2(n)$, allora

$$1-\alpha = \mathbb{P}\left(Z_n \in (\chi^-_n, \chi^+_n)\right) = \mathbb{P}\Big(\frac{n-1}{\sigma^2} \underbrace{\frac{1}{n-1} \sum_{j=1}^{n} (X_j - \mu)^2}_{S^2} \in (\chi^-_n, \chi^+_n)\Big)$$

$$= \mathbb{P}\left(\sigma^2 \in \left(\frac{S^2}{\chi^+_n/n-1}, \frac{S^2}{\chi^-_n/n-1}\right)\right)$$

c) Come visto in classe, il modello gaussiano

$$(\mathbb{R}, \mathcal{B}(\mathbb{R}), (\mathcal{N}(\mu, \sigma^2))_\mu)$$

è a verosimiglianza crescente rispetto a $\bar{X}$

regione critica $\quad C = \{\bar{X} \geq d_\alpha\}$

$$ST = \frac{\sqrt{n}}{S} \cdot (\bar{X} - \mu) \sim t(n-1) \qquad S = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X})^2}$$

Troviamo d_α:

$$\alpha = P_{\mu_0}(\bar{X} \geq d_\alpha) = P_{\mu_0}\Big(\underbrace{\frac{\sqrt{n}}{S}(\bar{X} - \mu_0)}_{T \sim t(n-1)} \geq \frac{\sqrt{n}}{S}(d_\alpha - \mu_0)\Big)$$

$$\iff \frac{\sqrt{n}}{S}(d_\alpha - \mu_0) = t_{\nu, 1-\alpha} \iff d_\alpha = \mu_0 + \frac{S}{\sqrt{n}} t_{\nu, 1-\alpha}$$

d) Usando la connessione tra test bilateri e intervalli di fiducia abbiamo

$$C = \{\omega \in \Omega : \theta_0 \notin D(\omega)\} = \{\omega \in \Omega : \underbrace{|\bar{X} - \mu_0| \notin \left(\frac{S}{\sqrt{n}} t^-_{\nu,\alpha}, \frac{S}{\sqrt{n}} t^+_{\nu,\alpha}\right)}_{\text{Esempio: } |\bar{X} - \mu_0| \geq t_{\nu, 1-\frac{\alpha}{2}}}\}$$

$\left(1-\frac{\alpha}{2}\right)$ quantile di $t(\nu)$

Esplicitamente:

$$\alpha = P(|T| \geq t_{\nu, 1-\frac{\alpha}{2}}) = P\left(\left|\frac{\bar{X} - \mu_0}{S/\sqrt{n}}\right| \geq t_{\nu, 1-\frac{\alpha}{2}}\right)$$

$$= P\Big(\{\omega : \bar{X}(\omega) \in \underbrace{\left[\mu_0 - \frac{S}{\sqrt{n}} t_{\nu, 1-\frac{\alpha}{2}}, \mu_0 + \frac{S}{\sqrt{n}} t_{\nu, 1-\frac{\alpha}{2}}\right]^c}_{\tilde{C}}\}\Big)$$

e) Modello statistico $\left(\mathbb{R}, \mathcal{B}(\mathbb{R}), (\mathcal{N}(\mu,\sigma^2))_{\sigma^2}\right)$

per $\sigma_1^2 < \sigma_2^2$ abbiamo

$$\frac{L_{\sigma_2^2}(x_1 \dots x_n)}{L_{\sigma_1^2}(x_1 \dots x_n)} = \sqrt{\frac{\sigma_1^2}{\sigma_2^2}} \cdot \exp\left(-\left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right) \underbrace{\sum_{j=1}^{n} (x_j - \mu)^2}_{T}\right)$$

$$= \sqrt{\frac{\sigma_1^2}{\sigma_2^2}} \exp\left(\frac{\sigma_2^2 - \sigma_1^2}{\sigma_1^2 \sigma_2^2} T\right)$$

$\Longrightarrow$ funzione crescente di T.

$\longrightarrow$ regione critica $C = \{T > d_\alpha\}$

$$ST = \frac{T}{\sigma^2} \sim \chi^2(n)$$

X

troviamo d_α:

$$\alpha = \mathbb{P}_{\sigma_0^2}(T > d_\alpha) = \mathbb{P}_{\sigma_0^2}\left(\frac{T}{\sigma_0^2} > \frac{d_\alpha}{\sigma_0^2}\right)$$

$$\Longleftrightarrow \quad \frac{d_\alpha}{\sigma_0^2} = \chi_{n,1-\alpha} \quad \Longleftrightarrow \quad d_\alpha = \sigma_0^2 \chi_{n,1-\alpha}$$

3. (Continua dal foglio precedente) Dato $n \geq 1$, sia $(X_1, X_2, \ldots, X_n)$ un campione estratto da una popolazione di densità

$$f_\theta(x) = \frac{1}{12\theta^5} x^9 e^{-x^2/\theta} \mathbb{1}_{(0,\infty)}(x),$$

dove $\mathbb{1}_{(0,\infty)}(x)$ rappresenta la funzione indicatrice sull'insieme $(0, \infty)$.

(a) Determinare lo stimatore Θ_n di massima verosimiglianza di θ.

(b) Formulare un test per l'ipotesi nulla $\mathcal{H}_0 : \theta \leq \theta_0$, ed ipotesi alternativa $\mathcal{H}_1 : \theta > \theta_0$ al livello di significatività α.

a) $\Theta_n = \frac{1}{5n} \sum_{j=1}^{n} X_j^2$

b) Verifichiamo che per $\theta_1 < \theta_2$

$$\frac{L_{\theta_2}(x_1, \ldots x_n)}{L_{\theta_1}(x_1, \ldots x_n)} = \frac{\theta_1^5}{\theta_2^5} e^{-\left(\frac{1}{\theta_2} - \frac{1}{\theta_1}\right) \sum_{j=1}^{n} x_j^2} = \frac{\theta_1^5}{\theta_2^5} e^{\frac{\theta_2 - \theta_1}{\theta_1 \theta_2} \sum_{j=1}^{n} x_j^2}$$

cresce strettamente in $T = \sum_{j=1}^{n} x_j^2$

$\Longrightarrow$ regione critica $c_\alpha = \{T > d_\alpha\}$ per livello α.

dal foglio precedente abbiamo:

$$Z_j = 2\frac{X_j^2}{\theta} \sim \Gamma\left(5, \frac{1}{2}\right) \Longrightarrow \sum_{j=1}^{n} Z_j = \frac{2}{\theta} \sum_{j=1}^{n} X_j^2 = \frac{2}{\theta} T \sim \chi^2(10n)$$

$$\Longrightarrow \alpha = P_{\theta_0}(T > d_\alpha) = P_{\theta_0}\Bigg(\underbrace{\frac{2T}{\theta_0}}_{\sum_{j=1}^{n} Z_j \sim \chi^2(10n)} > \frac{2 d_\alpha}{\theta_0}\Bigg)$$

$$\Longrightarrow \frac{2 d_\alpha}{\theta_0} = r_{10n, 1-\alpha} \iff d_\alpha = \frac{\theta_0}{2} r_{10n, 1-\alpha}.$$

4. (Continua dal foglio precedente) Dato $n \geq 1$, sia $(X_1, X_2, \dots, X_n)$ un campione estratto da una popolazione di densità
$$f_\theta(x) = \theta \exp(x - \theta \exp(x)), \qquad \text{per } \theta > 0\,.$$

(a) Determinare lo stimatore Θ_n di massima verosimiglianza di θ.

(b) Formulare un test per l'ipotesi nulla $\mathcal{H}_0 : \theta \leq \theta_0$, ed ipotesi alternativa $\mathcal{H}_1 : \theta > \theta_0$ al livello di significatività α.

a) dal foglio precedente: $\Theta_n = \left(\frac{1}{n}\sum_{j=1}^n e^{x_j}\right)^{-1}$

b) per $\vartheta_2 > \vartheta_1$ consideriamo

$$\frac{L_{\vartheta_2}(x_1, \dots, x_n)}{L_{\vartheta_1}(x_1, \dots, x_n)} = \frac{\vartheta_2}{\vartheta_1} \exp\left(-(\vartheta_2 - \vartheta_1)\sum_{j=1}^n \exp(x_j)\right)$$

$T = \left(\sum_{j=1}^n \exp(x_j)\right)^{-1}$

$$= \frac{\vartheta_2}{\vartheta_1} \exp\left(-(\vartheta_2 - \vartheta_1)\frac{1}{T}\right)$$

$\Longrightarrow$ rapporto di verosimiglianza crescente in T.

$\longrightarrow$ regione critica $C_\alpha = \{T > d_\alpha\}$.

Consideriamo la VA standardizzata

$$S = \vartheta \sum_{j=1}^n \exp(x_j) \sim \Gamma(n, 1) = \chi^2(2n)$$

$$\Longrightarrow \alpha = \mathbb{P}_{\vartheta_0}(T > d_\alpha) = \mathbb{P}_{\vartheta_0}\Big(\underbrace{\frac{1}{T}\cdot\vartheta_0}_{S} < \frac{\vartheta_0}{d_\alpha}\Big)$$

α-quantile di $\chi^2(2n)$

$$\Longleftrightarrow \frac{\vartheta_0}{d_\alpha} = \chi^2_{2n,\alpha} \Longleftrightarrow d_\alpha = \frac{\vartheta_0}{\chi^2_{2n,\alpha}}$$

5. Si consideri, per $a \in \mathbb{R}$ e $\theta > 0$ un campione $X_1, X_2, ..., X_n$ estratto da una popolazione di densità

$$f_{a,\theta}(x) = \frac{1}{\theta} e^{\frac{1}{\theta}(x-a)} \mathbb{1}_{[a,\infty)}(x).$$

(a) Determinare gli stimatori di massima verosimiglianza (A_n, Θ_n) dei parametri (a, θ).

(b) Per a costante, formulare un test unilatero con ipotesi

conosciuta

$$\mathcal{H}_0 : \theta \leq \theta_0, \qquad \mathcal{H}_1 : \theta > \theta_0$$

e livello $\alpha \in (0, 1)$.

(c) Stabilire se $(A_n)_{n \geq 2}$ è consistente.

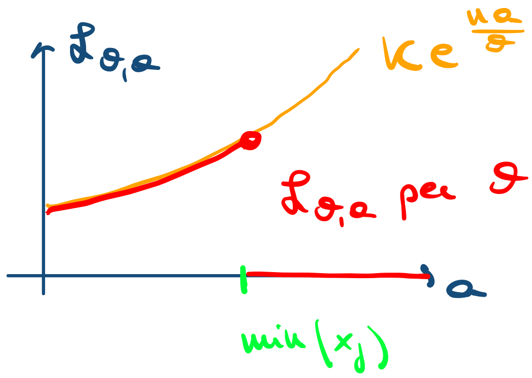
(d) Stabilire se $(\Theta_n)_{n \geq 2}$ è consistente.

a) $$\mathcal{L}_{a,\theta}(x_1, \dots x_n) = \prod_{j=1}^{n} \frac{1}{\theta} e^{-\frac{1}{\theta}(x_j - a)} \mathbb{1}_{[a,\infty)}(x_j)$$

$$= \frac{1}{\theta^n} e^{-\frac{1}{\theta}\left(\left(\sum_{j=1}^{n} x_j\right) - na\right)} \underbrace{\prod_{j=1}^{n} \mathbb{1}_{[a,\infty)}(x_j)}_{\mathbb{1}(a \leq \min x_j)}$$

$$= \frac{1}{\theta^n} \exp\left(-\frac{1}{\theta} \sum x_j + \frac{na}{\theta}\right) \mathbb{1}(a \leq \min(x_j))$$

$\Longrightarrow$ per θ fisso la funzione ha la forma



$\Longrightarrow$ massimo in $a = \min_j (x_j)$

$$\longrightarrow \ell_{\theta,\min(x_j)}(x_1 \dots x_n) = \log \mathcal{L}_{\theta, \min_j(x_j)} =$$

$$= -n \log \theta - \frac{1}{\theta} \sum_j x_j + \frac{n \min(x_j)}{\theta}$$

$\Longrightarrow$ minimizziamo in θ $\underset{\text{diff su } \theta}{\Longrightarrow}$ $\partial_\theta \ell_{\theta, \min x_j} = 0$

$$\Longleftrightarrow \quad -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_j x_j - \frac{1}{\theta^2}\, n \min(x_j) = 0$$

$$\Longleftrightarrow \begin{cases} \Theta_n = \dfrac{1}{n} \displaystyle\sum_{j=1}^{n} \left(x_j - \min_k x_k \right) \\ A_n = \min_k x_k \end{cases}$$

b) Verifichiamo il rapporto di verosimiglianza: $\theta_2 > \theta_1$

$$\frac{L_{\theta_2}(x_1 \dots x_n)}{L_{\theta_1}(x_1, \dots x_n)} = \frac{\theta_1}{\theta_2} \cdot e^{-\left(\frac{1}{\theta_2} - \frac{1}{\theta_1}\right) \underbrace{\left(\sum_{j=1}^{n} x_j - a\right)}_{T}}$$

$$= \frac{\theta_1}{\theta_2} \exp\left(\frac{\theta_2 - \theta_1}{\theta_1 \theta_2} T \right) \longrightarrow \text{strett cresc. in } T.$$

$\Longrightarrow$ regione critica è della forma

$$C_\alpha = \{ T > d_\alpha \}$$

Consideriamo la variabile standardizzata per $\theta = \theta_0$

$$\frac{T}{\theta_0} \sim \Gamma(n, 1) = \chi(2n) \sim S$$

$$\Longrightarrow \alpha = \mathbb{P}(S > x_{2n, 1-\alpha}) = \mathbb{P}(T > \underbrace{\theta_0 x_{2n, 1-\alpha}}_{d_\alpha})$$

c) $\mathbb{P}_{\theta,a}(|A_n - a| > \varepsilon) = \mathbb{P}(A_n > a+\varepsilon) =$

$$= \mathbb{P}(\min x_j > a+\varepsilon) = \mathbb{P}(x_j > a+\varepsilon)^n$$

$$= \Big(\underbrace{1 - \int_a^{a+\varepsilon} \frac{1}{\theta} e^{-\frac{1}{\theta}(x-a)}\, dx}_{>0}\Big)^n \xrightarrow{n\to\infty} 0$$

$\Longrightarrow$ consistente

d) $\mathbb{P}_{\theta,a}(|\Theta_n - \theta| > \varepsilon) = \mathbb{P}_{\theta,a}\left(\left|\frac{1}{n}\sum_{j=1}^{n}(X_j - \min_{e} X_e) - \theta\right| > \varepsilon\right)$

$$= \mathbb{P}_{\theta,a}\left(\left|\frac{1}{n}\sum_{j=1}^{n}(X_j - a) - \theta + a - \min_{e} X_e\right| > \varepsilon\right)$$

$$\leq \mathbb{P}_{\theta,a}\left(\left\{\left|\frac{1}{n}\sum(x_j - a) - \theta\right| > \frac{\varepsilon}{2}\right\} \cup \left\{\left|a - \min_{e} X_e\right| > \frac{\varepsilon}{2}\right\}\right)$$

$$\leq \mathbb{P}_{\theta,a}\left(\left|\frac{1}{n}\sum_{j=1}^{n}\underbrace{(x_j - a)}_{Y_j \sim Exp(\frac{1}{\theta})} - \underbrace{\theta}_{\mathbb{E}(Y_j)}\right| > \frac{\varepsilon}{2}\right) + \mathbb{P}_{\theta,a}\left(|A_n - a| > \frac{\varepsilon}{2}\right)$$

$$= \underbrace{\mathbb{P}_{\theta,a}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mathbb{E}(X_j)\right| > \frac{\varepsilon}{2}\right)}_{\xrightarrow{n\to\infty} 0} + \underbrace{\mathbb{P}_{\theta,a}\left(|A_n - a| > \frac{\varepsilon}{2}\right)}_{\xrightarrow{n\to\infty} 0}$$

(legge grandi num.)

$\xrightarrow[n\to\infty]{} 0 \quad \Longrightarrow$ consistente

6. * Sia dato un campione $X_1, \dots, X_n$ con gegge di Poisson con parametro $\lambda > 0$.

(a) Trovare la stima di massima verosimiglianza $\hat{\lambda}_n$ del parametro λ.

(b) Questa stima è corretta? asintoticamente corretta?

(c) La successione di stime $(\hat{\lambda}_n)_n$ è consistente?

(d) Trovare un intervallo di confidenza a livello $1-\alpha$ per il parametro λ.

(e) Esaminare il test unilatero di ipotesi nulla $H_\theta : \lambda \leq 1$ e ipotesi alternativa $\mathcal{H}_1 : \lambda > 1$ al livello α.

(f) Esaminare il test di ipotesi nulla $H_\theta : \lambda = 2$ e ipotesi alternativa $\mathcal{H}_1 : \lambda \neq 2$ al livello α.

a) $f_\theta(x_j) = e^{-\lambda} \dfrac{\lambda^{x_j}}{x_j!} \Longrightarrow \mathcal{L}_\lambda(x_1, \dots x_n) = e^{-n\lambda} \dfrac{\lambda^{\sum_{j=1}^n x_j}}{\prod_{j=1}^n x_j!}$

$\longrightarrow \ell_\lambda(x_1, \dots x_n) = \log(\mathcal{L}_\lambda(x_1, \dots x_n)) = -n\lambda + \log \lambda \cdot \sum_{j=1}^n x_j - \sum_{j=1}^n \log x_j!$

$\overset{\partial_\lambda = 0}{\Longrightarrow} \quad -n + \dfrac{1}{\lambda} \sum_{j=1}^n x_j = 0 \quad \Longleftarrow \quad \hat{\lambda} = \dfrac{1}{n} \sum_{j=1}^n x_j$

b) $\mathbb{E}_\lambda\left(\frac{1}{n}\sum_{j=1}^n X_j\right) = \mathbb{E}_\lambda(X_j) = \lambda \Longrightarrow$ corretta e asint. corretta.

c) $\mathbb{P}_\lambda\left(|\hat{\lambda}(x_1, \dots x_n) - \lambda| > \varepsilon\right) = \mathbb{P}_\lambda\left(\left|\frac{1}{n}\sum_{j=1}^n X_j - \mathbb{E}_\lambda(X_j)\right| > \varepsilon\right) \xrightarrow[LLN]{n \to \infty} 0$

$\Longrightarrow$ lo stimatore è consistente.

d) Approssimativamente $\overline{X}_n \overset{n \gg 1}{\approx} \mathcal{N}\left(\lambda, \frac{\lambda}{n}\right)$

$\Longrightarrow \dfrac{\overline{X}_n - \lambda}{\sqrt{\lambda/n}} \approx Z \sim \mathcal{N}(0,1)$

$\Longrightarrow$ Sia (z_α^-, z_α^+) un IC al livello $1-\alpha$ per Z

$1-\alpha = \mathbb{P}(Z \in (z_\alpha^-, z_\alpha^+)) = \mathbb{P}\left(\dfrac{\overline{X}_n - \lambda}{\sqrt{\lambda/n}} \in (z_\alpha^-, z_\alpha^+)\right)$

$\lambda \approx \overline{X}_n$

$\simeq \mathbb{P}\left(\lambda \in \left(\overline{X}_n - z_\alpha^+ \sqrt{\frac{\overline{X}_n}{n}}, \overline{X}_n + z_\alpha^- \sqrt{\frac{\overline{X}_n}{n}}\right)\right)$

e) il rapporto di verosimiglianza per $\lambda_1 < \lambda_2$

$$\frac{L_{\lambda_2}(x_1 \dots x_n)}{L_{\lambda_1}(x_1 \dots x_n)} = e^{-(\lambda_2-\lambda_1)} \left(\frac{\lambda_2}{\lambda_1}\right)^{\sum_{j=1}^{n} x_j}$$

è crescente rispetto a $T = \sum_{j=1}^{n} x_j$

$\Longrightarrow$ regione critica $C_\alpha = \{T > d_\alpha\}$ al livello α

asintoticamente $\frac{1}{n}\sum x_j \underset{n>>1}{\overset{\text{approx}}{\sim}} N\left(\lambda, \sqrt{\frac{\lambda}{n}}\right)$

$$\Longrightarrow \alpha = \mathbb{P}_{\lambda_0}\left(\sum_{j=1}^{n} x_j > d_\alpha\right) = \mathbb{P}_{\lambda_0}\left(\underbrace{\frac{T-\lambda_0}{\sqrt{\lambda_0/n}}}_{\simeq Z \sim N(0,1)} > \frac{d_\alpha-\lambda_0}{\sqrt{\lambda_0/n}}\right)$$

$$\Longrightarrow \frac{d_\alpha-\lambda_0}{\sqrt{\lambda_0/n}} = q_{1-\alpha} \iff d_\alpha = \lambda_0 + \sqrt{\frac{\lambda_0}{n}}\cdot q_{1-\alpha}$$

f) Siccome sotto H_0 $\quad \frac{\bar{X}-\lambda_0}{\sqrt{\lambda_0/n}} \approx Z \sim N(0,1)$

$$\alpha = \mathbb{P}_{\lambda_0}\left(|Z| > q_{1-\frac{\alpha}{2}}\right) \hat{=} \mathbb{P}_{\lambda_0}\left(\left|\frac{\bar{X}-\lambda_0}{\sqrt{\lambda_0/n}}\right| > q_{1-\frac{\alpha}{2}}\right)$$

$$= \mathbb{P}_{\lambda_0}\left(\bar{X} \in \underbrace{\left(\lambda_0 - \sqrt{\frac{\lambda_0}{n}}\,q_{1-\frac{\alpha}{2}}, \lambda_0 + \sqrt{\frac{\lambda_0}{n}}\,q_{1-\frac{\alpha}{2}}\right)}_{\tilde{C}}\right)$$

$$\Longrightarrow C_\alpha = \{\omega \in \Omega : \bar{X} \in \tilde{C}\}$$