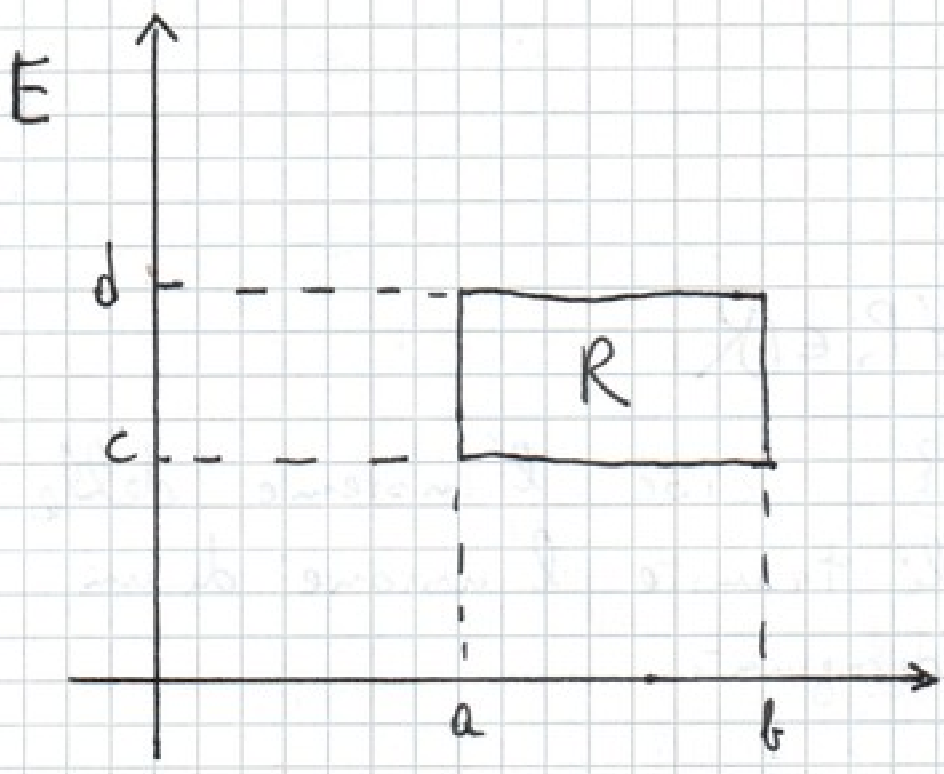
Teoria della Misura

Cominciamo col misurare i rettangoli



Intuitivamente la misura di R sarà:

$$m(R) = (b-a)(d-c)$$

Chiamiamo:

$$\mathcal{R} = \{\text{rettangoli in } E\}$$

la misura:

$$m : \mathcal{R} \to \mathbb{R}$$

$$\forall R \in \mathcal{R} \Rightarrow m(R) \geq 0$$

Posso anche misurare R considerandolo come una unione di rettangoli: $R = \bigcup R_i$

$$\{R_i\}^n_{i=1\dots n} \qquad R_i \cap R_j = \emptyset \quad \forall i \neq j$$

$$m\left(\bigcup_{i=1}^{n} R_i\right) = m(R)$$

Una collezione non vuota di insiemi $\mathcal{R}$ è detta ANELLO se soddisfa le seguenti condizioni:

- dati $A, B \in \mathcal{R} \Rightarrow A \Delta B \in \mathcal{R}$ e $A \cap B \in \mathcal{R}$

Ciò implica anche che $A \cup B \in \mathcal{R}$ e $A \setminus B \in \mathcal{R}$ poiché:

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

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

Ciascun anello contiene l'insieme vuoto $\emptyset$, poiché $A \setminus A = \emptyset$

Una collezione di insiemi G è un semianello sse:

- $\emptyset \in G$
- dati $A, B \in G \Rightarrow A \cap B \in G$
- dati $A, A_1 \in G$ con $A_1 \subseteq A \Rightarrow A = \bigcup_{i=1}^{n} A_i$, $A_i \in G$, $A_i \cap A_j = \emptyset \ \forall i, j$

Ciascun anello è anche un semianello, non viceversa.

Come esempio di un semianello che non è un anello possiamo considerare l'insieme di tutti gli intervalli (a, b), $[a, b]$, $(a, b]$ e $[a, b)$.

LEMMA 1: Siano gli insiemi $A_1, A_2, \dots, A_n, A \in G$ con $A_j \cap A_l = \emptyset$ $j \neq l$, $j, l = 1 \dots n$ $\forall A_i \subseteq A$ allora gli insiemi A_i $(i = 1 \dots n)$ possono essere inclusi come primi n elementi della decomposizione finita:

$$A = \bigcup_{k=1}^{s} A_k \qquad s \geq n$$

dell'insieme A, dove tutti gli $A_k \in G$

Lemma 2:

$$A_1, A_2, \dots, A_n \in G$$

$$\exists\, B_1, B_2, \dots, B_t \in G \text{ tale che } A_k = \bigcup_{s \in M_k} B_s$$

Th: Dato un insieme X e una famiglia $\mathcal{F}$ di sottoinsiemi di X

$$\exists!\ \mathcal{R}(\mathcal{F}):\quad ①\ \mathcal{R}(\mathcal{F}) \supseteq \mathcal{F}$$

$$②\ \mathcal{R}(\mathcal{F}) \subseteq \forall\, \mathcal{R}^*(\mathcal{F}) \supseteq \mathcal{F}$$

In altre parole: $\mathcal{R}(\mathcal{F})$ è l'anello minimale di $\mathcal{F}$; $\mathcal{R}(\mathcal{F})$ è generato da $\mathcal{F}$

Dim:

$\mathcal{P}(X)$ insieme delle parti

$\mathcal{P}(X) \in \mathcal{E}$, $\mathcal{P}(X) \supseteq \mathcal{F}$

Definiamo $\mathcal{X}$ come l'insieme di tutti gli anelli in $\mathcal{P}(X)$ che contengono $\mathcal{F}$

$\mathcal{X}$ è non vuoto, poiché contiene almeno $\mathcal{P}(X)$

Definiamo:

$$A^* = \bigcap_{A_i \in \mathcal{X}} A_i \implies \mathcal{F} \subseteq A^*$$

Devo dimostrare che:

1) A^* è un anello
2) A^* è il più piccolo anello che include $\mathcal{F}$

1) A^* è chiaramente un anello perché è intersezione di anelli.

2) Dato un anello $\mathcal{A} \supseteq \mathcal{F}$ per definizione di $\mathcal{X}$

$\mathcal{A} \in \mathcal{X}$

Definiamo:

$$\mathcal{A}' = \mathcal{A} \cap \mathcal{P}(X)$$

Ne segue che: $\mathcal{A}' \subseteq \mathcal{A}$

$$\underset{\substack{\uparrow \\ \text{perché è intersezione di anelli}}}{\mathcal{A}' \in \mathcal{X}} \Rightarrow \mathcal{A}' \geq \bigcap_{i} \mathcal{A}_i \quad \mathcal{A}_i \in \mathcal{X}$$

Quindi:

$$\mathcal{F} \subseteq \mathcal{A}^* \subseteq \mathcal{A}' \subseteq \mathcal{A}$$

CVD

Th: Se la famiglia $\mathcal{F}$ è un semianello

$\mathcal{R}(\mathcal{F}) = \{$tutti i sottoinsiemi di X che ammettono una partizione finita$\}$

ovvero:

$$X \supseteq A = \bigcup_{j=1}^{n} \underset{\substack{\uparrow \\ \text{semianello}}}{A_j} \in \mathcal{F}$$

Possiamo, adesso, definire la misura in modo alternativo:

$m : \mathcal{S}_m \to \mathbb{R}$

1) $\mathcal{S}_m$ è un semianello

2) $\forall A \in \mathcal{S}_m$, $m(A) \geq 0$

3) $A = \bigcup_{i=1}^{n} A_i$ con $A_i \cap A_j = \emptyset$ $\forall i \neq j \Rightarrow m(A) = \sum_{i=1}^{n} m(A_i)$

Th: Qualsiasi misura $m : \mathcal{S}_m \to \mathbb{R}$ ammette un'unica estensione a $\mathcal{R}(\mathcal{S}_m)$

Dim:

Dobbiamo dimostrare che detto

$$A \in \mathcal{R}(\mathcal{S}_m) \Rightarrow \mu(A) = \sum_{i=1}^{n} m(A_i)$$

dove

$$A = \bigcup_{i=1}^{n} \underset{\mathcal{S}_m}{\underset{\in}{A_i}} \qquad A_i \cap A_j = \emptyset \quad \forall i \neq j$$

E' banale vedere che $\mu(A)$ non dipende dalla partizione di A che scelgo e anche che è una funzione positiva e additiva

Non resta da dimostrare che sia l'unica estensione:

supponiamo per assurdo che ne esista un'altra che chiamiamo μ' estensione di m:

$$\mu'(A) = \sum_{j=1}^{n} \mu'(B_j) = \sum_{j=1}^{n} m(B_j) = \mu(A)$$

CVD

Def: Sia $m : \mathcal{S}_m \to \mathbb{R}$, m è σ-additiva sse:

$$A_1, A_2, \dots \in \mathcal{S}_m \qquad A_i \cap A_j = \emptyset \quad \forall i \neq j$$

$$A = \bigcup_n A_n \Rightarrow m(A) = \sum_n m(A_n)$$

Th: Se $m : \mathcal{S}_m \to \mathbb{R}$ e m è σ-additiva allora la sua estensione a $\mathcal{R}(\mathcal{S}_m)$ è σ-additiva

Possiamo procedere e ridefinire la misura di Lebesgue data $m : \mathcal{R}(\mathcal{S}_m) \to \mathbb{R}$ e $A \subseteq X$

$$\mu^*(A) = \inf \{ \bigcup_{\substack{\in \\ \mathcal{R}(\mathcal{S}_m)}} B_i \supseteq A, \sum m(B_i) \}$$

segue come per i rettangoli ...

Definiamo come:

$\mathcal{M} = \{$ tutti i sottoinsiemi di X approssimabili bene a piacere con elementi di $\mathcal{R}(\mathcal{S}_X) \}$

$\mathcal{M}$ è una σ-algebra

CONTINUITÀ DELLA MISURA   1° lemma Borel-Cantelli

Th: Sia data una successione monotona decrescente

$$A_1 \supset A_2 \supset A_3 \supset \dots \quad \text{tale che} \quad A_i \in \mathcal{M}$$

definiamo $A = \bigcap_n A_n \Rightarrow \mu(A) = \lim_{n \to +\infty} \mu(A_n)$

DIM:

- Se $A = \emptyset \Rightarrow \mu(A) = 0$

  Possiamo scrivere:

$$A_i = (A_i \setminus A_{i+1}) \cup (A_{i+1} \setminus A_{i+2}) \cup \dots$$

$$\mu(A_1) = \sum_{n=1}^{\infty} \mu(A_n \setminus A_{n+1})$$

$$\mu(A_k) = \sum_{n=k}^{\infty} \mu(A_n \setminus A_{n+1})$$

converge perchè A_k è misurabile e perchè il residuo converge

$$\mu(A_k) \xrightarrow[k \to +\infty]{} 0 = \mu(A)$$

- Se $A \neq \emptyset$

  $A'_i = A_i \setminus A$

  si procede come prima

DEF: σ-finitezza

$$S_1 \subseteq S_2 \subseteq S_3 \subseteq \dots$$

$S_i \in \mathcal{A}$ con $\mathcal{A}$ σ-algebra

$\mu : \mathcal{A} \to \mathbb{R}$ è σ-finita sse

$$\mu(S_m) < +\infty \quad \forall m \in \mathbb{N}$$

DEF: Dati due insiemi X e Y e $\mathcal{S}_X$, $\mathcal{S}_Y$ siano famiglie di sottoinsiemi costruite su queste, sia:

$$f : X \to Y$$

f è $(X, \mathcal{S}_X), (Y, \mathcal{S}_Y)$-misurabile sse:

$$\text{dato } A \in \mathcal{S}_Y \Rightarrow f^{-1}(A) \in \mathcal{S}_X$$

Th: Affinché una funzione f sia μ-misurabile è necessario e sufficiente che comunque preso $c \in \mathbb{R}$ l'insieme:

$$\{x : f(x) < c\} \in \mathcal{S}_\mu$$

ovvero sia μ-misurabile

DIM:

- f è μ-misurabile $\Rightarrow \{x : f(x) < c\} \in \mathcal{S}_\mu$

  $f^{-1}(A) \in \mathcal{S}_\mu \quad \forall A \in \mathbb{B}_{\mathbb{R}}$

  $f^{-1}((-\infty, c), c \in \mathbb{R}) \in \mathcal{S}_\mu$ Poiché $(-\infty, c) \in \mathbb{B}_{\mathbb{R}}$ allora è vero

- Se $\forall c \in \mathbb{R}$ $\{x : f(x) < c\} \in \mathcal{S}_\mu \Rightarrow f$ è μ-misurabile

  Prendiamo $\Sigma = \{(-\infty, c), c \in \mathbb{R}\} \Rightarrow \mathcal{R}(\Sigma) = \mathbb{B}_{\mathbb{R}}$

  Per ipotesi vale che $f^{-1}(\Sigma) \in \mathcal{S}_\mu$

possiamo scrivere:

$$\underbrace{f^{-1}(\mathcal{R}(\Sigma)) = \mathcal{R}(f^{-1}(\Sigma))}_{\text{da dimostrare}} \subseteq \mathcal{R}(S_\mu) \underset{\text{perché } S_\mu \text{ è una } \sigma\text{-algebra minimale}}{=} S_\mu$$

quindi f è μ-misurabile

CVD

Def: FUNZIONE SEMPLICE

Una funzione $f: X \to \mathbb{R}$ si dice semplice se:

1) f è (X, S_μ)-$(\mathbb{R}, \mathbb{B}(\mathbb{R}))$ misurabile
2) f assume al più una quantità numerabile di valori distinti: $y_1, y_2, y_3, \ldots$

Th: La funzione f per cui vale la (2) è μ-misurabile sse tutti gli insiemi:

$$A_n = \{ x : f(x) = y_n \}$$

sono μ-misurabili.

Dim:

- NECESS: f è μ-mis $\Rightarrow$ A_n è μ-mis

$$y_n \in \mathbb{B}_\mathbb{R} \Longrightarrow \underbrace{f^{-1}(\{y_n\})}_{= A_n} \in S_\mu$$

- SUFF: A_n è μ-misurabile $\Rightarrow$ f è μ-misurabile

Dato $C \in \mathbb{R}$

$$f^{-1}(C) = f^{-1}(C \cap (\{y_1\} \cup \{y_2\} \cup \ldots)) = f^{-1}(C \cap \{y_1\}) \cup$$
$$\cup f^{-1}(C \cap \{y_2\}) \cup \ldots = A_1 \cup A_2 \cup \ldots =$$
$$= \bigcup_j A_j \in S_\mu$$

CVD

Convergenza Puntuale

$\{f_n(x)\}_{n\in\mathbb{N}}$ è puntualmente convergente ad una funzione f su un insieme A sse

$$\forall x \in A,\ \forall \varepsilon > 0 \quad \exists\, \tilde{n} = \tilde{n}(\varepsilon, x) : \forall n > \tilde{n}$$

$$|f_n(x) - f(x)| < \varepsilon$$

WIKIPEDIA

$$\lim_{n\to\infty} f_n(x) = f(x)$$

Convergenza Uniforme

$\{f_n(x)\}_{n\in\mathbb{N}}$ è uniformemente convergente ad una funzione f su un insieme A sse:

$$\forall \varepsilon > 0 \quad \exists\, \tilde{n} : \tilde{n}(\varepsilon) : \forall n > \tilde{n}$$

$$a_n = \sup_{x\in A} |f_n(x) - f(x)|$$

$$a_n < \varepsilon$$

Uniforme più forte, comprende anche la puntuale

WIKIPEDIA

$$\lim_{n\to\infty} \sup_{x\in A} \{f_n(x) - f(x)\}^2 = 0$$

Esempio

$f_n(x) = x^n \qquad f(x) = 0 \qquad A = [0, 1]$

converge puntualmente?

$$\lim_{n\to\infty} f_n(x) = \begin{cases} 0 & \text{se } x \in [0,1) \\ 1 & \text{se } x = 1 \end{cases}$$

in questi punti converge a $f(x)$

converge uniformemente?

$$\lim_{n\to\infty} \sup_{x\in A} \{x^n - 0\} = 1 \neq 0$$

non converge uniformemente

Th: Le funzioni μ-misurabili sono tutte e solo quelle che si possono rappresentare come limite di successioni di funzioni semplici uniformemente convergenti

---

$\mathcal{L} = \{ f : X \to \mathbb{R} ,\ f$ sia μ-misurabile $\}$

$\mathcal{L}$ è chiuso sotto:

1) SOMMA $(f+g) \in \mathcal{L}$
2) PRODOTTO $(fg) \in \mathcal{L}$
3) COMPOSIZIZIONE $(f \circ g) \in \mathcal{L}$ sse $X = \mathbb{R}$ e μ è la misura di Lebesgue sui boreliani

(3) Th: date le funzioni $f : X \to \mathbb{R}$ μ-misurabile e $g : \mathbb{R} \to \mathbb{R}$ $\mathbb{B}_{\mathbb{R}}$-misurabile allora $g \circ f$ è μ-misurabile

DIM

$D \in \mathbb{B}_{\mathbb{R}} \Rightarrow g^{-1}(D) \in \mathbb{B}_{\mathbb{R}}$

$f^{-1}(g^{-1}(D)) \in \mathcal{S}_\mu$

CVD

(2) $fg = \frac{1}{4}\left((\underbrace{f+g}_{h})^2 - (\underbrace{f-g}_{K})^2 \right)$

diamo per scontato che la somma sia μ-misurabile

h e K sono come la composizione $s(x) = x^2$

Voglio misurare oltre oltre rettangoli.

Il primo passo è estendere m a una misura m' che abbia queste proprietà:

- $m' : \mathcal{E} \to \mathbb{R}$
- $\mathcal{E} \supseteq \mathcal{R}$
- $m'(R) = m(R) \qquad \forall R \in \mathcal{R}$

dove $\mathcal{E} = \{$insiemi elementari$\}$ cioè l'insieme delle regioni del piano descrivibili tramite l'unione di un numero finito di rettangoli disgiunti.

Th: $\mathcal{E}$ è chiuso sotto operazioni di $\cup, \cap, \setminus, \Delta$

Dim:

1) $\mathcal{E}$ è chiuso per $\cap$

$A, B \in \mathcal{E} \implies A \cap B \in \mathcal{E}$

Partizioniamo A e B in rettangoli:

$A = \bigcup_k P_k \qquad B = \bigcup_j Q_j$

$$A \cap B = \bigcup_{k,j} (P_k \cap Q_j)$$

È chiaro che l'intersezione di due rettangoli è un rettangolo, quindi: $P_k \cap Q_j \in \mathcal{R}$

L'unione di ~~un~~ numero finito di rettangoli, come da definizione, appartiene ad $\mathcal{E}$

2) $\mathcal{E}$ è chiuso per $\setminus$

$E \setminus A \in \mathcal{E} \qquad A \setminus B = A \cap (E \setminus B)$

Togliendo da (un) rettangolo (un) insieme elementare otteniamo un insieme elementare

DEF: EQUIVALENZA tra funzioni

Date $g, f : X \to \mathbb{R}$ μ-misurabili diciamo che

$$g \sim f \iff \mu(\{x \in X : f(x) \neq g(x)\}) = 0$$

equivalente

Th: Se $f : X \to \mathbb{R}$ è μ-mis e $g \sim f \Rightarrow g$ è μ-mis

DEF:

Una funzione semplice f è pesabile su A sse la serie:

$$\sum_i y_i \, \mu\left(f^{-1}\left(f(A) \cap \{y_i\}\right)\right) < +\infty$$

converge

Se questo è vero allora possiamo scrivere:

$$v(f, A) = \sum_i y_i \, \mu\left(f^{-1}\left(f(A) \cap \{y_i\}\right)\right)$$

Se $A = \bigcup_i A_i$ $\quad \{A_i\}_n$ partizione di A

$$v\left(f, \bigcup_i A_i\right) = \sum_i v(f, A_i)$$

Definiamo integrale di Lebesgue per le funzioni semplici:

$$\int_A f \, d\mu = v(f, A) = \sum_i y_i \, \mu\left(f^{-1}\left(f(A) \cap \{y_i\}\right)\right) =$$

$$= \sum_i y_i \, \mu(A_n)$$

dove $A_n = \{x : x \in A, f(x) = y_n\}$

Th: $\int_A \cdot \, d\mu$ è positivo, cioè $f \geq 0 \Rightarrow \int_A f \, d\mu \geq 0$

Th: $\int_A \cdot \, d\mu$ è monotono, cioè se $f \geq g$ su A $\Rightarrow \int_A f \, d\mu \geq \int_A g \, d\mu$

DEF: $\mathcal{M}_A = \left\{ f : X \to \mathbb{R} , \int_A f \, d\mu \right\}$

Th: Dato $A = \bigcup_{i=1}^{+\infty} B_i \qquad B_j \cap B_i = \emptyset \quad \forall i \neq j$

$$\int_A f \, d\mu = \sum_{i=1}^{+\infty} \int_{B_i} f \, d\mu$$

Dim:

Definiamo $B_{nk} = \{ x \in B_n : f(x) = y_k \}$

$$\int_A f \, d\mu = \sum_k y_k \, \mu(B_k) = \sum_k y_k \, \mu\left(\bigcup_n B_{nk}\right) =$$

$$= \sum_k y_k \sum_n \mu(B_{nk}) = \sum_n \sum_k y_k \, \mu(B_{nk}) =$$

$$= \sum_n \int_{B_n} f \, d\mu$$

CVD ne caso f sia semplice

Valore Atteso

$Z : \Omega \to \mathbb{R}$ è una v.c., una funzione $(\Sigma, \mathbb{B}_{\mathbb{R}})$-misurabile

Il valore atteso sarà:

$\int_\Omega Z \, dP$ dove P è la funzione di probabilità

Proprietà:

1) Lineare: $E(\lambda f + \gamma g) = \lambda E(f) + \gamma E(g)$

2) Positivo: vero, si vede 3 teoremi indietro

3) Additivo: vero, si vede teorema precedente.

Dimostriamo adesso le tre disuguaglianze presenti nel Mod utilizzando solo le proprietà del valore atteso:

DISUGUAGLIANZA DI CHEBYCHEV

$X \in \mathcal{M}$, $g(X) \geq 0$, $\forall c > 0 \implies$

$$\implies P(g(x) \geq c) \leq \frac{E(g(x))}{c}$$

DIM

Prendiamo una g che $\forall c$ partiziona Ω in 2 sottoinsiemi:

$\Omega_+ = \{\omega : g(X(\omega)) \geq c\}$ $\qquad g^+ = g_{|\Omega_+}$ ristretto

$\Omega_- = \{\omega : g(X(\omega)) < c\}$ $\qquad g^- = g_{|\Omega_-}$

$E(g(x)) = E(g^+(x)) + E(g^-(x)) =$ per la linearità

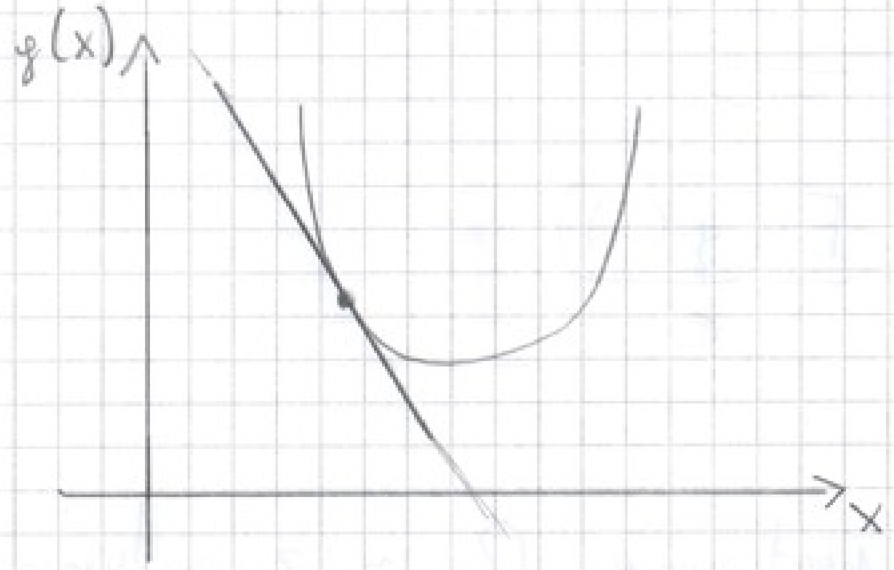
$\geq E(g^+(x))$ perché sono positive

continua ...

Disuguaglianza di Jensen

$g: \mathbb{R} \to \mathbb{R}$ convessa $\Rightarrow E(g(x)) \geq g(E(x))$

Dim



Prendiamo una retta tangente a $g(x)$ che sta sempre sotto il grafico.

Prendiamo la retta $L(x) = mx + q$ che interseca $g(x)$ nel punto $(E(x), g(E(x)))$

Allora vale che:

$$L(E(x)) = g(E(x))$$

$$E(L(x)) \underset{\uparrow}{=} q + E(mX) = q + m\,E(X) = L(E(x)) =$$

per linearità

$$= g(E(x)) \underset{\uparrow}{\leq} E(g(x))$$

questo per la positività di E poiché $L(x) \leq g(x)$, abbiamo scelto $L(x)$ in modo che stia sempre sotto il grafico di $g(x)$

CVD

Disuguaglianza di Cauchy-Schwarz

Se $E(XY)$ esiste

1) $|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$

oppure

$$(E(XY))^2 \leq E(X^2)E(Y^2)$$

2) L'uguaglianza vale se $X = cY$ per qualche $c \in \mathbb{R}$

Dim

$\varphi : \mathbb{R} \to \mathbb{R}$

1) $\varphi(t) = E((X-tY)^2) = E(X^2 - 2tXY + t^2Y^2) =$

$$= E(X^2) - 2tE(XY) + t^2E(Y^2)$$

Questo polinomio non ammette soluzioni reali perché è >0

allora: $\Delta = (b^2 - 4ac) < 0$

Allora possiamo scrivere:

$$E(XY)^2 - E(X^2)E(Y^2) < 0$$

$$E(XY)^2 < E(X^2)E(Y^2)$$

2) $\varphi(t) = E((X-tY)^2)$

$\varphi(t) = 0$ quando $X = tY$

non ci interessa proprio l'uguaglianza ma l'equivalenza:

$$X \sim tY \Rightarrow P(\{X - tY \neq 0\}) = 0$$

CVD

Spazio L_1

$$L_1 = \{ f : X \to \mathbb{R} ,\ \exists \int_A f \, d\mu \}$$

L_1 è chiuso per:

- Somma $\quad f, g \in L_1 \Rightarrow f+g \in L_1$
- Prodotto per uno scalare:

$$\lambda \in \mathbb{R},\ f \in L_1 \Rightarrow \int \lambda f \, d\mu = \lambda \int f \, d\mu$$

Non è chiuso per:

- Prodotto

$$f, g \in L_1 \Rightarrow f \cdot g \in L_1 \qquad \text{No}$$

Controesempio:

$$f(x) = \frac{1}{\sqrt{x}} \qquad g(x) = \frac{1}{\sqrt{x}}$$

- Composizione

Vale solo se g è concava

Def: Spazio Vettoriale

Dato V un insieme non vuoto ed un campo K (es. $\mathbb{R}$ e $\mathbb{C}$)

V è uno spazio vettoriale su K sse

1) La somma è interna, ovvero $(V, +)$ è un gruppo abeliano

2) Esiste il prodotto esterno $K \times V \to V$

Da ciò si evince che L_1 è uno spazio vettoriale su $\mathbb{R}$

DEF: NORMA

$\| \cdot \| : V \longrightarrow \mathbb{R}^+$

1) $\|x\| \geq 0 \quad \forall x \in V$
2) $\|x\| = 0$ sse $x \equiv 0$ identicamente 0
3) Soddisfa la disuguaglianza triangolare

$$\|x+y\| \leq \|x\| + \|y\| \qquad \forall x, y \in V$$

Provo a definire su L_1 la norma:

$\forall f \in L_1$

$$\|f\| = \int_X |f| \, d\mu$$

Il modulo mi serve perché la norma per definizione è sempre ≥ 0

Manca qualcosa!

DEF: $f \sim g$ sse $\mu(\{x \in X : f(x) \neq g(x)\}) = 0$

La norma di funzioni non identicamente zero può essere 0

Introduco la classe di equivalenza $\sim$ su L_1

Cioè quando considero una $f \in L_1$, prendo f e anche tutte le funzioni $\sim$ equivalenti.

Su questo L_1 la norma è ben definita, quindi L_1 è uno SV normato.

Vediamo il prodotto interno:

$f \cdot g \in L_1$? NO

Per cauchy-swarz $\left| \int_X f g \, d\mu \right| \leq \sqrt{\int_X f^2 d\mu \int_X g^2 d\mu}$

Quindi $f \cdot g$ è integrabile se f^2 e g^2 sono integrabili

SPAZIO L_2

$$L_2 \equiv \left\{ f : X \to \mathbb{R} \;:\; \exists \int_X f^2 \, d\mu \right\}$$

$$L_2 \subset L_1$$

L_2 è uno SV

DEF: PRODOTTO INTERNO

$f, g \in L_2$

$$(f, g) = \int_X fg \, d\mu$$

Def: Norma su L_2

$$\|f\| = \sqrt{(f, f)} = \sqrt{\int_X f^2 \, d\mu}$$

L_2 è uno spazio euclideo perché è uno spazio vettoriale dotato di prodotto interno

$L_1(\Omega, P)$: variabili casuali aventi valore atteso

$L_2(\Omega, P)$: variabili casuali aventi momento secondo finito e varianza

DEF: VARIANZA

$X \in L_2(\Omega, P)$

$var(X) = \| (X - E(X)\mathbb{1}) \|^2$

DEF: DEVIAZIONE STANDARD

$$\| X - E(X)\mathbb{1} \|$$

dove $\mathbb{1}$ è il rappresentante delle v.c. costanti:

$C(\omega) = c \quad \forall \omega \in \Omega \Rightarrow C = c \cdot \mathbb{1}(\omega) = 1 \cdot c$

Possiamo dire che: $E(X) = (\mathbb{1}, X)$

I vettori $\mathbb{1}$ e $X - E(X)\mathbb{1}$ sono ortogonali:

$(\mathbb{1}, X - E(X)\mathbb{1}) = (\mathbb{1}, X) - (\mathbb{1}, E(X)\mathbb{1}) = (\mathbb{1}, X) - E(X)(\mathbb{1}, \mathbb{1}) =$
$= E(X) - E(X) = 0$

DEF: COVARIANZA

$X, Y \in L_2(\Omega, P)$

$cov(X, Y) = ((X - E(X)\mathbb{1}), (Y - E(Y)\mathbb{1}))$

Indica quanto del vettore Y può essere espresso in termini di X

DEF: COEFFICIENTE DI CORRELAZIONE

$$\rho(X, Y) = \frac{cov(X, Y)}{\| X - E(X)\mathbb{1} \| \cdot \| Y - E(Y) \|}$$

Varia da -1 a 1 per Cauchy-Swartz:

$$|((X - E(X)\mathbb{1}, (Y - E(X)\mathbb{1}))| \leq \| X - E(X)\mathbb{1} \| \cdot \| Y - E(Y)\mathbb{1} \|$$

Posso esprimerlo come $\rho(X, Y) = \cos\alpha$

3) $\mathcal{E}$ è chiuso per $\cup$

$A, B \in \mathcal{E} \Rightarrow A \cup B \in \mathcal{E}$

$A \cup B \equiv (A^c \cap B^c)^c \equiv E \setminus (E \setminus A \cap E \setminus B)$

poichè $E \setminus A = A^c$ e $E \setminus B = B^c$.

Essendo $\mathcal{E}$ chiuso per $\cap$ e $\setminus$ allora CVD

4) Δ è esprimibile tramite $(A \cup B) \setminus (A \cap B)$

Ritorniamo ad m', è definita:

$$A = \bigcup_K R_K \qquad R_K \in \mathcal{R}$$

$$m'(A) = \sum_K m(R_K) \qquad R_K \cap R_j = \phi \quad \forall k \neq j$$

Affinchè m' sia una <u>buona misura</u>, non deve dipendere dalla partizione di rettangoli di un elemento del dominio.

Th: m' è una buona misura

Dim:

Siano $\bigcup_K P_K$ e $\bigcup_j Q_j$ due partizioni di A

$$m'(A) \;\; \begin{cases} m'\left(\bigcup_K P_K\right) = \sum_K m(P_K) = \sum_K \sum_j m(P_K \cap Q_j) \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \parallel \\ m'\left(\bigcup_j Q_j\right) = \sum_j m(Q_j) = \sum_j \sum_K m(Q_j \cap P_K) \end{cases}$$

quindi m' non dipende dalla partizione che scelgo

Dati due vettori paralleli o antiparalleli $\underline{v}$ e $\underline{w}$ posso dire che $\underline{w} = K\,\underline{v}$

Prendo X e Y paralleli (o antiparallele), cioè $|\rho(X,Y)| = 1$ allora:

$$(X - E(X)\mathbb{1}) = K\,(Y - E(Y)\mathbb{1})$$

$$\mathrm{var}(X) = E\left((X - E(X)\mathbb{1})^2\right) = E\left((X - E(X)\mathbb{1})(X - E(X)\mathbb{1})\right) =$$

$$= E\left((X - E(X)\mathbb{1}) \cdot K \cdot (Y - E(Y)\mathbb{1})\right) = K\,\mathrm{cov}(X,Y)$$

$$\Rightarrow \quad K = \frac{\mathrm{var}(X)}{\mathrm{cov}(X,Y)}$$

Nell'ipotesi che siano paralleli posso quindi scrivere:

$$Y = E(Y)\mathbb{1} + \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(X)}\,(X - E(X)\mathbb{1})$$

quasi ovunque

APPROSSIMAZIONE LINEARE (REGRESSIONE LINEARE)

Date due variabili casuali X e Y, supponiamo che Y sia una funzione lineare di X:

$$Y = \beta X + \alpha$$

Imposto una campagna sperimentale volta a minimizzare la distanza tra Y e $\alpha + \beta X$.

Cioè devo minimizzare $\| Y - (\alpha + \beta X) \|$ che equivale a minimizzare:

$$\| Y - (\alpha + \beta X) \|^2$$

$$\varphi(\alpha, \beta) = E\left(\left(Y - (\alpha + \beta X) \right)^2 \right) = E\left(Y^2 + (\alpha + \beta X)^2 - 2Y(\alpha + \beta X) \right) =$$

$$= E(Y^2) + E\left(\alpha^2 + 2\alpha\beta X + \beta^2 X^2 \right) - 2E\left(Y(\alpha + \beta X) \right) =$$

$$= E(Y^2) + \alpha^2 + 2\alpha\beta E(X) + \beta^2 E(X^2) - 2\alpha E(Y) - 2\beta E(XY)$$

Cerco i punti di minimo con le derivate; poiché ho due variabili uso le derivate parziali

$$\begin{cases} \dfrac{\partial \varphi(\alpha,\beta)}{\partial \alpha} = 2\alpha + 2\beta E(X) - 2E(Y) \\[2ex] \dfrac{\partial \varphi(\alpha,\beta)}{\partial \beta} = 2\alpha E(X) + 2\beta E(X^2) - 2E(XY) \end{cases}$$

$$\begin{cases} \alpha + \beta E(X) - E(Y) = 0 \\ \alpha E(X) + \beta E(X^2) - E(XY) = 0 \end{cases}$$

$$\begin{cases} \alpha = E(Y) - \beta E(X) \\[2ex] E(X)E(Y) - \beta (E(X))^2 + \beta E(X^2) - E(XY) = 0 \end{cases} \Rightarrow$$

$$\Rightarrow \beta = \frac{E(X)E(Y) - E(XY)}{(E(X))^2 - E(X^2)} = \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(X)}$$

$$\begin{cases} \alpha = E(Y) - E(X)\dfrac{cov(X,Y)}{var(X)} = \alpha^* \\ \beta = \dfrac{cov(X,Y)}{var(X)} = \beta^* \end{cases}$$

Quindi la distanza tra Y e $\alpha + \beta X$ è minimizzata dalla scelta $\alpha = \alpha^*$ e $\beta = \beta^*$

Adesso prendiamo la funzione:

$$Z = \alpha^* + \beta^* X$$

e calcoliamo:

$$\| Y - Z \|^2$$

$$\varphi(\alpha^*, \beta^*) = E\left((Y-Z)^2\right) = E\left(\left(Y - \overbrace{E(Y) + E(X)\frac{cov(X,Y)}{var(X)}}^{\alpha^*} - X \overbrace{\frac{cov(X,Y)}{var(X)}}^{\beta^*}\right)^2\right) =$$

$$= E\left(\left((Y - E(Y)) - \frac{cov(X,Y)}{var(X)}(X - E(X))\right)^2\right) =$$

ma $(a - b - c)^2 = a^2 + b^2 + c^2 - 2ab - 2ac + 2bc = (a-b)^2 + c^2 - 2(a+b)c$

$$= E\left((Y - E(Y))^2 + \left(\frac{cov(X,Y)}{var(X)}\right)^2 (X - E(X))^2 - 2(X - E(X))(Y - E(Y))\frac{cov(X,Y)}{var(X)}\right) =$$

$$= var(Y) + var(X)\left(\frac{cov(X,Y)}{var(X)}\right)^2 - 2\frac{(cov(X,Y))^2}{var(X)} =$$

$$= Var(Y) - \frac{(cov(X,Y))^2}{Var(X)} = \left(\frac{var(Y)}{var(Y)} - \frac{(cov(X,Y))^2}{var(X)\,var(Y)}\right) var(Y) =$$

$$= \underbrace{var(Y)}_{\text{non è nulla}} \underbrace{\left(1 - (\rho(X,Y))^2\right)}_{\text{questo si annulla quando } (Y - E(Y)) - K(X - E(X))} = \varphi(\alpha^*, \beta^*)$$

Abbiamo così scoperto quale errore commetto

Se l'errore è piccolo, viene amplificato quando Y ha una grande varianza

Fissata una tolleranza $\varepsilon > 0$ qual'è la probabilità di non rispettare il limite che mi sono prefissato?

$$P(|Y - Z| \geq \varepsilon) \leq ?$$ qual'è l'upper bound?

Poiché il:

$$E(Y - Z) = 0$$

possiamo usare Tchebycheff:

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{var(X)}{\varepsilon^2}$$

Troviamo la varianza:

$$var(Y - Z) = E\left(\left(Y - Z - \overset{0}{E(Y - Z)}\right)^2\right) = E\left((Y - Z)^2\right) =$$

$$= var(Y) \cdot \left(1 - \left(\rho(X, Y)\right)^2\right)$$

Allora per Tchebycheff possiamo dire che:

$$P(|Y - Z| \geq \varepsilon) \leq \frac{var(Y)\left(1 - (\rho(X, Y))^2\right)}{\varepsilon^2}$$

TEORIA DELLA REGRESSIONE LINEARE

$$P(|Y - Z| \geq K\, var(Y - Z)) \leq \frac{1}{K^2}$$

APPROSSIMAZIONE GENERICA

Date due v.c. X e Y definite in (Ω, Σ, P) voglio trovare una funzione g che approssimi Y a X, senza che g sia lineare; cioè deve minimizzare:

$$\| Y - g(X) \|^2_{L_2(\Omega, \Sigma, P)}$$

Facciamo delle ipotesi di lavoro:

$$X, Y \quad (\Omega, \Sigma, P)$$

• X funzione semplice	• Y funzione semplice
• $X(\omega) \in \{x_1, x_2, \dots, x_K\}$ $x_i \neq x_j$	• $Y(\omega) \in \{y_1, y_2, \dots, y_h\}$ $y_i \neq y_j$
• $A_j = X^{-1}(\{x_j\})$	• $B_i = X^{-1}(\{y_i\})$
• $P(A_j) > 0$ $j = 1, \dots, K$	• $P(B_i) > 0$ $i = 1, \dots, h$

Per semplicità da qui in avanti indicheremo con g_j $g(x_j)$

Dobbiamo minimizzare

$$E\left(\left(Y - g(X)\right)^2\right) =$$

tramite $g : \mathbb{R} \to \mathbb{R}$

Possiamo scrivere:

$$X(\omega) = \sum_{j=1}^{K} x_j \, I_{A_j}(\omega)$$

$$g(X(\omega)) = \sum_{j=1}^{K} g_j \, I_{A_j}(\omega)$$

Lo stesso vale per Y, quindi:

$$= E\left(Y^2 + (g(X))^2 - 2Y g(X)\right) = E(Y^2) + E\left((g(X))^2\right) - 2E\left(Y g(X)\right) =$$

$$= E\left(\left(\sum_{i=1}^{h} y_i I_{B_i}(\omega)\right)^2\right) + E\left(\left(\sum_{j=1}^{K} g_j I_{A_j}(\omega)\right)^2\right) - 2E\left(\left(\sum_{i=1}^{h} y_i I_{B_i}(\omega)\right)\left(\sum_{j=1}^{K} g_j I_{A_j}(\omega)\right)\right) =$$

Osservazione: $I_A \cdot I_B = I_{A \cap B}$

$$= E\left(\sum_{i=1}^{h} y_i^2 I_{B_i}\right) + E\left(\sum_{j=1}^{K} g_j^2 I_{A_j}\right) - 2E\left(\sum_{i=1}^{h}\sum_{j=1}^{K} y_i g_j I_{A_j \cap B_i}\right) =$$

la ... allora:

$$= \sum_{i=1}^{h} y_i^2 \, E(I_{B_i}) + \sum_{j=1}^{k} g_j^2 \, E(I_{A_j}) - 2 \sum_{i=1}^{h} \sum_{j=1}^{k} y_i \, g_j \, E(I_{A_j \wedge B_i}) =$$

Questa grandezza è funzione di g quindi la chiamo

$$= \varphi(g_1, g_2, \dots, g_K)$$

Cerco i punti in cui le derivate parziali si annullano:

$$\frac{\partial \varphi}{\partial g_r} = 2 g_r E(I_{A_r}) - 2 \sum_{i=1}^{h} y_i \, E(I_{A_r \wedge B_i}) = 0 \qquad r = 1, \dots, k$$

è un punto di minimo?

$$\Rightarrow \quad g_r = \frac{\sum_{i=1}^{h} y_i \, E(I_{A_r \wedge B_i})}{E(I_{A_r})} =$$

Ma vale per le indicatrici che:

$$E(I_A) = 1 \cdot P(\{\omega \in \Omega : I_A(\omega) = 1\}) + 0 \cdot P(\{\omega \in \Omega : I_A(\omega) = 0\}) =$$
$$= P(\{\omega \in \Omega : I_A(\omega) = 1\}) = P(A)$$

allora:

$$= \frac{\sum_{i=1}^{h} y_i \, P(A_r \wedge B_i)}{P(A_r)} = \sum_{i=1}^{h} y_i \, P(B_i | A_r)$$

per definizione della probabilità condizionata

Quindi scelgo $g^*(X)$ che minimizzi $E((Y - g(X))^2)$

$$g^*(X_r) = \sum_{i=1}^{h} y_i \, P(B_i | A_r) = \sum_{i=1}^{h} y_i \, P(Y = y_i | X = x_r)$$

allora:

$$g^*(X) = \sum_{i=1}^{h} y_i \, P(Y = y_i | X)$$

E' una variabile casuale!

DEF: ATTESA CONDIZIONATA

$$E(Y|X) = g^*(X)$$

$$E(Y|X)(\omega) = \sum_{j=1}^{k} \left(\sum_{i=1}^{h} y_i \, P(B_i | A_j) \right) I_{A_j}(\omega)$$

Adesso chiediamoci quanto vale:

$$E\left((Y-g^*(X))^2\right)=E\left(\left(Y-E(Y|X)\right)^2\right)$$

ma lo sviluppo di questa equazione non porta a niente di intelligibile. Quindi quantifichiamola con Tchebycheff.

Cerchiamo allora quanto vale $E\left(Y-\left(E(Y|X)\right)\right)$ e $\operatorname{var}(Y-E(Y|X))$

Concentriamoci su $E(E(Y|X))$

$$E\left(E(Y|X)\right)=E\left(\sum_{j=1}^{k}\left(\sum_{i=1}^{h}y_i\,P(B_i|A_j)\right)I_{A_j}\right)=\sum_{j=1}^{k}\sum_{i=1}^{h}y_i\,P(B_i|A_j)\,E(I_{A_j})=$$

$$=\sum_{j=1}^{k}\sum_{i=1}^{h}y_i\,P(B_i|A_j)\,P(A_j)\underset{\downarrow}{=}\sum_{j=1}^{k}\sum_{i=1}^{h}y_iP(B_i\cap A_j)=\sum_{i=1}^{h}y_i\,P(B_i)=E(Y)\quad(1)$$

def prob condizionata $P(Z|W)=\dfrac{P(Z\cap W)}{P(W)}$

$$\operatorname{var}\left(Y-E(Y|X)\right)=E\left(\left(\left(Y-E(Y|X)\right)-\underbrace{\left(E(Y)-\underset{\overset{\shortparallel}{E(Y)}}{E(E(Y|X))}\right)}_{=0}\right)^2\right)=$$

$$=E\left(\left(Y-E(Y|X)\right)^2\right)=E(Y^2)+E\left(\left(E(Y|X)\right)^2\right)-2E\left(Y\,E(Y|X)\right)=$$

Th: Cmq scelta una v.c. $h(X)$ vale

$$E\left(Y\,h(X)\right)=E\left(E(Y|X)\,h(X)\right)$$

*Dim pagina seguente

$$=E(Y^2)+E\left(\left(E(Y|X)\right)^2\right)-2E\left(\left(E(Y|X)\right)^2\right)=E(Y^2)-E\left(\left(E(Y|X)\right)^2\right)=$$

$$=E(Y^2)-\left(E(Y)\right)^2+\left(E(Y)\right)^2-E\left(\left(E(Y|X)\right)^2\right)=$$

$$=\operatorname{var}(Y)+\left(E\left(E(Y\overset{(1)}{|}X)\right)\right)^2-E\left(\left(E(Y|X)\right)^2\right)=$$

$$=\operatorname{var}(Y)-\operatorname{var}\left(E(Y|X)\right)=\operatorname{var}\left(Y-E(Y|X)\right)$$

Come nel caso precedente (Y combinazione lineare di X) ci chiediamo che upper bound possiamo dare alla probabilità:

$$P(|Y - g^*(X)| \geq \varepsilon \sigma_Y)$$

$\uparrow$ fluttuazione di y

Possiamo applicare Tchebycheff perché $E(Y - g^*(X)) = 0$:

$$P(|Y - g^*(X)| \geq r \sigma_Y) \leq \frac{1}{r^2 \sigma_Y^2} \operatorname{var}(Y - g^*(X)) =$$

$$= \frac{1}{r^2 \operatorname{var}(Y)} \left(\operatorname{var}(Y) - \operatorname{var}(g^*(X)) \right) = \frac{1}{r^2} \left(1 - \frac{\operatorname{var}(g^*(X))}{\operatorname{var}(Y)} \right)$$

$\uparrow$ si vede pagina precedente

Quindi varrà:

$$P(|Y - g^*(X)| \geq \varepsilon \sigma_Y) \leq \frac{1}{\varepsilon^2} \left(1 - \frac{\operatorname{var}(E(Y|X))}{\operatorname{var}(Y)} \right)$$

Questo ci dice che quando le due v.c. sono uguali non ci sono errori.

* DIM: definisco $h(X(\omega)) = \sum_{j=1}^{K} h(x_j) I_{A_j}(\omega)$

$$E(Y|X)(\omega) \cdot h(X(\omega)) = \left(\sum_{j=1}^{K} \left(\sum_{i=1}^{h} y_i \, P(B_i | A_j) \right) I_{A_j}(\omega) \right) \cdot \sum_{r=1}^{K} h(x_r) I_{A_r}(\omega) =$$

$$= \left(\sum_{j=1}^{K} \left(\sum_{i=1}^{h} y_i \, P(B_i | A_j) \right) h(x_j) I_{A_j}(\omega) \right)$$

$\Rightarrow$

$$E\left(E(Y|X)\, h(X)\right) = \sum_{j=1}^{K} \left(\sum_{i=1}^{h} y_i \, P(B_i | A_j) \right) h(x_j) \underbrace{P(A_j)}_{E(I_{A_j})} =$$

$$= \sum_{j=1}^{K} \sum_{i=1}^{h} y_i \, h(x_j) \, P(X = x_j \cap Y = y_i) = E(Y \cdot h(X))$$

CVD

Confronto Integrale Riemann e Lebesgue

Th: Se l'integrale di Riemann esiste:

$$J = \int_a^b f(x)\, dx$$

allora $f(x)$ è Lebesgue integrabile su $[a,b]$ e vale che:

$$\int_{[a,b]} f(x)\, d\mu = J$$

dove μ è la misura di Lebesgue su $\mathbb{B}_{\mathbb{R}}$

Dim:

Dividiamo l'intervallo $[a,b]$ in 2^n parti usando i punti:

$$x_k = a + \frac{b-a}{2^n} k \qquad k = 0, \dots, 2^n$$

Definiamo:

$$M_{n,k} = \sup_{x \in [x_{k-1}, x_k]} \{f(x)\} \qquad m_{n,k} = \inf_{x \in [x_{k-1}, x_k]} \{f(x)\}$$

$$\overline{S}_n = \frac{b-a}{2^n} \sum_{k=1}^{2^n} M_{n,k}$$

$$\underline{S}_n = \frac{b-a}{2^n} \sum_{k=1}^{2^n} m_{n,k}$$

Per definizione dell'integrale di Riemann:

$$J = \lim_{n \to \infty} \overline{S}_n = \lim_{n \to \infty} \underline{S}_n = \int_a^b f(x)\, dx$$

Fissata

$$\overline{f}_n(x) = \sum_{k=1}^{2^n} M_{n,k} \cdot I_{[x_{k-1}, x_k)}(x) + f(b)\, I_{\{b\}}(x)$$

perché sono a 2 a 2 disgiunti

e:

$$\underline{f}_n(x) = \sum_{k=1}^{2^n} m_{n,k}\, I_{[x_{k-1}, x_k)}(x) + f(b)\, I_{\{b\}}(x)$$

E' chiaro che formano una partizione di $[a,b]$

Quindi è facile vedere che:

$$\int_{[a,b]} \overline{f}_n(x)\, d\mu = \overline{S}_n$$

$$\int_{[a,b]} \underline{f}_n(x)\, d\mu = \underline{S}_n$$

perché:

$$\int_{[a,b]} \underline{f}_n(x)\, d\mu = \sum_{k=1}^{2^n} m_{n,k} \cdot \mu([x_{k-1}, x_k]) = \sum_{k=1}^{2^n} m_{n,k} \frac{b-a}{2^n} =$$

$$= \frac{b-a}{2^n} \sum_{k=1}^{2^n} m_{n,k} = \underline{S}_n$$

Th: Dato un insieme A e:

$$f_1(x) \leq f_2(x) \leq \dots \leq f_n(x) \leq \dots$$

e vale che

$$\int_A f_n(x)\, d\mu \leq K$$

Allora il limite

$$f(x) = \lim_{n \to \infty} f_n(x)$$

esiste, la funzione $f(x)$ è integrabile su A, e:

$$\int_A f_n(x)\, d\mu \longrightarrow \int_A f(x)\, d\mu$$

Per le nostre $\underline{f}_n$ e $\overline{f}_n$ vale che:

$\exists K : \forall n \int_{[a,b]} f_n(x) d\mu \leq K$ perché sono limitate su $[a,b]$

e vale anche che:

$$\overline{f}_1 \leq \overline{f}_2 \leq \overline{f}_3 \quad \dots$$

$$\underline{f}_1 \leq \underline{f}_2 \leq \underline{f}_3 \leq \dots$$

Chiamiamo misura esterna la funzione:

$$\mu^* : A \to \mathbb{R} \qquad ~~A \subseteq \mathcal{E}~~ \quad A \text{ può essere qualsiasi con}$$

$$\mu^*(A) = \inf \{ \sum_k m(P_k) , A \subseteq \bigcup_k P_k \}$$

ovvero la più piccola copertura di A di finiti o numerabili rettangoli.

Chiamiamo invece misura interna

$$\mu_*(A) = m(E) - \mu^*(E \setminus A)$$

LEZ II 4/3/10

Th: Dato $A \in \mathcal{E}$ con $\{A_n\}^{\mathcal{E}}$ collezione numerabile e $A \subseteq \bigcup_n A_n$ allora:

$$m'(A) \leq \sum_n m'(A_n)$$

DIM

?

$\forall \varepsilon > 0 \quad \exists \bar{A} \in \mathcal{E}$ con $\bar{A}$ chiuso e $\bar{A} \subseteq A$

t.c. $m'(\bar{A}) \geq m'(A) - \frac{\varepsilon}{2}$

$$A = \bigcup_K P_K \qquad \text{con } P_K \in \mathcal{R}$$

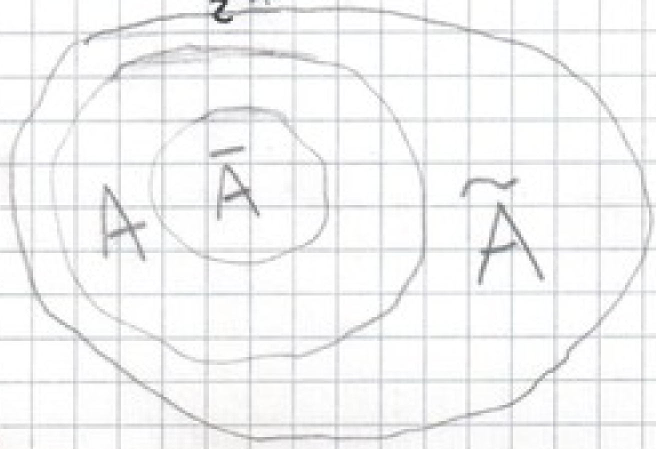
$$m(\bar{P}_i) \geq m(P_i) - \frac{\varepsilon}{2^{i+1}}$$

Dato $A_n \quad \exists \tilde{A}_n : A_n \subseteq \tilde{A}_n$ con $\tilde{A}_n$ aperto e $\tilde{A}_n \in \mathcal{E}$

tale che:

$$m'(\tilde{A}_n) \leq m'(A_n) + \frac{\varepsilon}{2^{n+1}}$$

vale che: $\bar{A} \subseteq \bigcup_n \tilde{A}_n$



Quindi possiamo applicare il teorema precedente

$$\bar{f}_n \xrightarrow[n \to \infty]{} \bar{f}$$

$$\underline{f}_n \xrightarrow[n \to \infty]{} \underline{f}$$

Allora

si vede teorema succ

$$\int_{[a,b]} \bar{f}(x)\, d\mu = \int_{[a,b]} \lim_{n \to \infty} \bar{f}_n(x)\, d\mu = \lim_{n \to \infty} \int_{[a,b]} \bar{f}_n(x)\, d\mu = \lim_{n \to +\infty} \bar{S}_n =$$

$$= J = \lim_{n \to \infty} \underline{S}_n = \lim_{n \to \infty} \int_{[a,b]} \underline{f}_n\, d\mu = \int_{[a,b]} \lim_{n \to +\infty} \underline{f}_n\, d\mu = \int_{[a,b]} \underline{f}\, d\mu$$

Th :

Se $f_n(x) \to f(x)$ allora:

$$\lim_{n \to +\infty} \int_{[a,b]} f_n\, d\mu = \int_{[a,b]} \lim_{n \to \infty} f_n\, d\mu = \int_{[a,b]} f\, d\mu$$

Misura di Lebesgue - Stieltjes

Prendiamo una funzione $F: \mathbb{R} \to \mathbb{R}$ che abbia le seguenti proprietà:

1) Monotona non decrescente
2) Continua da destra: $\lim_{\varepsilon \to 0} F(x+\varepsilon) = F(x)$

Definiamo la misura m come segue:

$$m : S_m \to \mathbb{R}$$

dove $S_m = \{ a, b \in \mathbb{R}, \ a \leq b, \ [a,b], (a,b], (a,b], [a,b), (a,b) \}$ è un semianello

$$m([a,b]) = F(b) - \lim_{\varepsilon \to 0} F(a-\varepsilon)$$

$$m((a,b]) = F(b) - F(a)$$

$$m([a,b)) = \lim_{\varepsilon \to 0} F(b-\varepsilon) - \lim_{\varepsilon \to 0} F(a-\varepsilon)$$

$$m((a,b)) = \lim_{\varepsilon \to 0} F(b-\varepsilon) - F(a)$$

Perché m sia una misura devono valere le seguenti proprietà:

- Non negativa? Sì, perché F è monotona
- Additiva? Vediamo:

Partizioniamo l'intervallo $[a,b]$ in questo modo:

$$\underbrace{[x_0, x_0]}_{A_0} \cup \underbrace{(x_0, x_1]}_{A_1} \cup \underbrace{(x_1, x_2]}_{A_2} \cup \dots \cup \underbrace{(x_{k-1}, x_k]}_{A_k}$$

dove $x_0 = a$ e $x_k = b$

Dobbiamo dimostrare che:

$$m([a,b]) = \sum_k m(A_k)$$

$$m([a,b]) = F(b) - \lim_{\varepsilon \to 0} F(a-\varepsilon) =$$

$$= m([x_0]) + m((x_0, x_1]) + \dots + m((x_{k-1}, x_k]) =$$

$$= F(x_0) - \lim_{\varepsilon \to 0} F(x_0 - \varepsilon) + F(x_1) - F(x_0) + \dots + F(x_K) - F(x_{K-1}) =$$

$$= F(b) - \lim_{\varepsilon \to 0} F(a - \varepsilon)$$

E' additiva

m è una misura!

Estendiamo $m : \mathcal{S}_m \to \mathbb{R}$ a $\mu_F : \mathcal{B}_{\mathbb{R}} \to \mathbb{R}$ e otteniamo la misura a Lebesgue-Stieltjes

Def: μ_F è assolutamente continua rispetto a μ sse:

$$\mu(A) = 0 \Rightarrow \mu_F(A) = 0 \quad A \in \mathbb{B}_{\mathbb{R}}$$

Si indica con:

$$\mu_F \ll \mu$$

Se aggiungiamo 2 proprietà ad F:

3) $\lim_{x \to -\infty} F(x) = 0$

4) $\lim_{x \to +\infty} F(x) = 1$

μ_F è una misura di probabilità

Th: Radon-Nikodym

μ misura di Lebesgue su $\mathbb{B}_{\mathbb{R}}$

μ_F misura di Lebesgue-Stieltjes su $\mathbb{B}_{\mathbb{R}}$

$\mu_F \ll \mu$

Allora $\exists f : \mathbb{R} \to \mathbb{R}$ μ-misurabile tale che:

$$\forall E \in \mathbb{B}_{\mathbb{R}} \quad \mu_F(E) = \int_E f \, d\mu$$

f si chiama derivata di Radon-Nikodym:

$$f = \frac{dF}{d\mu}$$ f è determinata quasi ovunque

La F che soddisfa le 4 proprietà è la funzione di ripartizione.

f è la densità di probabilità

La derivata di Radon-Nikodym esiste anche per le funzioni discrete (noi non lo vedremo).

TEOREMA LIMITE CENTRALE

Data una collezione $X_1, X_2, \dots, X_n$ di variabili casuali indipendenti e identicamente distribuite ($F_{X_i} = F_X \quad i = 1, \dots, n$) aventi momento 1° e 2° finiti, definite le variabili casuali:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad \bar{X}^*_n = \frac{\bar{X}_n - E(\bar{X}_n)}{\sigma_{\bar{X}_n}}$$

allora vale che:

$$\underbrace{\varphi(x)}_{F_{N(0,1)}} = \lim_{n \to +\infty} \underbrace{P(\bar{X}^*_n \leq x)}_{F_{\bar{X}^*_n}} \qquad \text{PUNTUALMENTE}$$

Questo teorema è utile per determinare la convergenza delle funzioni di ripartizione

DIMOSTRAZIONE

Cerchiamo di dimostrarlo con le funzioni di ripartizione:

$$F_{\bar{X}^*_n}(x) \simeq F_{\bar{X}_n} \simeq F_{\sum X_i} \quad \text{a meno di costanti}$$

$$F_{X_1+X_2}(z) \qquad P(X_1 + X_2 \leq z) = \int_{-\infty}^{+\infty} f_{X_1}(w)\, f_{X_2}(z-w)\, dw$$

Questo è un prodotto di convoluzioni. Ciò ci fa capire che non è facile intraprendere questa strada.

È più conveniente utilizzare le generatrici di momenti (MGF) poiché:

$$m_{X_1+X_2}(t) = E\left(e^{t(X_1+X_2)}\right) = E\left(e^{tX_1} e^{tX_2}\right) = E\left(e^{tX_1}\right) \cdot E\left(e^{tX_2}\right) =$$
$$= m_{X_1}(t)\, m_{X_2}(t)$$

Tuttavia, a volte, le MGF può non esistere perché:

$$m_X(t) = E\left(e^{tX}\right)$$

e l'integrale può non convergere

Esempio 1:

$X \sim$ Cauchy-standard

$$f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$

misura di probabilità

$$m_X(t) = E(e^{tX}) = \int_{\mathbb{R}} e^{tx}\, d\mu_F = \int_{-\infty}^{+\infty} e^{tx} f_X(x)\, dx =$$

perché: $\frac{d\mu_F}{d\mu} \cdot d\mu \Rightarrow f_X(x)\, d\mu \rightarrow dx$ passaggio integrale Riemann

$$= \int_{-\infty}^{+\infty} e^{tx}\frac{1}{\pi}\frac{1}{1+x^2}\, dx = \frac{1}{\pi}\left(\int_{-\infty}^{0} e^{tx}\frac{1}{1+x^2}\, dx + \int_{0}^{+\infty} e^{tx}\frac{1}{1+x^2}\, dx\right)$$

Se $t \geq 0$ $\int_0^{+\infty} e^{tx}\frac{1}{1+x^2}\, dx$ diverge, $\int_{-\infty}^{0} e^{tx}\frac{1}{1+x^2}\, dx$ converge

Se $t \leq 0$ // converge, // diverge

Se $t = 0$ convergono tutti e due, quindi $m_X(t)$ è definita solo per $t = 0$

Esempio 2:

$Y \sim$ Esponenziale

$$F_Y(y) = (1 - e^{-\lambda y})\, I_{(0,+\infty)}(y)$$

$$f_Y(y) = \frac{dF_Y}{dy} = \lambda e^{-\lambda y}\, I_{(0,+\infty)}(y)$$

$$m_Y(t) = E(e^{tY}) = \lambda\int_0^{+\infty} e^{ty} e^{-\lambda y}\, dy =$$

perché $\int_{-\infty}^{0}$ è nullo per via dell' $I_{(0,+\infty)}(y)$

$$= \lambda\int_0^{+\infty} e^{(t-\lambda)y}\, dy = \frac{\lambda}{t-\lambda}\left[e^{(t-\lambda)y}\right]_0^{+\infty}$$

Quindi perché converga deve valere che $\lambda \geq t \Rightarrow m_Y(t) = \frac{\lambda}{\lambda - t}$

Risolveremo tutti questi problemi usando la funzione caratteristica:

DEF: FUNZIONE CARATTERISTICA

$\varphi_X(t) = E\left(e^{itx}\right)$

dove 'i' è la componente immaginaria $e^{itx} : \Omega \to \mathbb{C}$

Se X è continua $\varphi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f_x(x)\, dx$

Quest'ultima è molto simile alla trasformata di Fourier con opportune precisazioni

Ritorniamo alla dimostrazione

Poiché $m_X(t)$ ha problemi di esistenza usiamo $\varphi_X(t) = E\left(e^{itx}\right)$

Dobbiamo chiederci: c'è corrispondenza tra φ_X e F_X?

Un teorema ci dice di sì:

TH DI INVERSIONE: Dati $\alpha < \beta$ punti di continuità di F_X allora:

$$F(\beta) - F(\alpha) = \lim_{\sigma \to 0} \frac{1}{2\pi} \int_{-\infty}^{+\infty} \varphi_X(\lambda)\, e^{\frac{-\sigma^2 \lambda^2}{2}} \frac{e^{-i\lambda\beta} - e^{-i\lambda\alpha}}{-i\lambda}\, d\lambda$$

$\varphi_X(t)$ esiste se $E(X)$ esiste

Continuare la dimostrazione ...

Funzione Esponenziale Complessa

Def: $z \in \mathbb{C}$

$$\exp(z) = e^z = \sum_{n=0}^{+\infty} \frac{z^n}{n!}$$

Questa serie converge assolutamente (cioè in valore assoluto)

Proprieta'

- $\exp(z)$ è continua in z
- $$\exp(a)\exp(b) = \sum_{k=0}^{+\infty} \frac{a^k}{k!} \cdot \sum_{h=0}^{+\infty} \frac{b^h}{h!} = \sum_{n=0}^{+\infty} \frac{1}{n!} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} a^k b^{n-k} =$$
  $$= \sum_{n=0}^{+\infty} \frac{1}{n!} (a+b)^n = e^{a+b}$$
- $e = \exp(1) = \sum_{n=0}^{+\infty} \frac{1}{n!} = e$
- $\exp(0) = 1$
- $\exp'(z) = \frac{d}{dz}\exp(z) = \exp(z)$

  Dimostrazione:

  $$\exp'(z) = \lim_{h \to 0} \frac{\exp(z+h) - \exp(z)}{h} = \exp(z) \lim_{h \to 0} \frac{\exp(h) - 1}{h} =$$
  $$= \exp(z) \lim_{h \to 0} \frac{1}{h}\left(\sum_{n=0}^{+\infty} \frac{h^n}{n!} - 1\right) = \exp(z) \lim_{h \to 0} \frac{1}{h}\left(\frac{h^0}{0} + \frac{h^1}{1!} + \frac{h^2}{2!} + \ldots - 1\right) =$$
  $$= \exp(z)$$
- $\lim_{z \to -\infty} e^z = 0$
- $\lim_{z \to +\infty} e^z = \infty$
- Complesso coniugato

  $$\overline{(e^{it})} = \overline{\sum_{n=0}^{+\infty} \frac{it^n}{n!}} = \sum \frac{(\overline{it})^n}{n!} = \sum \frac{(-it)^n}{n!} = e^{-it}$$

- $e^{it} e^{\overline{it}} = \exp(it - it) = \exp(0) = 1$

- $\exp(it) = \sum_{n=0}^{+\infty} \frac{it^n}{n!} = \sum_{n=0}^{+\infty} \frac{(it)^{2n}}{(2n)!} + \sum_{n=0}^{+\infty} \frac{(it)^{2n+1}}{(2n+1)!} =$

$$= \sum \frac{i^{2n} t^{2n}}{2n!} + i \sum \frac{i^{2n} t^{2n+1}}{(2n+1)!} = \cos(t) + i \operatorname{sen}(t)$$

- $\exists$ un numero positivo g tale che $e^{\frac{ig}{2}} = i$ e tale che $e^z = 1$ sse $\frac{z}{2ig}$ è intero

  Tale numero z è π

- $\forall t \in \mathbb{R} \quad |e^{it}| = 1$

- $w \in \mathbb{C} \quad \exists z \in \mathbb{C} : w = e^z$

$$\forall t \quad \cos^2(t) + \operatorname{sen}^2(t) = 1 \quad \Rightarrow |e^{it}| = 1$$

Esempio Funzione Caratteristica Exp

$X \sim Exp(\lambda)$

$F_X(x) = (1 - e^{-\lambda x}) I_{(0,+\infty)}(x)$

F_X è continua allora $\frac{dF_X}{dx} = f_X(x) = \lambda e^{-\lambda x} I_{(0,+\infty)}(x)$

$$\varphi_X(t) = E(e^{itX}) = \int_\Omega e^{itX(\omega)} d\mu_{F_X} = \int_{-\infty}^{+\infty} e^{itx} f(x)\, dx =$$

$$= \int_0^{+\infty} e^{itx} \lambda e^{-\lambda x} dx = \lambda \int_0^{+\infty} e^{-x(\lambda - it)} dx =$$

poniamo $y = x(\lambda - it) \Rightarrow \frac{dy}{dx} = \lambda - it \quad dx = \frac{dy}{\lambda - it}$

$$= \lambda \int_0^{+\infty} \frac{e^{-y}}{\lambda - it} dy = \frac{\lambda}{\lambda - it} \int_0^{+\infty} e^{-y} dy = \frac{\lambda}{\lambda - it}$$

Esempio Funzione Caratteristica Poisson

$Y \sim Pois(\lambda)$

$f_Y(y) = \frac{\lambda^y}{y!} e^{-\lambda} I(y)$

$$\varphi_Y(t) = E(e^{itY}) = \sum_{y=0}^{+\infty} e^{ity} \frac{\lambda^y}{y!} e^{-\lambda} = e^{-\lambda} \sum_{y=0}^{+\infty} \frac{(e^{it}\lambda)^y}{y!} = e^{-\lambda} e^{\lambda e^{it}} =$$

$$= e^{\lambda(e^{it} - 1)}$$

Esiste per tutti i t a differenza della generatrice dei momenti

Th: Poiché $\bar{A}$ è chiuso e $\bigcup_{n} \tilde{A}_n$ è una copertura aperta numerabile di $\bar{A}$ allora da $\{\tilde{A}_n\}_{n=1}^{+\infty}$ posso estrarre una collezione finita di elementi di $\{\tilde{A}_n\}$ tale che $\{\tilde{A}_{s_1}, \tilde{A}_{s_2}, \dots, \tilde{A}_{s_K}\}: \bigcup_{i=1}^{K} \tilde{A}_{s_i} \supseteq \bar{A}$

Poiché $\bar{A} \subseteq \bigcup_{i=1}^{K} \tilde{A}_{s_i} \Rightarrow m'(\bar{A}) \leq \sum_{i=1}^{K} m'(\tilde{A}_{s_i})$

$$m'(A) \leq m'(\bar{A}) + \frac{\varepsilon}{2} \leq \sum_{i=1}^{K} m'(\tilde{A}_{s_i}) + \frac{\varepsilon}{2} \leq \sum_{n} m'(\tilde{A}_n) + \frac{\varepsilon}{2} \leq$$
$$\leq \sum_{n} \left(m'(A_n) + \frac{\varepsilon}{2^{n+1}} \right) + \frac{\varepsilon}{2} \leq \sum_{n} m'(A_n) + \frac{\varepsilon}{2} \cdot \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n + \frac{\varepsilon}{2} =$$
$$= \sum_{n} m'(A_n) + \varepsilon$$

CVD

Th: $\mu^*(A) \geq \mu_*(A)$

Dim:

Supponiamo per assurdo che $\mu_*(A) > \mu^*(A)$

Per convenzione $m'(E) = 1$

Quindi varrà:

$$\mu^*(A) < 1 - \mu^*(E \setminus A)$$

$$\mu^*(A) + \mu^*(E \setminus A) < 1 = m'(E)$$

Prendiamo $\{P_i\}$ copertura di A e $\{Q_j\}$ copertura di $E \setminus A$, varrà che:

$\sum_i m(P_i) + \sum_j m(Q_j) < 1$ questo è un assurdo

perché:

- $A \cup (E \setminus A) = E$
- $E \subseteq \bigcup_i P_i \cup \bigcup_j Q_j$

assurdo

CVD

ESEMPIO FUNZIONE CARATTERISTICA NORMALE STANDARD

$N \sim Norm(0,1)$

$$\phi_N(t) = E(e^{itN}) = \int_{-\infty}^{+\infty} e^{itx} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx = \frac{1}{\sqrt{2\pi}} \int e^{itx - \frac{x^2}{2}} dx =$$

ma $\frac{1}{2}(x^2 - 2itx - t^2) = \frac{1}{2}(x - it)^2$

$$= \frac{1}{\sqrt{2\pi}} \int e^{-\frac{x^2}{2} + itx + \frac{t^2}{2} - \frac{t^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int e^{-\frac{1}{2}(x-it)^2} e^{-\frac{t^2}{2}} dx =$$

poniamo $y = x - it$

$$= e^{-\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} \int e^{-\frac{1}{2}y^2} dy = e^{-\frac{t^2}{2}}$$

ma questa quantità è la densità della normale standard su tutto $\mathbb{R}$, allora vale 1

RELAZIONE TRA I MOMENTI DELLA MGF E DELLA FUNZIONE CARATTERISTICA

Per la MGF vale che:

$$\frac{d^r m_X}{dt}\Big|_{t=0} = E(X^r)$$

vediamo se ciò si può applicare alla f. caratteristica:

MOMENTO PRIMO: $\dfrac{d\,\varphi_X(t)}{dt}\Big|_{t=0} \overset{?}{=} E(X)$

proviamo con $Y \sim Poiss$ $\quad E(Y) = \lambda$

$$\varphi_Y(t) = e^{\lambda(e^{it}-1)}$$

$$\varphi'_Y(t) = e^{\lambda(e^{it}-1)}\, i\lambda\, e^{it}\Big|_0 = i\lambda$$

~~sem~~ vale che $\dfrac{\varphi'(t)|_{t=0}}{i} = E(Y)$

proviamo con le derivate seconde

$$\varphi''_Y(t) = e^{\lambda(e^{it}-1)}\, e^{it}\, i\lambda\, e^{it}\, i\lambda + e^{\lambda(e^{it}-1)} \cdot i^2\lambda\, e^{it}\Big|_{t=0} =$$

$$= -\lambda^2 - \lambda$$

$m''_Y(t)\Big|_{t=0} = \lambda^2 + \lambda$ $\quad$ allora vale che:

$$\frac{\varphi''_Y(t)\,|_{t=0}}{i^2} = E(Y^2)$$

In generale varrà che

$$\frac{1}{i^n}\,\frac{d^n \varphi_X(t)}{dt^n}\Big|_{t=0} = E(X^n)$$

INVARIANZA PER ROTAZIONE DELLA NORMALE STANDARD

$$X_1 \sim N(0,1) \qquad X_2 \sim X_1 \qquad X_2 \perp X_1 \qquad \varphi_{X_1}(t) = e^{-\frac{t^2}{2}}$$

Dato un $\alpha \in \mathbb{R}$ con $0 < \alpha < 2\pi$ definiamo

$$X_1(\alpha) = X_1 \cos\alpha + X_2 \,\mathrm{sen}\,\alpha$$

$$X_2(\alpha) = -X_1 \,\mathrm{sen}\,\alpha + X_2 \cos\alpha$$

che possiamo scrivere:

$$\underline{X}(\alpha) = \begin{pmatrix} X_1(\alpha) \\ X_2(\alpha) \end{pmatrix} = \begin{pmatrix} \cos\alpha & \mathrm{sen}\,\alpha \\ -\mathrm{sen}\,\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

è come se facessimo una rotazione sul piano di α gradi

Calcoliamo:

$$\varphi_{\underline{X}(\alpha)}(\underline{u}) = \varphi_{X_1(\alpha)X_2(\alpha)}(u_1, u_2) = E\left(e^{i(u_1 X_1(\alpha) + u_2 X_2(\alpha))}\right) =$$

$$= E\left(e^{i(u_1 X_1 \cos\alpha + u_1 X_2 \,\mathrm{sen}\,\alpha - u_2 X_1 \,\mathrm{sen}\,\alpha + u_2 X_2 \cos\alpha)}\right) =$$

$$= E\left(e^{i\left(X_1(u_1\cos\alpha - u_2 \,\mathrm{sen}\,\alpha) + X_2(u_1 \,\mathrm{sen}\,\alpha + u_2 \cos\alpha)\right)}\right) =$$

$$= \varphi_{X_1}(u_1 \cos\alpha - u_2 \,\mathrm{sen}\,\alpha)\, \varphi_{X_2}(u_1 \,\mathrm{sen}\,\alpha + u_2 \cos\alpha) =$$

$$= e^{-\frac{1}{2}(u_1\cos\alpha - u_2 \,\mathrm{sen}\,\alpha)^2} e^{-\frac{1}{2}(u_1 \,\mathrm{sen}\,\alpha + u_2\cos\alpha)^2} =$$

$$= e^{-\frac{1}{2}(u_1^2 + u_2^2)}$$

Questo risultato ci dice che se prendo una coppia di v.c. normali e le ruoto ottengo ancora una coppia di v.c. normali identiche.

Vediamo se per $\alpha = \frac{\pi}{4}$ vale quanto detto sopra:

$$\underline{X}\left(\frac{\pi}{4}\right) = \begin{pmatrix} X_1 \frac{1}{\sqrt{2}} + X_2 \frac{1}{\sqrt{2}} \\ -X_1 \frac{1}{\sqrt{2}} + X_2 \frac{1}{\sqrt{2}} \end{pmatrix} \quad \text{quindi} \quad X_1\left(\frac{\pi}{4}\right) = \frac{X_1 + X_2}{\sqrt{2}} \qquad X_2\left(\frac{\pi}{4}\right) = \frac{X_2 - X_1}{\sqrt{2}}$$

$$\varphi_{X_1(\frac{\pi}{4})}(t) = E\left(e^{it\frac{X_1+X_2}{\sqrt{2}}}\right) = E\left(e^{i\frac{t}{\sqrt{2}}X_1}\right) E\left(e^{i\frac{t}{\sqrt{2}}X_2}\right) = \varphi_{X_1}\left(\frac{t}{\sqrt{2}}\right)\varphi_{X_2}\left(\frac{t}{\sqrt{2}}\right) = e^{-\frac{1}{2}t^2}$$

Th: Date due v.c. $A \sim B$, $A \perp B$, di cui non conosco la distribuzione, per cui vale $\frac{A+B}{\sqrt{2}} \sim A$ allora A e B sono v.c. normali standard.

DIM:

Introduciamo una funzione $\psi(t) = \ln \varphi(t)$

$$\varphi_A(u) = \varphi_B(u) = \varphi_{\frac{A+B}{\sqrt{2}}}(u) = \quad \text{per ipotesi del teorema}$$

$$= E\left(e^{iu\frac{A+B}{\sqrt{2}}}\right) = E\left(e^{\frac{iuA}{\sqrt{2}}} e^{\frac{iuB}{\sqrt{2}}}\right) = \varphi_A\left(\frac{u}{\sqrt{2}}\right)\varphi_B\left(\frac{u}{\sqrt{2}}\right) = \left(\varphi\left(\frac{u}{\sqrt{2}}\right)\right)^2$$

Vediamo per $\psi(u)$:

$$\psi(u) = \ln(\varphi(u)) = \ln\left(\varphi\left(\frac{u}{\sqrt{2}}\right)\right)^2 = 2\ln\varphi\left(\frac{u}{\sqrt{2}}\right) = 2\psi\left(\frac{u}{\sqrt{2}}\right)$$

Dimostriamo che vale $\frac{\psi(u)}{u^2} = \frac{\psi\left(\frac{u}{2^{\frac{n}{2}}}\right)}{\left(\frac{u}{2^{\frac{n}{2}}}\right)^2}$

$$\frac{\psi(u)}{u^2} = \frac{\ln(\varphi(u))}{u^2} = \frac{1}{u^2}\ln\left(\varphi\left(\frac{u}{\sqrt{2}}\right)\varphi\left(\frac{u}{\sqrt{2}}\right)\right) =$$

$$\text{ma } \varphi\left(\frac{u}{\sqrt{2}}\right) = \left(\varphi\left(\frac{u}{\sqrt{2}}\cdot\frac{1}{\sqrt{2}}\right)\right)^2 \quad (1)$$

$$= \frac{1}{u^2}\ln\left(\varphi\left(\frac{u}{(\sqrt{2})^2}\right)\right)^4 = \frac{4}{u^2}\ln\varphi\left(\frac{u}{(\sqrt{2})^2}\right) = \frac{4}{u^2}\psi\left(\frac{u}{(\sqrt{2})^2}\right) = \frac{\psi\left(\frac{u}{2^{\frac{2}{2}}}\right)}{\left(\frac{u}{2^{\frac{2}{2}}}\right)^2} =$$

In generale si può iterare sulla (1) e ottenere così

$$= \frac{\psi\left(\frac{u}{2^{\frac{n}{2}}}\right)}{\left(\frac{u}{2^{\frac{n}{2}}}\right)^2}$$

Adesso calcoliamo il limite $\lim_{u\to 0}\frac{\psi(u)}{u^2}$

Proviamo con la serie di Taylor centrata nello 0

$$\frac{1}{u^2}\psi(u-0) = \frac{1}{u^2}\psi(0) + \frac{1}{u^2}\psi'(0)\cdot u + \frac{1}{u^2}\frac{\psi''(0)}{2}u^2 + \frac{1}{u^2}\frac{\psi'''(0)}{3!}u^3 + \frac{TOS}{u^2}$$

è chiaro che calcolare il limite si riduce a considerare queste quantità, poiché le altre si annullano per $u \to 0$

$$\lim_{u\to 0}\frac{\psi(u)}{u^2} = \cancel{\frac{1}{u^2}}\psi(0) + \cancel{\frac{1}{u}}\psi'(0) + \frac{1}{2}\psi''(0)$$

Calcoliamo le derivate di ψ

$$\psi'(0) = \frac{d}{dt} \ln(\varphi(t))\Big|_{t=0} = \frac{\varphi'(t)}{\varphi(t)}\Big|_{t=0}$$

$$\psi''(0) = \frac{\varphi''(t)\varphi(t) - (\varphi'(t))^2}{(\varphi(t))^2}\Big|_{t=0}$$

Abbiamo bisogno delle derivate di φ

$$\varphi_A(0) = E(e^{i0x}) = E(1) = 1$$

supponiamo che $E(A) = 0$ $\quad var(A) = 1$

$$\varphi_A'(0) = i\,E(A) = 0$$

$$\varphi_A''(0) = ?$$

sappiamo che $var(A) = 1 = E(A^2) - (\overset{0}{\overset{\shortparallel}{E(A)}})^2 = E(A^2) = \frac{\varphi_A''(t)}{i^2}$ allora

$$\varphi_A''(0) = i^2 = -1$$

Adesso possiamo calcolare:

$$\psi'(0) = \frac{0}{1} = 0$$

$$\psi''(0) = \frac{-1\cdot 1 - 0^2}{1^2} = -1$$

Allora il limite vale:

$$\lim_{u\to 0} \frac{\psi(0)}{u^2} = -\frac{1}{2}$$

Calcoliamo il limite:

$$\lim_{n\to+\infty} \frac{\psi(u/2^{n/2})}{(u/2^{-n/2})^2} = \frac{\psi(u)}{u^2} \overset{u\to 0}{=} -\frac{1}{2} \qquad ?$$

allora

$$\psi(u) = -\frac{1}{2}u^2 = \ln(\varphi(u))$$

$$\varphi(u) = e^{-\frac{1}{2}u^2} = \varphi_A(u) = \varphi_B(u) = \varphi_{\frac{A+B}{\sqrt{2}}}(u)$$

CVD

Il seguente teorema generalizza ancora di più:

Th: X_1 X_2 v.c.

$$Y_1 = a_{11} X_1 + a_{12} X_2$$

$$Y_2 = a_{21} X_1 + a_{22} X_2$$

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \underbrace{\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}}_{\Delta = 1} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

perché deve essere una matrice di rotazione

Se $X_1 \perp X_2$ e $Y_1 \perp Y_2$ e $a_{ij} \neq 0 \quad \forall\, i,j = 1,2$

Allora $X_1, X_2, Y_1, Y_2 \sim N$

Corollario: se il th vale in un punto allora vale su tutto Ω

Poiché noi abbiamo dimostrato per un punto $\alpha = \frac{\pi}{4}$ vorrei per tutto Ω

Th BOCHNER

$\varphi : \mathbb{R} \to \mathbb{C}$ è una funzione caratteristica di una v.c. sse:

1) è semi definita positiva: se $x > 0 \Rightarrow f(x) \geq 0$

2) è continua nell'origine: $\lim_{x \to 0^+} f(x) = \lim_{x \to 0^-} f(x)$

3) $\varphi(0) = 1$

Abbiamo visto che è possibile passare

$$X \xrightarrow{F_X} \varphi_X$$

ma anche che:

$$\varphi_X \longrightarrow F_X$$ grazie al teorema d'inversione

Esiste un caso particolare in cui il th d'inversione assume una forma più semplice:

se X è un v.c. continua posso scrivere:

$$\varphi_X(t) = E\left(e^{itX}\right) = \int_{-\infty}^{+\infty} e^{itx} f_X(x)\, dx$$

ma anche

$$f_X(x) = \int_{-\infty}^{+\infty} e^{-itx} \varphi_X(t)\, dt$$ solo perché X è continua

Assomigliano molto rispettivamente alla trasformata e alla antitrasformata di Fourier.

TRASFORMATA DI FOURIER:

$$(\mathcal{F} f)(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x)\, e^{itx}\, dx$$

ANTITRASFORMATA DI FOURIER

$$(\mathcal{F}^{-1} \varphi)(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \varphi(t)\, e^{-itx}\, dt$$

la costante $\frac{1}{\sqrt{2\pi}}$ normalizza le cose, per la gaussiana

Esempio 1

$X \sim N(0,1) \qquad \varphi_X(t) = e^{-\frac{1}{2}t^2} \qquad f_X(x) = ?$

$$f_X(x) = \int_{-\infty}^{+\infty} e^{-itx} \varphi_X(t)\, dt = \int_{-\infty}^{+\infty} e^{-itx} e^{-\frac{1}{2}t^2} dt =$$

$$= \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(t^2 - 2itx)} dt =$$

ma $(t - ix)^2 = t^2 - x^2 - 2itx$ quindi:

$$(t - ix)^2 + x^2 = t^2 - 2itx$$

$$= \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(x^2 + (t - ix)^2\right)} dt = \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} e^{-\frac{1}{2}(t-ix)^2} dt =$$

$$= e^{-\frac{x^2}{2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(t-ix)^2} dt \overset{?}{=} e^{-\frac{x^2}{2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}t^2} dt = \sqrt{2\pi}\, e^{-\frac{x^2}{2}}$$

Esempio 2

$X \sim$ Cauchy-standard

Per questa distribuzione non esiste il valore medio

$$f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2} \qquad \int_{-\infty}^{+\infty} \frac{1}{\pi} \frac{x}{1+x^2} dx \quad \text{non è definito}$$

Vediamo se esiste la $\varphi_X(t)$

$$\varphi_X(t) = E(e^{itx}) = \int_{-\infty}^{+\infty} e^{itx} f_X(x)\, dx \leq \int_{-\infty}^{+\infty} |e^{itx}|\, f_X(x) dx =$$

$$= \int_{-\infty}^{+\infty} f_X(x)\, dx = 1$$

Questo ci dice che la funzione caratteristica esiste (anche se non esiste il valore medio) a differenza della mgf

Proviamo a calcolarla:

$$\varphi_X(t) = \int_{-\infty}^{+\infty} \frac{1}{\pi} \frac{1}{1+x^2} e^{itx} \, dx = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{e^{itx}}{1+x^2} \, dx$$

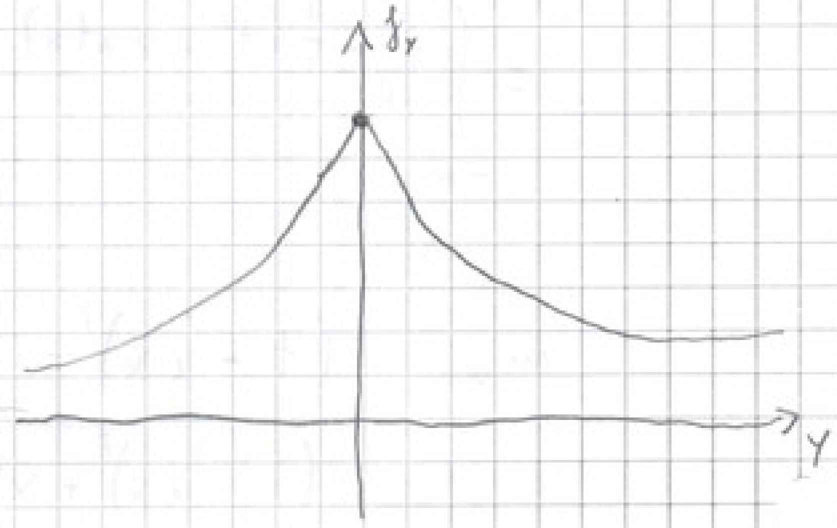
questa è un'integrazione complessa, non riusciamo a tirare fuori la i, ci fermiamo

Vediamo di calcolarla tramite un'altra distribuzione

ESEMPIO 3

$Y \sim \text{LAPLACE}(\lambda = 1)$

$$f_Y(y) = e^{-|y|} = \begin{cases} e^{-y} & \text{se } y \geq 0 \\ e^{y} & \text{se } y < 0 \end{cases}$$



$$\varphi_Y(t) = \int_{-\infty}^{+\infty} e^{ity} e^{-|y|} \, dy = \int_0^{+\infty} e^{-y(1-it)} \, dy + \int_{-\infty}^{0} e^{y(1+it)} \, dy =$$

poniamo $y(it-1) = x \qquad dy = \frac{dx}{it-1}$

$$= \int_0^{+\infty} \frac{e^{-x}}{1-it} \, dx + \int_{-\infty}^{0} \frac{e^{x}}{1+it} \, dx = \frac{1}{1-it} \left[e^{-x} \right]_0^{+\infty} + \frac{1}{1+it} \left[e^{x} \right]_{-\infty}^{0} =$$

$$= \frac{1}{1-it} \left(e^{-\infty} - e^{0} \right) + \frac{1}{1+it} \left(e^{0} - e^{-\infty} \right) =$$

$$= \frac{1}{1-it} (0 - 1) + \frac{1}{1+it} (1 - 0) \overset{?}{=} -\frac{1}{1-it} + \frac{1}{1+it} = \frac{2}{1+t^2}$$

È molto simile alla f_X della Cauchy

Si può dimostrare che vale la seguente uguaglianza

$$\int_{-\infty}^{+\infty} e^{-|y|} e^{ity} \, dy = \int_{-\infty}^{+\infty} e^{-|y|} e^{-ity} \, dy$$

Abbiamo visto che $(\mathcal{F} f_Y)(t) \simeq f_X(t)$

vale inoltre:

$$\int f_Y \, e^{iyt} \, dy = \int f_Y \, e^{-iyt} \, dy$$

ne segue che:

$$(\mathcal{F} \, f_X) = f_Y$$

cioè $\varphi_X(t) = f_Y(y) = e^{-|y|}$

Def: Dato un $A \subseteq E$, se $\mu^*(A) = \mu_*(A) = \mu(A)$ dove μ è detta misura di Lebesgue, A si dice Lebesgue-misurabile.

Th: Se $A \subseteq \bigcup_n A_n$ dove $\{A_n\}$ è una collezione numerabile, allora:

$$\mu^*(A) \leq \sum_n \mu^*(A_n)$$

Th: Tutti gli insiemi elementari sono Lebesgue-misurabili e tale misura coincide con m', cioè:

$$\forall A \in \mathcal{E} \Rightarrow \mu(A) = m'(A)$$

Dim:

Dato $A \in \mathcal{E}$ e $A = \bigcup_{i=1}^{K} P_i \in \mathcal{R}$ con $P_j \cap P_l = \phi \quad \forall j \neq l$

per definizione vale che:

$$m'(A) = \sum_{i=1}^{K} m(P_i)$$

Poiché i rettangoli P_i coprono tutto A, vale che

$$\mu^*(A) \leq \sum_i m(P_i) = m'(A) \qquad (1)$$

per definizione è la più piccola copertura

Presa una qualsiasi copertura di A $\{Q_j\}$ finita o numerabile, siccome m' è monotona varrà che:

$$m'(A) \leq \sum_j m(Q_j)$$

quindi se vale per qualsiasi copertura vale anche per la più piccola:

$$m'(A) \leq \mu^*(A) \qquad (2)$$

allora per (1) e (2)

$$\mu^*(A) = m'(A)$$

Vediamo anche per μ_*:

Poisson

Si indica con la coppia $N(t_i, t_g)$
Per comodità prenderemo $t_i = 0$ e quindi per un generico t:

$N(t) = N(0,t)$

Assiomi:

1) $N(0) = 0$ ovvero $P(N(0) = 0) = 1$
2) $N(t)$ ha incrementi indipendenti
3) $\forall t > 0 \quad P(N(t) > 0) < 1$

4) $$\lim_{h \to 0} \frac{P(N(t+h) - N(t) \geq 2)}{P(\underbrace{N(t+h) - N(t)}_{\text{incrementi}} = 1)} = 0$$

5) • $N(t) - N(s)$
• $N(t+h) - N(s+h)$
dovrebbero essere uguali per omogeneità sui tempi: invarianza per traslazione.

Possiamo indicare un processo di Poisson con la dicitura:
$\{N(t), t \geq 0\}$

Th: Un processo di conteggio che soddisfi i 5 assiomi è un processo di Poisson, cioè

$$\exists\, \nu > 0 : P(N(t) = k) = e^{-\nu t} \frac{(\nu t)^k}{k!} I_{\{0,1,\dots\}}(k)$$

Dim

Dimostrare il teorema si riduce a dimostrare che un processo di conteggio $N(t)$ che soddisfa i 5 assiomi ha funzione caratteristica:

$$\exists \nu : \varphi_{N(t)}(z) = e^{\nu t (e^{iz} - 1)}$$

questa è la funzione caratteristica di una poissoniana (che l'ho avevo calcolata prima)

Supponiamo che $\exists\ r>0$ tale che:

- $\lim_{h\to 0} \frac{1-P(N(h)=0)}{h} = r$
- $\lim_{h\to 0} \frac{P(N(h)=1)}{h} = r$
- $\lim_{h\to 0} \frac{P(N(h)\geq 2)}{h} = 0$

Osservazione:

$$\varphi_{N(t+h)}(z) = E\left(e^{izN(t+h)}\right) = E\left(e^{iz(N(t+h)-N(t)+N(t))}\right) =$$

$$= E\left(e^{iz(N(t+h)-N(t))}\, e^{izN(t)}\right) \underset{\substack{\uparrow\\ \text{incrementi} \\ \text{indipendenti}}}{=} E\left(e^{iz(N(t+h)-N(t))}\right) E\left(e^{izN(t)}\right) =$$

$$\underset{\text{omogeneità}}{=} E\left(e^{izN(h)}\right) E\left(e^{izN(t)}\right) = \varphi_{N(h)}(z)\, \varphi_{N(t)}(z)$$

Adesso scriviamo il rapporto incrementale del processo $N(t)$

$$\frac{1}{h}\Big(\underbrace{\varphi_{N(t+h)}(z) - \varphi_{N(t)}(z)}_{\text{incremento del processo } N(t,\,t+h)}\Big) = \frac{1}{h}\left(\varphi_{N(h)} - 1\right)\varphi_{N(t)}(z) =$$

Per comodità cambiamo notazione: $\varphi_{N(t)}(z) = \psi(z,t)$

$$= \underbrace{\frac{1}{h}\left(\psi(z,h) - 1\right)}\psi(z,t)$$

Studiamo il primo pezzo:

$$\frac{1}{h}\left(\psi(z,h) - 1\right) = \frac{1}{h}\left(E\left(e^{izN(h)}\right) - 1\right) =$$

$$= \frac{1}{h}\left(-1 + e^{iz0} P(N(h)=0) + e^{iz1} P(N(h)=1) + \sum_{k=2}^{+\infty} e^{izk} P(N(h)=k)\right)$$

Ora studiamo il limite di queste quantità:

$$\lim_{h \to 0} \frac{1}{h}\left(-1 + P(N(h)=0) + e^{iz} P(N(h)=1) + \sum_{K=2}^{+\infty} e^{izK} P(N(h)=K)\right) =$$

$$= -\nu + e^{iz}\nu + \lim_{h \to 0} \frac{1}{h} \sum_{K=2}^{+\infty} e^{izK} P(N(h)=K) =$$

OSSERVAZIONE:

$$\sum_{K=2}^{+\infty} e^{iKz} P(N(h)=K) \leq \sum_{K=2}^{+\infty} \underbrace{|e^{iKz}|}_{=1} P(N(h)=K)$$

allora:

$$\lim_{h \to 0} \frac{1}{h} \sum e^{izK} P(N(h)=K) = \lim_{h \to 0} \frac{1}{h} \sum P(N(h)=K) =$$

$$= P(N(h) \geq 2) \xrightarrow{h \to 0} 0$$

$$= \nu(e^{iz}-1)$$

Adesso studiamo il limite del rapporto incrementale (derivata):

$$\lim_{h \to 0} \frac{1}{h}\left(\psi(z,h)-1\right)\psi(z,t) = \nu\left(e^{iz}-1\right) \lim_{h \to 0} \psi(z,t) =$$

$$= \nu\left(e^{iz}-1\right)\psi(z,t)$$

quindi: $\frac{\partial}{\partial t}\psi(z,t) = \nu\left(e^{iz}-1\right)\psi(z,t)$

Valutiamo: $\psi(z,0) = E\left(e^{izN(0)}\right) = E\left(e^{iz0}\right) = 1$

Allora per trovare $\psi(z,t)$ basta risolvere l'equazione differenziale:

$$\psi(z,t) = \begin{cases} \frac{\partial}{\partial t}\psi(z,t) = \nu\left(e^{iz}-1\right)\psi(z,t) \\ \psi(z,0) = 1 \end{cases} = e^{\nu t(e^{iz}-1)}$$

che è proprio la funzione caratteristica della Poisson

Restano da dimostrare le supposizioni:

Indichiamo per comodità: $P_0(t) \equiv P(N(t)=0)$

Osservazione:

$$P(N(t_1+t_2)=0) = P(N(t_1+t_2)+N(t_1)-N(t_1)=0) =$$

$$= P(\underbrace{(N(t_1+t_2)-N(t_1))}=0 \wedge \underbrace{N(t_1)}=0) =$$

$$= P_0(t_2)\,P_0(t_1) \quad \text{per omogeneità e indipendenza di}$$

Th: $t>0$ $\quad f(t)$ reale e limitata su ogni intervallo limitato

Presi $t_1, t_2 > 0$ e vale che $f(t_1+t_2) = f(t_1)f(t_2)$

Allora:

- o $f(t) \equiv 0 \quad \forall t$
- o $\exists \nu : f(t) = e^{-\nu t}$

Quindi o $P_0(t) \equiv 0$ oppure $P_0(t) = e^{-\nu t}$

Possiamo escludere il primo caso perché:

$$0 < (P(N(t)>0)) < 1 \quad \forall t > 0$$

Adesso possiamo dimostrare le supposizioni:

- $\lim_{h \to 0} \frac{1-P_0(h)}{h} = \nu$

$$\frac{1-e^{-\nu h}}{h} = \frac{1-\overbrace{\left(1-\nu h+\frac{\nu^2 h^2}{2}+TOS\right)}^{\text{serie di Taylor}}}{h} = \frac{\nu \cancel{h} - \nu^2 h^{\cancel{2}} + TOS}{\cancel{h}}$$

$$\lim_{h\to 0} \nu - \nu^2 h + TOS = \nu$$

la prima supposizione è dimostrata

- $\lim_{h \to 0} \frac{P(N(h)=1)}{h} = \nu$

Possiamo scrivere che

$$\frac{1}{h} = \frac{P(N(h)=0) + P(N(h)=1) + P(N(h)\geq 2)}{h}$$

quindi

$$\frac{1 - P(N(h)=0)}{h} = \frac{1}{h}\Big(P(N(h)=1) + P(N(h)\geq 2)\Big) =$$

$$= \frac{P(N(h)=1)}{h}\left(1 + \frac{P(N(h)\geq 2)}{P(N(h)=1)}\right)$$

sappiamo che $\lim_{h \to 0} \frac{1 - P(N(h)=0)}{h} = \nu$

il $\lim_{h \to 0} \frac{P(N(h)\geq 2)}{P(N(h)=1)} = 0$ per l'assioma 4

Allora

$$\lim_{h \to 0} \frac{P(N(h)=1)}{h} = \nu$$

- $\lim_{h \to 0} \frac{P(N(h)\geq 2)}{h} = 0$

Possiamo scrivere che:

$$\frac{P(N(h)=0)}{h} + \frac{P(N(h)=1)}{h} + \frac{P(N(h)\geq 2)}{h} = \frac{1}{h}$$

$$\frac{1 - P(N(h)=0)}{h} = \frac{P(N(h)=1)}{h} + \frac{P(N(h)\geq 2)}{h}$$

$$\lim_{h \to 0} \frac{P(N(h)\geq 2}{h} = \lim_{h \to 0} \frac{1 - P(N(h)=0)}{h} - \lim_{h \to 0} \frac{P(N(h)=1)}{h} =$$

$$= \nu - \nu = 0$$

CVD

INFINITAMENTE DIVISIBILI

- Sia $N(t) \sim Poiss(\lambda)$

$$\varphi_{N(t)}(z) = e^{\lambda(e^{iz}-1)}$$

se ne calcolo la radice n-esima è ancora una Poisson?

$$\sqrt[n]{\varphi_{N(t)}(z} = \sqrt[n]{e^{\lambda(e^{iz}-1)}} = e^{\frac{\lambda}{n}(e^{iz}-1)}$$

Si : $N(t) \sim Poiss\left(\frac{\lambda}{n}\right)$

- Sia $X \sim Norm(0, \sigma^2)$

$$\varphi_X(t) = e^{-\frac{1}{2}t^2\sigma^2}$$

$$\sqrt[n]{\varphi_X(t)} = \sqrt[n]{e^{-\frac{1}{n}t^2\sigma^2}} = e^{-\frac{1}{2}t^2\frac{\sigma^2}{n}}$$

$X' \sim Norm(0, \frac{\sigma^2}{n}$

Proprietà: infinitamente divisibili

Se $\sqrt[n]{\varphi_Y(t)}$ è ancora la funzione caratteristica della stessa distribuzione di Y allora Y si dice infinitamente divisibile

Un esempio di distribuzione non infinitamente divisibile è la bernoulliana

Stimatori di massima verosimiglianza (ML)

Dati n campioni casuali:

$$X_1, X_2, \dots, X_n$$

dove $X \sim f(\cdot\,; \underline{\theta})$

definiamo funzione di verosimiglianza:

$$L(\theta) = f_{X_1, X_2, \dots, X_n}(x_1, \dots, x_n; \underline{\theta}) = \prod_{i=1}^{n} f_X(x_i; \theta)$$

Il valore $\hat{\theta}(\underline{x})$ che massimizza $L(\theta)$ è detto stima di massima verosimiglianza di $\underline{\theta}$ per il campione $\underline{x} = x_1, \dots, x_n$

Esempio

Date le v.c. $X_1, X_2, \dots, X_n \xrightarrow{\text{esperimento}} x_1, x_2, \dots x_n$

$\forall X_i \sim Bern(p)$

$$L(p) = f_{X_1, \dots X_n}(x_1, \dots, x_n; p) = \left(p^{x_1}(1-p)^{1-x_1}\right)\left(p^{x_2}(1-p)^{1-x_2}\right) \cdot \ldots =$$

$$= p^{\sum_{i=1}^{n} x_i} (1-p)^{n - \sum_{i=1}^{n} x_i} = p^{S_n}(1-p)^{n-S_n}$$

$$\frac{d}{dp} p^{S_n}(1-p)^{n-S_n} = S_n p^{S_n - 1}(1-p)^{n-S_n} - p^{S_n}(n-S_n)p^{n-S_n-1} =$$

$$= p^{S_n-1}(1-p)^{n-S_n-1}\left(S_n(1-p) - p(n-S_n)\right)$$

Si annulla quando:

$$S_n(1-p) - p(n-S_n) = 0$$

$$S_n - \cancel{S_n p} - pn + \cancel{p S_n} = 0$$

$$S_n - pn = 0$$

$$p = \frac{S_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Quindi lo stimatore ML di p è la statistica $\frac{1}{n}\sum_{i=1}^{n} X_n$

Indichiamo con $\mathcal{F}$ la famiglia di stimatori di massima verosimiglianza

Th: Dato $S \in \mathcal{F}$ stimatore di $\theta \Rightarrow g(S)$ è in $\mathcal{F}$ ed è stimatore di $g(\theta)$, per ogni funzione g

DEF: Stimatori Best Asymtotically Normal (BAN)

Una successione di stimatori $T_1, T_2, \dots, T_n, \dots$ di θ viene definita BAN sse soddisfa le 3 condizioni:

1) $\sqrt{n}\,(T_n - \theta) \underset{n\to\infty}{\sim} N(0, \sigma^2(\theta))$ Asymtotically Normal

2) $\forall \varepsilon > 0 \quad P(|T_n - \theta| > \varepsilon) = 0$

3) $\forall\ T_1', T_2', \dots, T_n', \dots \quad \sqrt{n}\,(T_n' - \theta) \underset{n\to\infty}{\sim} N(0, (\sigma'(\theta))^2)$

$$(\sigma'(\theta))^2 \geq (\sigma(\theta))^2$$ Best

Th: Data una successione $T_1, T_2, \dots, T_n, \dots$ di stimatori per θ dove ogni $T_j \in \mathcal{F}$

$$\{T_k\}_{k=1}^{+\infty} \text{ è BAN}$$

OSSERVAZIONE: $L(\theta)$ e $\log L(\theta)$ hanno i loro massimi per lo stesso valore di θ. A volte è più facile trovare il massimo del logaritmo della verosimiglianza

Questo è vero perché il logaritmo è una funzione monotona crescente

ESEMPIO

$\underline{N} = N_1, \dots, N_K$ v.c. identicamente distribuite

$N_i \sim N(0, \sigma^2)$ σ^2 sarà il parametro da stimare

Stimare σ^2 è come stimare σ

$$f_{\underline{N}}(\underline{x}, \sigma) = L(\sigma) = \prod_{j=1}^{K} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_j}{\sigma}\right)^2} = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^K e^{-\frac{1}{2}\frac{1}{\sigma}\sum_{j=1}^{K} x_j^2} =$$

chiameremo $S_K^2 = \sum_{j=1}^{K} x_j^2$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^K e^{-\frac{1}{2\sigma^2} S_K^2}$$

Ora dobbiamo calcolare i punti stazionari di $L(\sigma)$ per vedere dove è massima. Cioè dobbiamo calcolare la derivata:

$$\frac{d\,L(\sigma)}{d\sigma}$$

È sicuramente più comodo trattare il $\log L(\sigma)$

$$\ln(L(\sigma)) = \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^K + \ln\left(e^{-\frac{1}{2}\frac{S_K^2}{\sigma^2}}\right) =$$

$$= -K\left(\ln\sigma + \ln\sqrt{2\pi}\right) - \frac{1}{2}\frac{S_K^2}{\sigma^2}$$

$$\frac{d\ln(L(\sigma))}{d\sigma} = -\frac{K}{\sigma} + \frac{1}{2} S_K^2 \frac{2}{\sigma^3} = -\frac{K}{\sigma} + \frac{S_K^2}{\sigma^3} = \frac{1}{\sigma}\left(\frac{S_K^2}{\sigma^2} - K\right) = 0$$

Si annulla quando

$$\frac{S_K^2}{\sigma^2} - K = 0 \implies \sigma^2 = \frac{S_K^2}{K}$$

La $\underline{\text{stima}}$ sarà: $\hat{\sigma}^2 = \frac{S_K^2}{K} = \frac{\sum_{j=1}^{K} x_j^2}{K}$

Lo stimatore: $S_K = \frac{1}{K}\sum_{j=1}^{K} N_j^2$ è distorto, ma è BAN

?

CONFRONTO DI 2 MODELLI PROBABILISTICI PER LO STESSO FENOMENO

modello 1: (Ω, Σ_x, P_0)

modello 2: (Ω, Σ_x, P_1)

Useremo le seguenti ipotesi di lavoro:

- $|\Omega| < +\infty$
- $|X(\Omega)| = k$ $\quad$ X è una v.c. discreta

$$X(\Omega) = \{a_1, a_2, \dots, a_k\} \quad a \in \mathbb{R}$$

- $p_0 : \mathbb{R} \to \mathbb{R} \qquad p_0(x) = P_0(\{X = x\})$
- $p_1 : \mathbb{R} \to \mathbb{R} \qquad p_1(x) = P_1(\{X = x\})$
- $p_j(a_i) \neq 0 \qquad j = 0, 1 \qquad i = 1, \dots, k$

Abbiamo ottenuto il campione $x_1, x_2, \dots, x_n$ possiamo dire che:

$$\mathcal{L}_0(x_1, \dots, x_n) = \prod_{i=1}^{n} p_0(x_i) \qquad \mathcal{L}_1(x_1, \dots, x_n) = \prod_{i=1}^{n} p_1(x_i)$$

$$L_0(X_1, \dots, X_n) = \prod_{i=1}^{n} p_0(X_i) \qquad L_1(X_1, \dots, X_n) = \prod_{i=1}^{n} p_1(X_i)$$

Le ultime due funzioni sono v.c., le indicheremo rispettivamente con L_0 e L_1

Ci chiediamo adesso quanto valgono i valori attesi:

$$E_0(L_0) = \sum_j v_j \, P_0(L_0 = v_j)$$

$$E_1(L_0) = \sum_j v_j \, P_1(L_0 = v_j)$$

L'idea di base che ci guiderà è:

Se $a > b \Rightarrow \frac{a}{b} > 1 \Rightarrow \ln\left(\frac{a}{b}\right) > 0$

Se $a < b \Rightarrow \frac{a}{b} < 1 \Rightarrow \ln\left(\frac{a}{b}\right) < 0$

Scriviamo:

$$\mu^*(E \setminus A) = m'(E \setminus A)$$
$$1 - \mu_*(A) = m'(E \setminus A) = 1 - m'(A)$$

Allora vale anche

$$\mu_*(A) = m'(A)$$

quindi: $$m'(A) = \mu^*(A) = \mu_*(A) = \mu(A)$$

CVD

Il teorema precedente ci dice che μ è estensione di m'.

Vediamo le proprietà di μ:

- Positiva
- Monotona $\quad A \subseteq A' \implies \mu(A) \leq \mu(A')$

  e anche:

  $$A \subseteq \bigcup_n A_n \qquad \mu(A) \leq \mu(\bigcup_n A_n) \leq \sum_n \mu(A_n)$$

- Additiva: dati $B, C \subseteq E$ $\quad A = B \cup C$ e $B \cap C = \emptyset$

  vale che:

  $$\mu(A) = \mu(B) + \mu(C)$$

Poiché in (Ω, Σ_X, P_0) $L_0 > L_1$ vale che:

$$E_0\left(\frac{L_0}{L_1}\right) > 1 \qquad E_1\left(\frac{L_0}{L_1}\right) < 1$$

con $\frac{L_0}{L_1}$ indichiamo il rapporto di verosimiglianza

Per semplificare i calcoli dimostreremo che:

$$E_0\left(\ln \frac{L_0}{L_1}\right) > 0 \qquad E_1\left(\ln \frac{L_0}{L_1}\right) < 0$$

DIM

$$E_0\left(\ln \frac{L_0}{L_1}\right) = E_0\left(\ln \frac{\prod_{j=1}^{n} f_0(X_j)}{\prod_{j=1}^{n} f_1(X_j)}\right) = E_0\left(\ln\left(\prod_{j=1}^{n} \frac{f_0(X_j)}{f_1(X_j)}\right)\right) =$$

$$= E_0\left(\sum_{j=1}^{n} \ln\left(\frac{f_0(X_j)}{f_1(X_j)}\right)\right) = \sum_{j=1}^{n} E_0\left(\ln \frac{f_0(X_j)}{f_1(X_j)}\right) =$$

poiché X_j sono identicamente distribuite a una v.c. X

$$= n\, E_0\left(\ln \frac{f_0(X)}{f_1(X)}\right) = n \sum_{i=1}^{k} \ln \frac{f_0(a_i)}{f_1(a_i)}\, f_0(a_i) =$$

$$= n \sum_{i=1}^{k} \left(f_1(a_i)\, \frac{f_0(a_i)}{f_1(a_i)}\, \ln \frac{f_0(a_i)}{f_1(a_i)}\right) =$$

Per Jensen data una funzione g convessa vale che:

$$E\left(g(X)\right) \geq g\left(E(X)\right)$$

prendiamo $g(x) = x \ln x$ avremo che:

$$\sum_{i=1}^{k} f_1(a_i)\, g\left(\frac{f_0(a_i)}{f_1(a_i)}\right) = E_1\left(g\left(\frac{f_0(X)}{f_1(X)}\right)\right)$$

quindi:

$$= n\, E_1\left(g \frac{f_0(X)}{f_1(X)}\right) \geq n\, g\left(E_1\left(\frac{f_0(X)}{f_1(X)}\right)\right) = n\, E_1\left(\frac{f_0(X)}{f_1(X)}\right) \ln\left(E_1\left(\frac{f_0(X)}{f_1(X)}\right)\right) =$$

$$= n \sum_{i=1}^{K} p_1(a_i) \frac{p_0(a_i)}{p_1(a_i)} \quad \ln\left(\sum_{i=1}^{K} \cancel{p_1(a_i)} \frac{p_0(a_i)}{\cancel{p_1(a_i)}} \right) \leq n \cdot 1 \cdot \ln 1 = 0$$

Abbiamo dimostrato che:

$$E_0\left(\ln \frac{L_0}{L_1} \right) \geqslant 0$$

Ma Jensen mette l'uguale per coprire ipotesi banali; noi abbiamo modelli diversi allora possiamo usare il maggiore stretto; ovvero $\exists i : p_1(x_i) \neq p_0(x_i)$

CVD

Abbiamo visto che:

$$E_0\left(\ln \frac{L_0}{L_1} \right) = n\, E_0\left(\boxed{\ln \frac{p_0(x)}{p_1(x)}} \right)$$

questa qui giocherà un ruolo centrale nella nostra discussione

Possiamo scrivere che

$$\boxed{\frac{L_0}{L_1}} = e^{n J_n}$$

perché questa quantità è positiva, e J_n sarà:

$$J_n = \frac{1}{n} \ln \frac{L_0}{L_1} = \underbrace{\frac{1}{n} \sum_{j=1}^{n} \ln \frac{p_0(x_j)}{p_1(x_j)}}_{\text{media campionaria}}$$

Introduciamo due grandezze:

$$I_{0,1} = E_0\left(\ln \frac{p_0(x)}{p_1(x)} \right) = \sum_{i=1}^{k} p_0(a_i) \ln \frac{p_0(a_i)}{p_1(a_i)} > 0$$

$$-I_{1,0} = E_1\left(\ln \frac{p_0(x)}{p_1(x)} \right) < 0$$ in modo che $I_{1,0}$ sia positivo

Calcoliamo

$$E_0(J_n) = \frac{1}{n} E_0\left(\sum_{j=1}^{n} \ln \frac{p_0(x_j)}{p_1(x_j)} \right) = E_0\left(\ln \frac{p_0(X)}{p_1(X)} \right) = I_{0,1}$$

$$E_1(J_n) = -I_{1,0}$$

$$Var_0(J_n) = \frac{1}{n} \, var_0\left(\ln \frac{f_0(x)}{f_1(x)}\right) = \frac{1}{n} \sigma^2_{0,1}$$

$$Var_1(J_n) = \frac{1}{n} \sigma^2_{1,0}$$

Per il teorema del limite centrale possiamo approssimare:

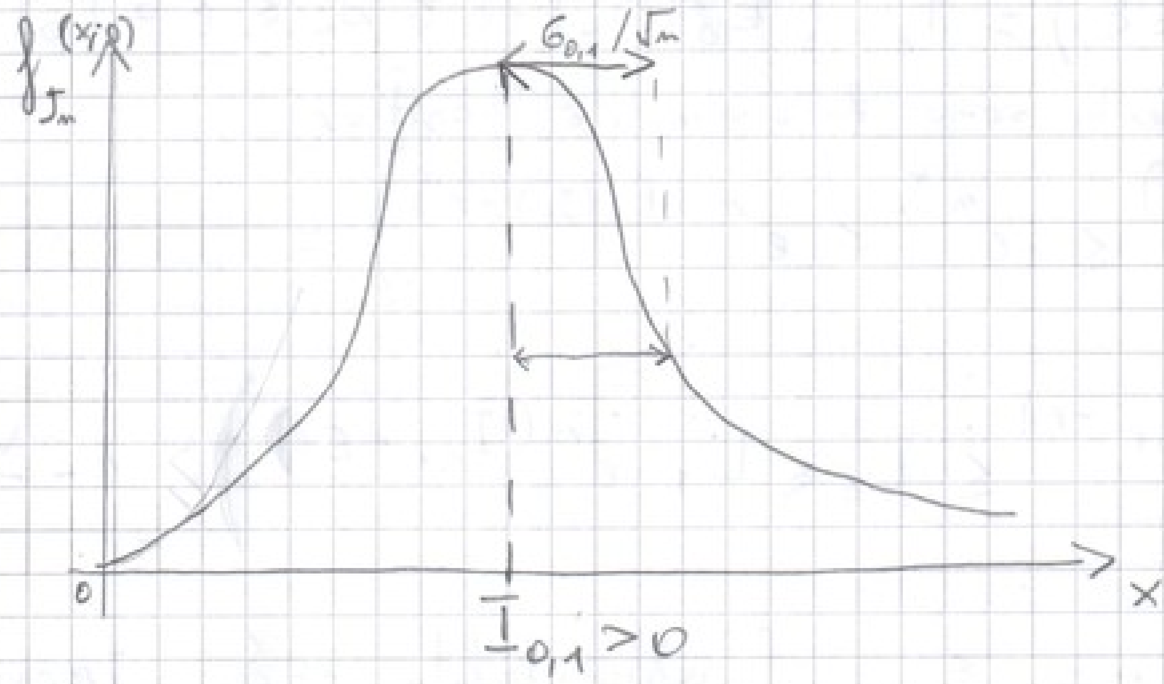
$$J_n \sim N\left(I_?, \frac{\sigma^2_?}{n}\right)$$

Avendo la struttura della media campionaria, J_n, per un n grande, ha varianza piccola e comportamento NORMALE

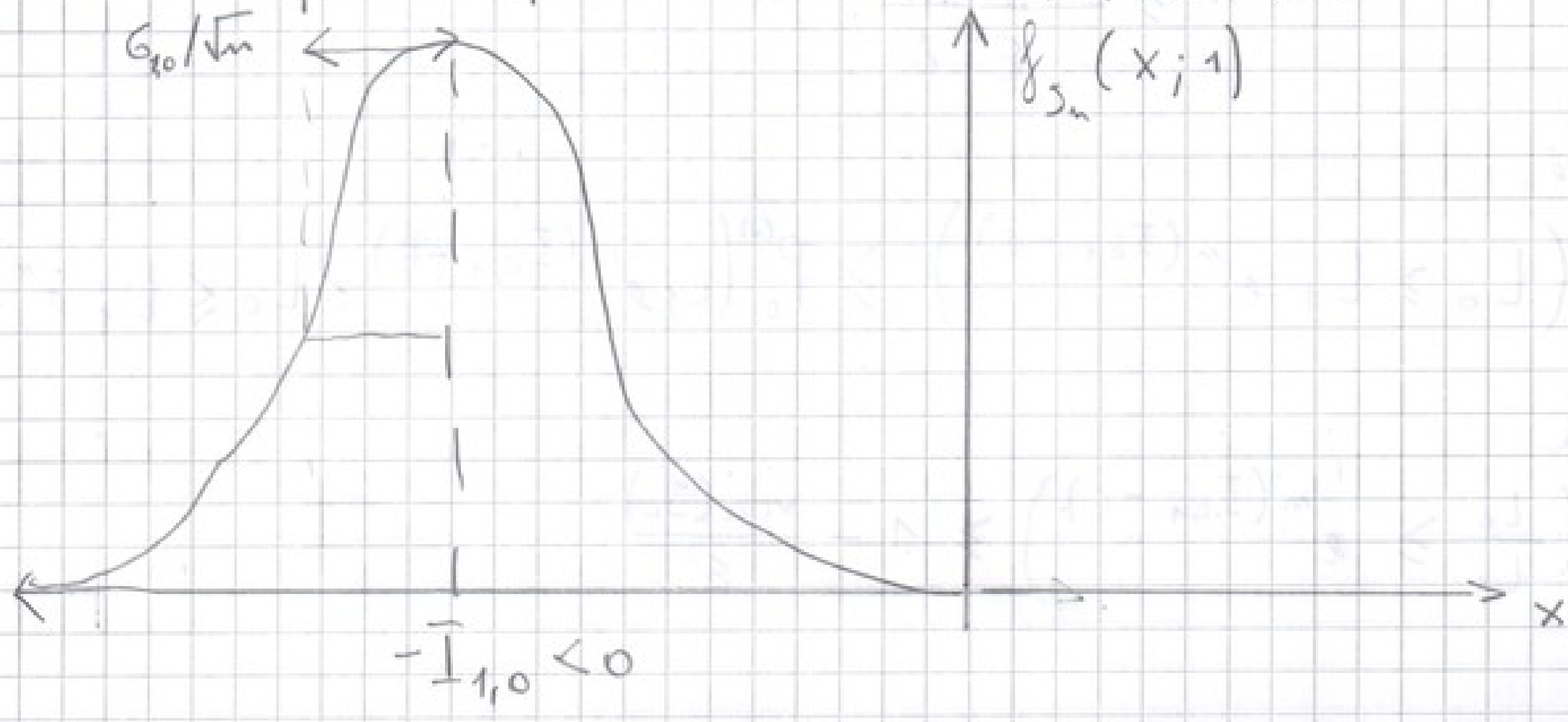
Riassumendo:

1) Sotto l'ipotesi P_0 l'andamento qualitativo di J_n, nell'approssimazione normale è il seguente:

$$J_n \sim N\left(I_{0,1}, \frac{1}{n}\sigma^2_{0,1}\right)$$



2) Sotto l'ipotesi P_1 $\quad J_n \sim N\left(I_{1,0}, \frac{1}{n}\sigma^2_{1,0}\right)$



La diseguaglianza di Tchebychev ci aiuterà a rendere quantitativo le nostre considerazioni qualitative sul comportamento delle fluttuazioni intorno al valore medio:

$$P(|J_n - E(J_n)| > \varepsilon) < \frac{var(J_n)}{\varepsilon^2}$$

Ma sia il valore atteso che la varianza dipendono dal modello

$$\frac{L_0}{L_1} = e^{nJ_n} \overset{H_0}{\simeq} e^{nI_{0,1}} \Rightarrow L_0 \overset{H_0}{\simeq} L_1 e^{nI_{0,1}}$$

$$\frac{L_0}{L_1} = e^{nJ_n} \overset{H_1}{\simeq} e^{-nI_{1,0}} \Rightarrow L_0 \overset{H_1}{\simeq} L_1 e^{-nI_{1,0}}$$

dove H_0 e H_1 sono ipotesi che siamo rispettivamente nel modello e nel modello 1

Allora posso dire che:

$$P_0^{(n)}(|J_n - E_0(J_n)| \leq \varepsilon) = P_0^{(n)}(E_0(J_n) - \varepsilon \leq J_n \leq \varepsilon + E_0(J_n)) =$$

moltiplico tutto per n tanto sono tutte quantità positive

$$= P_0^{(n)}\left(e^{n(E_0(J_n)-\varepsilon)} \leq e^{nJ_n} \leq e^{n(E_0(J_n)+\varepsilon)}\right) =$$

$$= P_0^{(n)}\left(L_1 e^{n(I_{0,1}-\varepsilon)} \leq L_0 \leq L_1 e^{n(I_{0,1}+\varepsilon)}\right) \geq 1 - \overbrace{\frac{var_0(J_n)}{\varepsilon^2}}^{\delta}$$

Fissate la tolleranza ε e l'affidabilità $1-\delta$ troviamo un n:

$$n \geq \frac{var(J_n)}{\varepsilon^2 \delta}$$

Poiché:

$$P_0^{(n)}\left(L_0 \geq L_1 e^{n(I_{0,1}-\varepsilon)}\right) \geq P_0^{(n)}\left(L_1 e^{n(I_{0,1}-\varepsilon)} \leq L_0 \leq L_1 e^{n(I_{0,1}+\varepsilon)}\right)$$

allora

$$P_0^{(n)}\left(\frac{L_0}{L_1} \geq e^{n(I_{0,1}-\varepsilon)}\right) \geq 1 - \frac{var(J_n)}{\varepsilon^2}$$

$$\forall \varepsilon, \delta, n : P_0^{(n)}\left(\frac{L_0}{L_1} \geq e^{n(I_{0,1}-\varepsilon)}\right) \geq 1-\delta \qquad (1)$$

Per P_1 si fa lo stesso ragionamento e si ottiene:

$$\forall \varepsilon, \eta, n : P_1^{(n)}\left(\frac{L_0}{L_1} \leq e^{n(\varepsilon - I_{1,0})}\right) \geq 1-\eta \qquad (2)$$

Supponiamo che $\varepsilon < I_{0,1}$ allora $e^{n(I_{0,1}-\varepsilon)} > 1$

Supponiamo che $\varepsilon < I_{1,0}$ allora $e^{n(\varepsilon - I_{1,0})} < 1$

Definisco due eventi:

$$A_n(\varepsilon) = \left\{ \omega \in \Omega^n : \frac{L_0}{L_1} \geq \underbrace{e^{n(I_{0,1}-\varepsilon)}}_{>1} \right\}$$

$$C_n(\varepsilon) = \left\{ \omega \in \Omega^n : \frac{L_0}{L_1} \leq \underbrace{e^{n(\varepsilon - I_{1,0})}}_{<1} \right\}$$

$A_n(\varepsilon) \cap C_n(\varepsilon) = \emptyset$ per via delle supposizioni precedenti

Questi due eventi possono aiutarmi a discriminare il modello giusto.

Fissate le affidabilità $\delta > 0$ e $\eta > 0$, e preso un campione di dimensione sufficiente a soddisfare la (1) e la (2) ovvero

$$n \geq \max\left\{ \frac{\sigma_{0,1}^2}{\delta \varepsilon^2}, \frac{\sigma_{1,0}^2}{\eta \varepsilon^2} \right\}$$

allora:

$$P_0^{(n)}\left(A_n(\varepsilon)\right) \geq 1-\delta \qquad P_1^{(n)}\left(C_n(\varepsilon)\right) \geq 1-\eta$$

e quindi

$$P_0^{(n)}\left((A_n(\varepsilon))^c\right) < \delta \qquad P_1^{(n)}\left((C_n(\varepsilon))^c\right) < \eta$$

Ma

$$\left(A_n(\varepsilon)\right)^c = \left\{ \omega \in \Omega^n : \frac{L_0}{L_1}(\omega) < e^{n(I_{0,1}-\varepsilon)} \right\} \supseteq C_n$$

perché

$$e^{n(I_{0,1}-\varepsilon)} > e^{-n(I_{1,0}-\varepsilon)} \qquad \text{se } I_{0,1} > I_{1,0}$$

allora

$$P_0^{(n)}\left(C_n(\varepsilon)\right) \leq \delta$$

ovvero sotto H_0 è piccola la probabilità di ottenere valori bassi nel rapporto di verosimiglianza $\frac{L_0}{L_1}$. E' invece prossima a 1 la probabilità di avere rapporti di verosimiglianza grandi.

Riassumendo:

$$\begin{cases} P_1^{(n)}\left(A_n(\varepsilon)\right) \leq e^{-n(I_{0,1}-\varepsilon)} \\ P_0^{(n)}\left(A_n(\varepsilon)\right) \geq 1 - \dfrac{\sigma_{0,1}^2}{\varepsilon^2 n} \end{cases}$$

$$\begin{cases} P_0^{(n)}\left(C_n(\varepsilon)\right) \leq e^{-n(I_{1,0}-\varepsilon)} \\ P_1^{(n)}\left(C_n(\varepsilon)\right) \geq 1 - \dfrac{\sigma_{1,0}^2}{\varepsilon^2 n} \end{cases}$$

Abbiamo quindi individuato 2 sottoinsiemi di Ω^n, A_n e C_n che godono delle seguenti proprietà

- in H_0 la probabilità di $A_n(\varepsilon)$ è alta, mentre quella di C_n è bassa
- in H_1 la probabilità di $A_n(\varepsilon)$ è bassa, mentre quella di C_n è alta

L'esistenza di questi due insiemi ci permette di verificare sperimentalmente la validità delle due ipotesi H_0 e H_1:

a partire da un campione

$$x_1, \dots, x_n \longrightarrow \frac{\mathcal{L}_0(\underline{x})}{\mathcal{L}_1(\underline{x})} \longrightarrow \text{VALORE NUMERICO} \begin{cases} \in \underline{X}\left(A_n(\varepsilon)\right) & \text{accetto } H_0 \\ \in \underline{X}\left(C_n(\varepsilon)\right) & \text{accetto } H_1 \end{cases}$$

Tuttavia la probabilità di sbagliare è non nulla:

- è possibile rifiutare H_0 quando questo è vero con probabilità δ
Questo si chiama errore di I tipo

$$\begin{cases} P_0(A_n(\epsilon)) \geq 1-\delta \\ P_0(C_n(\epsilon)) < \delta \end{cases}$$

- è possibile rifiutare H_1 quando questo è vero con probabilità η
Questo si chiama errore di II tipo

$$\begin{cases} P_1(C_n(\epsilon)) \geq 1-\eta \\ P_1(A_n(\epsilon)) < \eta \end{cases}$$

E' chiaro che se diminuiamo δ η aumenterà e viceversa
Bisogna trovare un compromesso

CRITERIO DI OTTIMALITÀ

Dato un evento B_n e le due condizioni:

a) il modello P_0 attribuisce a B_n probabilità BASSA

b) il modello P_1 attribuisce a B_n probabilità ALTA

Voglio determinare B_n^* tale che:

1) B_n^* gode della proprietà (a)

2) Tra tutti gli eventi che godono della proprietà (a), B_n^* gode al massimo grado della proprietà (b)

Il lemma seguente dimostrerà che C_n è ottimale in questo senso!

Lemma di Neyman-Pearson

Un qualsiasi evento $B_n \subset \Omega_n$ che soddisfi la condizione:

$$P_0^{(n)}(B_n) \leq P_0^{(n)}(C_n(\epsilon))$$

soddisfa anche la condizione:

$$P_1^{(n)}(B_n) \leq P_1^{(n)}(C_n(\epsilon))$$

$C_n(\epsilon)$ si dice massimamente discriminante

Dim:

$$\tau = e^{-n(I_{1,0}-\varepsilon)}$$

$$\underline{X}(C_n(\epsilon)) = \left\{ \underline{x} \in \mathbb{R}^n : \frac{\mathcal{L}_0(\underline{x})}{\mathcal{L}_1(\underline{x})} \leq \tau \right\}$$

Analogamente

$$\underline{X}(B_n) = \left\{ \underline{x} \in \mathbb{R}^n : x_i = X_i(\underline{\omega}),\ \underline{\omega} \in B_n,\ i = 1 \dots n \right\}$$

$$P_1^{(n)}(C_n(\epsilon)) - P_1^{(n)}(B_n) = \sum_{\underline{x} \in \underline{X}(C_n(\epsilon))} \mathcal{L}_1(\underline{x}) - \sum_{\underline{x} \in \underline{X}(B_n)} \mathcal{L}_1(\underline{x}) =$$

osserviamo che $\underline{X}(C_n(\epsilon)) \cap \underline{X}(B_n) \neq \phi$ tuttavia le probabilità dell'intersezione si eliminano a due a due nella differenza; possiamo quindi scrivere:

$$= \sum_{x \in \underline{X}(C_n(\epsilon)) \setminus \underline{X}(B_n)} \mathcal{L}_1(\underline{x}) - \sum_{\underline{x} \in \underline{X}(B_n) \setminus \underline{X}(C_n(\epsilon))} \mathcal{L}_1(\underline{x})$$

Ricordiamo che $\forall\, \underline{x} \in \underline{X}(C_n(\epsilon))$ vale:

$$\frac{\mathcal{L}_0(\underline{x})}{\mathcal{L}_1(\underline{x})} \leq e^{-n(I_{1,0}-\varepsilon)}$$

cioè $\mathcal{L}_1(\underline{x}) \geq \frac{\mathcal{L}_0(\underline{x})}{\tau}$

$$\sum_{x \in \underline{X}(C_n(\epsilon)) \setminus \underline{X}(B_n)} \mathcal{L}_1 \geq \sum_{x \in \underline{X}(C_n(\epsilon)) \setminus \underline{X}(B_n)} \mathcal{L}_0 / \tau$$

Ma allora:

$$P_1(C_n(\varepsilon)) - P_1(B_n) \geq \frac{1}{\tau}\left(\sum_{x \in \underline{X}(C_n(\varepsilon)) \setminus \underline{X}(B_n)} \mathcal{L}_0(\underline{x}) - \sum_{x \in \underline{X}(B_n) \setminus \underline{X}(C_n(\varepsilon))} \mathcal{L}_0(\underline{x})\right) =$$

$$= \frac{1}{\tau}\left(P_0^{(n)}(C_n(\varepsilon)) - P_0^{(n)}(B_n)\right)$$

Per ipotesi del lemma

$$P_0^{(n)}(B_n) \leq P_0^{(n)}(C_n(\varepsilon))$$

quindi:

$$\frac{1}{\tau}\left(P_0^{(n)}(C_n(\varepsilon)) - P_0^{(n)}(B_n)\right) \geq 0$$

CVD

Corollario (si ottiene scambiando gli indici 0 e 1...)

Un qualsiasi evento $D_n \subset \Omega_n$ che soddisfa la condizione

$$P_1^{(n)}(D_n) \leq P_1^{(n)}(A_n(\varepsilon))$$

soddisfa la condizione

$$P_0^{(n)}(A_n(\varepsilon)) \geq P_0^{(n)}(D_n)$$

TEST DI IPOTESI: contesto applicativo

In questo corso abbiamo lavorato sotto ipotesi del tutto generali, cioè non abbiamo fatto ipotesi sulla distribuzione del campione.

Nel mondo reale i test di ipotesi si fanno a partire da un'ipotesi di normalità (o approssimativamente).

Th: CNS affinché A sia misurabile, cioè $A \in \mathcal{M}$, è

$$\forall \varepsilon > 0 \quad \exists B \in \mathcal{E} \quad \text{tale che} \quad \mu^*(A \triangle B) < \varepsilon$$

LEMMA: Dati due insiemi arbitrari $A, B \subseteq E$ vale:

$$|\mu^*(A) - \mu^*(B)| \leq \mu^*(A \triangle B)$$

Un insieme è misurabile sse lo si può approssimare con una precisione a piacere mediante insiemi elementari.

DIM LEMMA:

$$A \subset B \cup (A \triangle B)$$

possiamo dunque scrivere

$$\mu^*(A) \leq \mu^*(B) + \mu^*(A \triangle B)$$

CVD

LEZ III 8/3/10

Th: L'unione e l'intersezione di un'infinità numerabile di insiemi misurabili è misurabile

~~DIM UNIONE: ?~~

~~Dato $A = \bigcup_{n=1}^{\infty} A_n$ con ogni $A_n \in M$ deve valere che $A \in M$~~

~~Scriviamo:~~

$$A'_n = A_n \setminus \bigcup_{K=1}^{n-1} A_K$$

$$A = \bigcup_n A_n = \bigcup_n A'_n$$

~~Per il teorema precedente A è misurabile, scelto un $B \in M$ vale che:~~ $\mu^*(A \triangle B) < \varepsilon$

$$A = \bigcup_n A'_n = \underbrace{\boxed{\bigcup_{n=1}^{N} A'_n}}_{= C_N} \underbrace{\boxed{\bigcup_{n=N+1}^{\infty} A'_n}}_{= K_N}$$

$$\mu^*(C_N \triangle B) < \varepsilon$$

... booooo!

Disuguaglianza di Cramer-Rao

Nel corso della discussione sul test di ipotesi è comparsa la grandezza

$$I_{0,1} = E_0\left(\ln \frac{L_0}{L_1}\right) = n\, E_0\left(\ln \frac{p_0(x)}{p_1(x)}\right) =$$

Supponiamo ora che P_0 e P_1 differiscano solo per il valore θ_0 e θ_1, quindi:

$$P_0 \quad f(\cdot\,;\theta_0)$$
$$P_1 \quad f(\cdot\,;\theta_1)$$

in questo caso possiamo riscrivere:

$$I_{0,1} = \sum_x f_X(x,\theta_0)\ \ln \frac{f_X(x;\theta_0)}{f_X(x;\theta_1)} = \underline{I}(\theta_0,\theta_1)$$

Adesso fissiamo θ_0 e studiamo $\underline{I}(\theta_0,\theta_1)$ con lo sviluppo di Mc Laurin

$$\underline{I}_{\theta_0}(\theta_1) = \underline{I}(\theta_0,\theta_1)\big|_{\theta_1=\theta_0} + \frac{d}{d\theta_1} I(\theta_0,\theta_1)\big|_{\theta_0}\cdot(\theta_1-\theta_0) +$$
$$+ \frac{d^2}{d\theta_1^2} I(\theta_0,\theta_1)\big|_{\theta_1=\theta_0}\frac{(\theta_1-\theta_0)^2}{2} + TOS =$$

$$= \underbrace{\underline{I}(\theta_0,\theta_0)}_{\substack{=0 \\ \text{perché } \ln 1 = 0}} - (\theta_1-\theta_0)\sum_x f'(x;\theta_0) + \frac{(\theta_1-\theta_0)^2}{2}\left(\sum_x \frac{f'(x;\theta_0)^2}{f(x;\theta_0)} - f''(x,\theta_0)\right) +$$

$$+ TOS =$$

$$\text{ma } \sum_x f'(x,\theta) = \frac{d}{d\theta}\sum_x f(x,\theta) = \frac{d}{d\theta} 1 = 0$$

possiamo farlo perché non abbiamo problemi di convergenza

lo stesso vale per $\sum f''(x,\theta)$

$$= \frac{(\theta_1-\theta_0)^2}{2}\sum_x \frac{f'(x,\theta_0)^2}{f(x,\theta_0)} + TOS$$

Una buona approssimazione sarà dunque:

$$I_{\theta_0}(\theta_1) \simeq n\frac{(\theta_1-\theta_0)^2}{2}\sum_x \frac{f'(x;\theta_0)^2}{f(x;\theta_0)} = \frac{n(\theta_1-\theta_0)^2}{2}\sum_x\left(\frac{f'(x;\theta_0)}{f(x;\theta_0)}\right)^2 f(x;\theta_0) =$$

$$= \frac{n(\theta_1-\theta_0)^2}{2} E_0\left(\left(\frac{d}{d\theta_0}\ln f_x(x;\theta_0)\right)^2\right)$$

In particolare

$$I(\theta,\theta+\delta) \simeq \frac{n\delta^2}{2} E_0\left(\left(\frac{d}{d\theta}\ln f_x(x;\theta)\right)^2\right)$$

Il log di una probabilità fa pensare all' INFORMAZIONE

E' come se studiassi la media della variazione (derivata) dell' informazione associata alla probabilità rispetto a θ

Infatti l' informazione di Fisher è proprio così definita:

$$I(\theta) = E\left(\left(\frac{d}{d\theta}\log(f(x,\theta))\right)^2\right) \qquad \text{se esiste}$$

misura la dipendenza media dell' informazione dal parametro θ

Cioè, è tanto più alta quanto più alta è la dipendenza dal parametro della v.c. X.

Quindi, intuitivamente, se una legge dipende in media fortemente dal suo parametro, parametri diversi denoteranno distribuzioni MOLTO diverse.

ESEMPIO 1

$Y \sim Exp(\lambda)$ $\qquad\qquad f_y(x,\lambda) = \lambda e^{-\lambda x} I_{(0,+\infty)}(x)$

Studiamo il logaritmo:

$$\log(\lambda e^{-\lambda x}) I_{(0,+\infty)}(x) = \log(\lambda)-\lambda x)\, I_{(0,+\infty)}$$

ora la derivata

$$\frac{d}{d\lambda}\log(\lambda)-\lambda x) = \frac{1}{\lambda} - Y$$

$$E_\lambda\left(\left(\frac{1}{\lambda}-Y\right)^2\right) \underset{\uparrow}{=} Var(Y) = \frac{1}{\lambda^2}$$

perché $\frac{1}{\lambda}$ è il valore atteso dell' exp

ESEMPIO 2

$X \sim N(\theta, \sigma^2)$ si suppone varianza nota

$$f_N(x;\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2\sigma^2}}$$

$$\ln f_N(x,\theta) = -\ln\left(\sigma\sqrt{2\pi}\right) - \frac{(x-\theta)^2}{2\sigma^2}$$

$$\frac{d}{d\theta} \ln f_N(x,\theta) = \frac{d}{d\theta}\left(-\frac{x^2+\theta^2-2x\theta}{2\sigma^2}\right) = -\frac{-\cancel{2}x+\cancel{2}\theta}{\cancel{2}\sigma^2} = \frac{\theta - x}{\sigma^2}$$

$$E\left(\left(\frac{\theta-x}{\sigma^2}\right)^2\right) = \frac{1}{\sigma^4} E\left((\theta-x)^2\right) = \frac{1}{\sigma^4}\sigma^2 = \frac{1}{\sigma^2}$$

OSSERVAZIONE: la variazione media dell'informazione è zero, ovvero

$$E_\theta\left(\frac{d}{d\theta}\ln f(x,\theta)\right) = 0$$

DIM: per le v.c. continue:

$$E_\theta\left(\frac{d}{d\theta}\ln f(x,\theta)\right) = \int_{-\infty}^{+\infty}\left(\frac{d}{d\theta}\ln f(x,\theta)\right) f(x,\theta)\,dx = \int_{-\infty}^{+\infty}\frac{d}{d\theta} f(x,\theta)\,dx =$$

$$= \frac{d}{d\theta}\int_{-\infty}^{+\infty} f(x,\theta)\,dx = \frac{d}{d\theta} 1 = 0$$

Dimostrato questo possiamo dire che

$$E_\theta\left(\left(\frac{d}{d\theta}\ln f(x,\theta)\right)^2\right) = E_\theta\left(\left(\frac{d}{d\theta}\ln f(x,\theta) - 0\right)^2\right) = \mathrm{var}_\theta\left(\frac{d}{d\theta}\ln f(x,\theta)\right) = I(\theta)$$

Questo ci dice che l'informazione di Fisher è pari alle fluttuazioni medie della dipendenza della distribuzione dal parametro.

OSSERVAZIONE: dato un campione $X_1, \dots, X_n$ l'informazione del campione è

$$I_n(\theta) = n I(\theta)$$

Borto pag 391

Consideriamo, per semplicità, stimatori non distorti per il parametro θ o per una sua funzione $\tau(\theta)$

Che relazione sussiste tra le fluttuazioni della statistica e le fluttuazioni del campione?

DISUGUAGLIANZA DI CRAMER-RAO:

$$\mathrm{var}_\theta(T_n) \geq \frac{\left(\frac{d\tau(\theta)}{d\theta}\right)^2}{n\,\underline{I}(\theta)}$$

DIM

Osservazione 1: $\int_{\mathbb{R}^n} \prod_{i=1}^{n} f(x_i,\theta)\, d^n x = 1$

Osservazione 2: $\tau(\theta) = E_\theta(T_n)$

Cominciamo:

$$\frac{d\tau(\theta)}{d\theta} = \frac{d}{d\theta} E_\theta(T_n) = \frac{d}{d\theta}\int_{\mathbb{R}^n} \underbrace{t(x_1,\dots,x_n)}_{\substack{\text{valore assunto dalla statistica} \\ \text{una volta estratto il campione}}} \prod_{i=1}^{n} f(x_i,\theta)\, d^n x =$$

$$= \int_{\mathbb{R}^n} d^n x\; t(x_1,\dots,x_n)\, \frac{d}{d\theta}\prod_{i=1}^{n} f(x_i,\theta) =$$

$$= \int_{\mathbb{R}^n} d^n x\; t(x_1,\dots,x_n)\, \frac{\prod_{i=1}^{n} f(x_i,\theta)}{\prod_{i=1}^{n} f(x_i,\theta)}\, \frac{d}{d\theta}\prod_{i=1}^{n} f(x_i,\theta) =$$

$$= \int_{\mathbb{R}^n} d^n x\; t(x_1,\dots,x_n) \prod_{i=1}^{n} f(x_i,\theta)\, \frac{d}{d\theta} \ln\left(\prod_{i=1}^{n} f(x_i,\theta)\right) =$$

$$= E_\theta\left(T\, \frac{d}{d\theta} \ln\left(\prod_{i=1}^{n} f(X_i,\theta)\right)\right) = E_\theta\left((T-\tau(\theta))\frac{d}{d\theta}\ln\left(\prod_{i=1}^{n} f(X_i,\theta)\right)\right)$$

è come se avessimo sottratto θ perché $E\left(\tau(\theta)\ln\left(\prod f(x_i,\theta)\right)\right)=0$?

Calcoliamo adesso $\left(\frac{d\tau(\theta)}{d\theta}\right)^2$:

Cauchy-Swarz

$$\left(\frac{d\tau(\theta)}{d\theta}\right)^2 = \left(E_\theta\left((T-\tau(\theta))\frac{d}{d\theta}\ln\left(\prod_{i=1}^{n} f(X_i,\theta)\right)\right)\right)^2 \leq$$

$$\leq E_\theta\left((T-\tau(\theta))^2\right)\cdot E_\theta\left(\left(\frac{d}{d\theta}\ln\prod f(X_i,\theta)\right)^2\right) = var(T)\cdot nI(\theta)$$

CVD

Def: Uno stimatore non distorto per $\tau(\theta)$ è detto EFFICIENTE sse la sua varianza satura la disuguaglianza di Cramer-Rao, ovvero quando vale l'uguaglianza:

$$\left(\frac{d\tau(\theta)}{d\theta}\right)^2 = E_\theta\left((T-\tau(\theta))^2\right) nI(\theta)$$

la quantità $\frac{nI(\theta)}{var(T_n)}$ è detta efficienza di T_n

L'uguaglianza varrà sse le v.c. $T-\tau(\theta)$ e $\frac{d}{d\theta}\ln\prod_{i=1}^{n} f(X_i,\theta)$ sono parallele, quindi deve valere che:

$$\frac{d}{d\theta}\ln\prod_{i=1}^{n} f(X_i,\theta) = K(n,\theta)\cdot\left(t(x_1,\dots,x_n)-\tau(\theta)\right)$$

Per il momento fissiamo $n=1$:

$$\frac{d}{d\theta}\ln f(x,\theta) = K(\theta)\left(t(x)-\tau(\theta)\right)$$

OSSERVAZ. 1) Supponiamo che $f(x,\theta) = a(\theta)\, b(x)\cdot e^{c(\theta)\, d(x)}$ allora

$$\frac{d}{d\theta}\ln f(x;\theta) = \frac{d}{d\theta}\ln\left(a(\theta)\, b(x)\cdot e^{c(\theta)\, d(x)}\right) =$$

$$= \frac{d}{d\theta}\ln a(\theta)\ln b(x)\cdot c(\theta)\cdot d(x) = \frac{a'(\theta)}{a(\theta)} + d(x)\cdot c'(\theta) =$$

$$= c'(\theta)\left(\frac{a'(\theta)}{a(\theta)\,c'(\theta)} + d(x)\right)$$

quindi possiamo fissare:

$$K(\theta) = c'(\theta) \qquad \tau(\theta) = -\frac{a'(\theta)}{c'(\theta)\,a(\theta)} \qquad t(x) = d(x)$$

OSSERVAZIONE 2) Supponiamo che $\exists K(\theta) : \frac{d}{d\theta} \ln f(x;\theta) = K(\theta)\left(t(x) - \tau(\theta)\right) =$

$$= \underbrace{K(\theta)\, t(x)}_{\text{sarà della forma } c(\theta)\, d(x)} - \underbrace{K(\theta)\, \tau(\theta)}_{\text{sarà della forma } h(\theta)\, \boxed{+\, g(x)}}$$

ne segue che

$$e^{d(x)c(\theta)} \underbrace{e^{h(\theta)}}_{>0,\ a(\theta)} \underbrace{e^{g(x)}}_{>0,\ b(x)} = f(x;\theta)$$

NOTA: questa è la forma della famiglia esponenziale

Adesso mettiamo insieme le due osservazioni.

Abbiamo dimostrato che se $f(x,\theta) = a(\theta)\, b(x)\, e^{c(\theta) d(x)}$ allora esiste uno stimatore efficiente per: $\tau(\theta) = -\frac{a'(\theta)}{a(\theta)\, c'(\theta)}$

Abbiamo anche dimostrato che se saturo Cramer-Rao allora la distribuzione deve appartenere alla famiglia esponenziale

Fissato $\tau(\theta) = -\frac{a'(\theta)}{a(\theta)\, c'(\theta)}$

$$f(x,\theta) = a(\theta)\, b(x)\, e^{c(\theta) d(x)}$$

$$f_{X_1 \dots X_n}(x_1, \dots, x_n; \theta) = \prod_{i=1}^{n} f(x_i, \theta) = (a(\theta))^n \prod_{i=1}^{n} b(x_i) \quad e^{c(\theta) \prod d(x_i)}$$

$$\ln\left(f_{X_1 \dots X_n}(x_1, \dots, x_n; \theta)\right) = n \ln a(\theta) + \sum_{i=1}^{n} \ln(b(x_i)) + c(\theta) \sum_{i=1}^{n} d(x_i)$$

$$\frac{d}{d\theta} \ln\left(f_{\underline{X}}(\underline{x}, \theta)\right) = n \frac{a'(\theta)}{a(\theta)} + c'(\theta) \sum_{i=1}^{n} d(x_i) = \underbrace{n\, c'(\theta)}_{K(n,\theta)} \cdot \underbrace{\frac{1}{n} \sum d(x_i)}_{t(x_1, \dots, x_n)} + \underbrace{\frac{a'(\theta)}{a(\theta) c'(\theta)}}_{-\tau(\theta)} \ ?$$

Quindi $T_n(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^{n} d(X_i)$ è stimatore efficiente di $\tau(\theta) = -\frac{a'(\theta)}{a(\theta)\, c'(\theta)}$

Adesso chiediamoci se T_n è non distorto per $\tau(\theta)$

$$E_\theta\left(\frac{1}{n}\sum_i d(X_i)\right) = \frac{1}{n}\int dx_1 \int dx_2 \cdots \int dx_n \sum_{i=1}^{n} d(x_i) \prod_{j=1}^{n} f(x_j, \theta) =$$

$$= \frac{1}{n}\int d^n x \left(\sum_{i=1}^{n} d(x_i)\right)\left(\prod_{j=1}^{n} b(x_j)\right)(a(\theta))^n e^{c(\theta)\sum_{i=1}^{n} d(x_i)} =$$

$$= \frac{1}{n}\int d^n x \, (a(\theta))^n \left(\prod_{j=1}^{n} b(x_j)\right) \frac{d}{d\theta}\left(e^{c(\theta)\sum d(x_i)}\right)\frac{1}{c'(\theta)} =$$

Osservazione:

$$(a(\theta))^n \prod b(x_i) \frac{d}{d\theta} e^{c(\theta)\sum d(x_i)} + n\, a'(\theta)(a(\theta))^{n-1}\prod b(x_i)\, e^{c(\theta)\sum d(x_i)} =$$

$$= \frac{d}{d\theta}\left((a(\theta))^n \prod b(x_i)\, e^{c(\theta)\sum d(x_i)}\right)$$

$$= \frac{1}{n c'(\theta)}\left(\int d^n x \; - n\, a'(\theta)(a(\theta))^{n-1}\prod b(x_i)\, e^{c(\theta)\sum d(x_i)} + \right.$$

$$\left. + \frac{d}{d\theta}\underbrace{\int d^n x \left((a(\theta))^n \prod b(x_i)\, e^{c(\theta)\sum d(x_i)}\right)}_{=1}\right) =$$

perché è f su tutto $\mathbb{R}$ della famiglia dell'esponenziale

$$= \frac{-1}{n c'(\theta)}\int d^n x \; n\, a'(\theta)(a(\theta))^{n-1}\prod b(x_i)\, e^{c(\theta)\sum d(x_i)} =$$

$$= -\frac{\cancel{n}\, a'(\theta)}{\cancel{n}\, c'(\theta)}\int d^n x \, (a(\theta))^{n-1}\prod b(x_i)\, e^{c(\theta)\sum d(x_i)} =$$

$$= -\frac{a'(\theta)}{c'(\theta)\,a(\theta)}\int d^n x \, (a(\theta))^{n}\prod b(x_i)\, e^{c(\theta)\sum d(x_i)} = -\frac{a'(\theta)}{c'(\theta)\,a(\theta)}$$

T_n è non distorto rispetto a $\tau(\theta)$

T_n è detto
Uniform
Minimum
Var
Unbiased
Est

per $\tau(\theta)$

ESEMPIO 1

$X_1, \dots, X_n \qquad X \sim Exp(\theta)$

$$f(x,\theta) = \underbrace{I_{(0,+\infty)}(x)}_{b(x)} \cdot \underbrace{\theta}_{a(\theta)} \cdot e^{\overbrace{-\theta}^{c(\theta)} \overbrace{x}^{d(x)}}$$

$$T(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$\tau(\theta) = - \frac{a'(\theta)}{a(\theta) c'(\theta)} = - \frac{1}{\theta(-1)} = \frac{1}{\theta}$$

ESEMPIO 2

$X_1, \dots, X_n \qquad X \sim N(\theta, \sigma)$

$$f(x,\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2} = \frac{e^{-\frac{1}{2\sigma^2}(x^2+\theta^2-2x\theta)}}{\sigma\sqrt{2\pi}} =$$

$$= \underbrace{\frac{e^{-\frac{x^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}}_{b(x)} \cdot \underbrace{e^{-\frac{\theta^2}{2\sigma^2}}}_{a(\theta)} \cdot e^{\overbrace{\frac{\theta}{\sigma^2}}^{c(\theta)} \cdot \overbrace{x}^{d(x)}}$$

$$T(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$\tau(\theta) = - \frac{a'(\theta)}{a(\theta) c'(\theta)} = - \frac{\cancel{e^{-\frac{\theta^2}{2\sigma^2}}}\left(-\frac{2\theta}{2\sigma^2}\right)}{\cancel{e^{-\frac{\theta^2}{2\sigma^2}}} \cdot \frac{1}{\sigma^2}} = \frac{\theta\sigma^2}{\sigma^2} = \theta$$

ESEMPIO 3

Distribuzione uniforme $U(0,a)$ con $a>0$

$$f(x,a) = \frac{1}{a} \underbrace{I_{(0,a)}(x)}$$

dipende da 2 parametri simultaneamente, non posso fattorizzare

la distribuzione uniforme non appartiene alla famiglia esponenziale

Dim:

$A_1, A_2, \dots, A_n, \dots$ sia una famiglia numerabile di elementi misurabili tale che:

$$A = \bigcup_{n=1}^{\infty} A_n$$

Poniamo $A'_n = A_n / \bigcup_{k=1}^{n-1} A_k$ è chiaro che gli A'_n sono a due a due disgiunti.

Poiché ogni A'_i è misurabile varrà, per qualsiasi n, che:

$$\sum_{k=1}^{n} \mu(A'_i) = \mu\left(\bigcup_{i=1}^{n} A'_i\right) \leq \mu^*(A)$$ per la monotonia delle misure

Pertanto la serie $\sum_{n=1}^{+\infty} \mu(A'_n)$ converge. perché sono a due a due disgiunti

Di conseguenza per ogni $\varepsilon > 0$ esiste un N tale che

$$\sum_{i=N}^{+\infty} \mu(A'_i) < \frac{\varepsilon}{2}$$

Definiamo

$$C = \bigcup_{i=1}^{N} A'_i$$

Th Kolmogorov 7

C è misurabile perché unione finita di elementi misurabili quindi esisterà un B tale che:

$$\mu^*(C \triangle B) < \frac{\varepsilon}{2}$$

Possiamo dire che:

$$A \triangle B \subseteq (C \triangle B) \cup \left(\bigcup_{i=N}^{\infty} A'_i\right)$$

perché $A = C \cup \left(\bigcup_{i=N}^{\infty} A'_i\right)$, allora risulta che

$\mu^*(A \triangle B) < \varepsilon$ cioè A è misurabile

Essendo misurabili i complementi degli insiemi misurabili, possiamo scrivere l'intersezione:

$$\bigcap_n A_n = E \setminus \bigcup_n (E \setminus A_n)$$

poiché l'unione è misurabile sarà misurabile

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

CVD

Th: $\mu(A) = \mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mu(A)$ $\bigcup A_j$ numerabile

Dim:

$$\mu\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} \mu(A_n) \leq \mu(A)$$

Per $N \to +\infty$

$$\sum_{n=1}^{+\infty} \mu(A_n) \leq \mu(A)$$

?

Però $A \subseteq \bigcup_n^{\infty} A_n$

$$\mu(A) = \mu^*(A) \leq \mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

Allora

$$\mu(A) = \mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

CVD