Probabilità e Processi Stocastici (455AA)

Lezione 9

Dario Trevisan

21/10/2024

Section 1

Covarianza

Richiami

- *Varianza* $\mathrm{Var}\left(X\right) = \mathbb{E}\left[(X - \mathbb{E}\left[X\right])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2$

$$\boxed{P(|X - \mathbb{E}\left[X\right]| \leq k\sigma_X) \geq 1 - \frac{1}{k^2}}$$

Richiami

- *Varianza* $\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$
- *Deviazione standard* $\sigma_X = \sqrt{\mathrm{Var}(X)}$

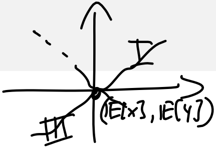
$$P(|X - \mathbb{E}[X]| \leq k\sigma_X) \geq 1 - \frac{1}{k^2}$$

Richiami

- *Varianza* $\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$
- *Deviazione standard* $\sigma_X = \sqrt{\mathrm{Var}(X)}$
- Diseguaglianza di Chebychev

$$P(|X - \mathbb{E}[X]| \leq k\sigma_X) \geq 1 - \frac{1}{k^2}$$

Richiami



- *Varianza* $\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 \geq 0$
- *Deviazione standard* $\sigma_X = \sqrt{\mathrm{Var}(X)}$
- Diseguaglianza di Chebychev

$$P(|X - \mathbb{E}[X]| \leq k\sigma_X) \geq 1 - \frac{1}{k^2}$$

- *Covarianza*
  $\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \in \mathbb{R}$
- Esempio $\theta \in [0, 2\pi]$ $X = \cos\theta$ $Y = \sin\theta$

Matrice delle covarianze

Dato un vettore aleatorio $X = (X_1, \ldots, X_d) \in \mathbb{R}^d$, si definisce la matrice delle covarianze di X la matrice di numeri reali $\Sigma_X \in \mathbb{R}^{d \times d}$

$$(\Sigma_X)_{i,j} = \mathrm{Cov}(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\,\mathbb{E}[X_j] \quad \text{per } i, j \in \{1, \ldots, d\}.$$

- Notazioni alternative: $\mathrm{Var}(X)$, K_{XX} o Q_X.

Proprietà

- La matrice delle covarianze è simmetrica $\Sigma_X = \Sigma_X^T$, dove T indica l'operazione di trasposizione.

$$\begin{pmatrix} \ddots & & (\Sigma_X)_{j,i} \\ & \ddots & \\ (\Sigma_X)_{i,j} & & \ddots \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 & 4 \\ 2 & 3 & 5 \\ 4 & 5 & -6 \end{pmatrix}$$

Proprietà

- La matrice delle covarianze è simmetrica $\Sigma_X = \Sigma_X^T$, dove T indica l'operazione di trasposizione.

- (trasformazioni affini) Sia $X \in \mathbb{R}^d$ e

$$Y = AX + b \quad \text{ossia} \quad \boxed{Y_i = \sum_{j=1}^{d} A_{ij} X_j + b_i,}$$

dove $A \in \mathbb{R}^{k \times d}$ è una matrice e $b \in \mathbb{R}^k$ è un vettore (costanti). Vale

$$\mathbb{R}^{k\times k} \ni \Sigma_{AX+b} = A \Sigma_X A^T. \qquad (A\Sigma_X A^T)_{ii'}$$

$$\mathrm{Cov}(Y_i, Y_{i'}) = \mathrm{Cov}\Big(\sum_j A_{ij}X_j + \cancel{b_i}, \sum_{j'} A_{i'j'}X_{j'} + \cancel{b_{i'}}\Big) =$$

$$\| $$

$$= \sum_{j'} A_{i'j'} \,\mathrm{Cov}\Big(\sum_j A_{ij}X_j + \cancel{b_i}, X_{j'}\Big) = \sum_{j,j'} A_{ij} (\Sigma_X)_{jj'} A_{i'j'}$$

Proprietà

- La matrice delle covarianze è simmetrica $\Sigma_X = \Sigma_X^T$, dove T indica l'operazione di trasposizione.
- (trasformazioni affini) Sia $X \in \mathbb{R}^d$ e

$$Y = AX + b \quad \text{ossia} \quad Y_i = \sum_{j=1}^{d} A_{ij} X_j + b_i,$$

dove $A \in \mathbb{R}^{k \times d}$ è una matrice e $b \in \mathbb{R}^k$ è un vettore (costanti). Vale

$$\Sigma_{AX+b} = A \Sigma_X A^T.$$

- In particolare, se $k = 1$ e $A = v^T$, con $v \in \mathbb{R}^d$, si ottiene che

$$\mathrm{Var}\left(v \cdot X\right) = \Sigma_{v \cdot X} = v^T \Sigma_X v, \geq 0$$

ossia Σ_X è (semi-)definita positiva.

Dimostrazione

Altra dimostrazione della formula $\Sigma_{AX} = A \Sigma_X A^T$

$(X - \mathbb{E}[X]) \in \mathbb{R}^d$ $\qquad \Sigma_X = \mathbb{E}\Big[\underbrace{(X-\mathbb{E}[X])(X-\mathbb{E}[X])^T}_{d \times d} \Big]$

$Y = AX \qquad \mathbb{E}[Y] = \mathbb{E}[AX] = A\mathbb{E}[X]$

$$\Sigma_Y = \mathbb{E}\big[(Y - \mathbb{E}[Y])(Y - \mathbb{E}[Y])^T \big]$$

$$= \mathbb{E}\big[A(X - \mathbb{E}[X]) \cdot (X - \mathbb{E}[X])^T A^T \big]$$

$$= A \Sigma_X A^T$$

Conseguenze: il coefficiente di correlazione

- Nel caso $d = 2$, scrivendo (X, Y) per la variabile congiunta di due variabili reali X, Y, la matrice delle covarianze è esplicitamente

$$\Sigma_{(X,Y)} = \begin{pmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X, Y) \\ \mathrm{Cov}(X, Y) & \mathrm{Var}(Y) \end{pmatrix}.$$

Conseguenze: il coefficiente di correlazione

- Nel caso $d = 2$, scrivendo (X, Y) per la variabile congiunta di due variabili reali X, Y, la matrice delle covarianze è esplicitamente

$$\Sigma_{(X,Y)} = \begin{pmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X, Y) \\ \mathrm{Cov}(X, Y) & \mathrm{Var}(Y) \end{pmatrix}.$$

- Essendo semidefinita positiva, il suo determinante è positivo (o nullo):

$$\det(\Sigma_{(X,Y)}) = \mathrm{Var}(X)\,\mathrm{Var}(Y) - (\mathrm{Cov}(X, Y))^2 \geq 0,$$

$$\left| \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} \right| \leq \frac{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}{\sigma_X \sigma_Y} = 1$$

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} \in [-1, 1]$$

Conseguenze: il coefficiente di correlazione

- Nel caso $d = 2$, scrivendo (X, Y) per la variabile congiunta di due variabili reali X, Y, la matrice delle covarianze è esplicitamente

$$\Sigma_{(X,Y)} = \begin{pmatrix} \mathrm{Var}\,(X) & \mathrm{Cov}\,(X, Y) \\ \mathrm{Cov}\,(X, Y) & \mathrm{Var}\,(Y) \end{pmatrix}.$$

- Essendo semidefinita positiva, il suo determinante è positivo (o nullo):

$$\det(\Sigma_{(X,Y)}) = \mathrm{Var}\,(X)\,\mathrm{Var}\,(Y) - (\mathrm{Cov}\,(X, Y))^2 \geq 0,$$

- ossia, dopo alcune operazioni elementari

$$\rho_{XY} := \frac{\mathrm{Cov}\,(X, Y)}{\sigma_X \sigma_Y} \in [-1, 1].$$

Conseguenze: il coefficiente di correlazione

- Nel caso $d = 2$, scrivendo (X, Y) per la variabile congiunta di due variabili reali X, Y, la matrice delle covarianze è esplicitamente

$$\Sigma_{(X,Y)} = \begin{pmatrix} \mathrm{Var}\,(X) & \mathrm{Cov}\,(X, Y) \\ \mathrm{Cov}\,(X, Y) & \mathrm{Var}\,(Y) \end{pmatrix}.$$

- Essendo semidefinita positiva, il suo determinante è positivo (o nullo):

$$\det(\Sigma_{(X,Y)}) = \mathrm{Var}\,(X)\,\mathrm{Var}\,(Y) - (\mathrm{Cov}\,(X, Y))^2 \geq 0,$$

- ossia, dopo alcune operazioni elementari

$$\rho_{XY} := \frac{\mathrm{Cov}\,(X, Y)}{\sigma_X \sigma_Y} \in [-1, 1].$$

$\rho \in \{-1, 1\}$ se e solo se esiste λ, a $Y = \lambda X + a$

- ρ_{XY} è il *coefficiente di correlazione* (o indice di correlazione di Pearson).

Standardizzazione nel caso vettoriale

Il teorema spettrale permette di decomporre

$$\Sigma_X = U^T D U,$$

dove $U \in \mathbb{R}^{d\times d}$, è ortogonale $U^T U = Id$ e D è diagonale (gli autovalori di Σ_X).

- La trasformazione $\overset{Y}{UX}$, corrisponde ad un cambio di coordinate e trasforma la covarianza

$$(\cancel{UU^T})D(\cancel{UU^T})$$

$$\Sigma_{UX} = U\Sigma_X U^T = D$$

ossia le componenti di UX sono a due a due non correlate.

- Se D è invertibile si può definire una *standardizzazione* di X

$$\hat{X} = \sqrt{D}^{-1} U(X - \mathbb{E}[X])$$

dove $\sqrt{D}$ è la matrice diagonale con entrate date dalla radice quadrata di quelle di D.

- Se D è invertibile si può definire una *standardizzazione* di X

$$\hat{X} = \sqrt{D}^{-1} U(X - \mathbb{E}[X])$$

dove $\sqrt{D}$ è la matrice diagonale con entrate date dalla radice quadrata di quelle di D.

- Usando le proprietà del vettore delle medie e della varianza, si ha

$$\mathbb{E}\left[\hat{X}\right] = 0 \in \mathbb{R}^d \quad \text{e} \quad \Sigma_{\hat{X}} = Id.$$

(supponendo $D > 0$)

Problema

Un'urna contiene N palline di cui R rosse e $B = N - R$ blu. Si effettuano $n \leq N$ estrazioni senza rimpiazzo e si pone $X_i \in \{0, 1\}$ la variabile indicatrice dell'evento "l'esito della estrazione i è una pallina rossa".

1. Calcolare il coefficiente di correlazione tra X_1 e X_2. Sono indipendenti?
2. Calcolare la matrice delle covarianze del vettore $(X_i)_{i=1}^n$.
3. Calcolare valor medio e varianza del numero totale Y di palline rosse estratte nelle n estrazioni.
4. La variabile Y è positivamente, negativamente o non correlata con X_1?

$$\mathrm{Cov}(Y, X_1) = \mathrm{Cov}\left(\sum_{i=1}^n X_i, X_1\right) = \sum_{i=1}^n \mathrm{Cov}(X_i, X_1) = \frac{R}{N}\frac{B}{N} - \frac{(n-1)}{N-1}\frac{R}{N}\frac{B}{N} = \frac{RB}{NN}\left(1 - \frac{n-1}{N-1}\right)$$

- X_1, X_2 variabili Bernoulli

<u>Negativamente correlate</u>

$$Var(X_1) = P(R_1)(1-P(R_1)) = \underline{\frac{R}{N} \cdot \frac{B}{N}}$$

$$Var(X_2) = Var(X_1) = \underline{\frac{R}{N} \cdot \frac{B}{N}}$$

$$Cov(X_1, X_2) = P(R_1 e R_2) - P(R_1)P(R_2) = \frac{R}{N} \cdot \frac{R-1}{N-1} - \frac{R}{N} \cdot \frac{R}{N} = \frac{R}{N} \frac{(R-1)\cdot N - (N-1)R}{N(N-1)} = \frac{R}{N} \frac{R-N}{N(N-1)} = \underline{-\frac{R}{N} \cdot \frac{B}{N(N-1)}}$$

$$\rho_{X_1 X_2} = - \frac{\cancel{\frac{R}{N}} \cdot \cancel{\frac{B}{N}} \cdot \frac{1}{N-1}}{\cancel{\frac{R}{N}} \cdot \cancel{\frac{B}{N}}} = -\frac{1}{N-1} < 0$$

$(N \geq 2)$ $\quad \rho \to 0$ se $N \to \infty$

- <u>OSS</u> $P(X_i = 1 \mid X_j = 1) = \frac{R-1}{N-1}$ $\quad \underline{\underline{\forall i \neq j}}$

$$P(X_j = 1) = \frac{R}{N}$$

$$Y = X_1 + X_2 + \dots + X_n \qquad \underline{\underline{n \leq N}}$$

$$\mathbb{E}[Y] = \mathbb{E}[X_1] + \dots + \mathbb{E}[X_n]$$

$$= n \cdot \frac{R}{N}$$

$$\mathrm{Var}(Y) < n \frac{R}{N}\left(1 - \frac{R}{N}\right) \leftarrow \text{caso con rimpiazzo}$$

$$\parallel$$

$$\mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right) = \sum_{i,j=1}^{n} \mathrm{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^{n} \frac{R}{N} \cdot \frac{B}{N} - \sum_{i \neq j} \frac{R}{N} \cdot \frac{B}{N} \cdot \frac{1}{N-1}$$

$$= n \frac{R}{N}\left(1 - \frac{R}{N}\right) - (n^2 - n) \frac{R}{N} \frac{B}{N} \frac{1}{N-1}$$

Problema

$$\Lambda \longrightarrow T$$

La durata della batteria di un drone è rappresentata tramite una variabile aleatoria T avente densità esponenziale di un parametro Λ, non del tutto noto e quindi rappresentato tramite una variabile aleatoria, anch'essa esponenziale, di parametro 1.

1. Dire se le variabili (Λ, T) sono tra loro indipendenti. No
2. Calcolare il coefficiente di correlazione tra Λ e T.
3. Si osserva che $T \geq 1$. Come cambia la densità di T? e di Λ? e il coefficiente di correlazione?

$\Rightarrow$ usare la densità $p(\Lambda=\lambda, T=t \mid T\geq 1) \propto \begin{cases} 0 & \text{se } t<1 \\ e^{-\lambda}\lambda e^{-\lambda t} & \text{se } t\geq 1 \end{cases}$

- Dal testo segue che $p(T=t \mid \Lambda=\lambda) = \lambda e^{-\lambda t}$ $t \geq 0$

T non è indipendente da λ dipende da λ

Se fossero indipendenti $p(T=t \mid \Lambda=\lambda) = p(T=t)$

Non dipende da λ

- $$\mathrm{Cov}(T, \Lambda) \overset{?}{=} \mathbb{E}[T\Lambda] - \mathbb{E}[T]\,\mathbb{E}[\Lambda]$$

$$\mathbb{E}[T\Lambda] = \int \mathbb{E}[T\Lambda \mid \Lambda=\lambda]\, p(\Lambda=\lambda)\, d\lambda = \int_0^{+\infty} \lambda\, \underbrace{\mathbb{E}[T \mid \Lambda=\lambda]}_{\frac{1}{\lambda}}\, e^{-\lambda}\, d\lambda$$

disintegriamo $\Lambda = \lambda$

$$= \int_0^{+\infty} \lambda \cdot \frac{1}{\lambda}\, e^{-\lambda}\, d\lambda = \int_0^{+\infty} e^{-\lambda}\, d\lambda = 1$$

$$\mathbb{E}[T] = \int \mathbb{E}[T \mid \Lambda = \lambda]\, p(\Lambda = \lambda)\, d\lambda = \int_0^{+\infty} \mathbb{E}[T \mid \Lambda = \lambda]\, e^{-\lambda}\, d\lambda$$

disintegriamo $\Lambda = \lambda$

$$= \int_0^{+\infty} \frac{1}{\lambda} e^{-\lambda} d\lambda = \int_0^{+\infty} \frac{e^{-\lambda}}{\lambda} d\lambda = +\infty$$

$\approx \frac{1}{\lambda}$ se $\lambda \to 0^+$

$\mathrm{Cov}(\Lambda, T) =$ "non esiste"

$= 1 - 1 \cdot \infty = -\infty$

Se $T \geq 1$ come cambiano le risposte?

dobbiamo calcolare $p(T=t, \Lambda=\lambda \mid T \geq 1) =$

$$= p(\Lambda = \lambda \mid T \geq 1) \cdot p(T = t \mid t \geq 1, \Lambda = \lambda)$$

$$= \frac{P(T \geq 1 \mid \Lambda = \lambda)\, p(\Lambda = \lambda)}{P(T \geq 1)} \cdot \frac{\lambda e^{-\lambda t}}{P(T \geq 1 \mid \Lambda = \lambda)} \quad \text{se } t \geq 1$$

Se X ha densità $p(X=x)$ e osserviamo $X \in U$

allora X condizionato a $X \in U$ ha densità

$$p(X=x) = \begin{cases} 0 & \text{se } x \notin U \\ p(X=x)/P(X \in U) & \text{se } x \in U \end{cases}$$

Section 2

Momenti

Approssimazione della legge tramite polinomi

Supponiamo di dover approssimare $\mathbb{E}[g(X)]$, dove $g : \mathbb{R} \to \mathbb{R}$ è una funzione regolare.

- Consideriamo uno sviluppo di Taylor per g

$$g(x) \sim a_0 + a_1 x + a_2 x^2 + \ldots + a_k x^k$$

dove $a_i \in \mathbb{R}$ sono costanti, e per la linearità del valor medio

$$\mathbb{E}[g(X)] \sim a_0 + a_1 \mathbb{E}[X] + a_2 \mathbb{E}\left[X^2\right] + \ldots + a_k \mathbb{E}\left[X^k\right].$$

Approssimazione della legge tramite polinomi

Supponiamo di dover approssimare $\mathbb{E}\left[g(X)\right]$, dove $g : \mathbb{R} \to \mathbb{R}$ è una funzione regolare.

- Consideriamo uno sviluppo di Taylor per g

$$g(x) \sim a_0 + a_1 x + a_2 x^2 + \ldots + a_k x^k$$

dove $a_i \in \mathbb{R}$ sono costanti, e per la linearità del valor medio

$$\mathbb{E}\left[g(X)\right] \sim a_0 + a_1 \mathbb{E}\left[X\right] + a_2 \mathbb{E}\left[X^2\right] + \ldots + a_k \mathbb{E}\left[X^k\right].$$

- Problemi:

Approssimazione della legge tramite polinomi

Supponiamo di dover approssimare $\mathbb{E}[g(X)]$, dove $g : \mathbb{R} \to \mathbb{R}$ è una funzione regolare.

- Consideriamo uno sviluppo di Taylor per g

$$g(x) \sim a_0 + a_1 x + a_2 x^2 + \ldots + a_k x^k$$

dove $a_i \in \mathbb{R}$ sono costanti, e per la linearità del valor medio

$$\mathbb{E}[g(X)] \sim a_0 + a_1 \mathbb{E}[X] + a_2 \mathbb{E}\left[X^2\right] + \ldots + a_k \mathbb{E}\left[X^k\right].$$

- Problemi:
  1. determinare un polinomio approssimante per g

Approssimazione della legge tramite polinomi

Supponiamo di dover approssimare $\mathbb{E}[g(X)]$, dove $g : \mathbb{R} \to \mathbb{R}$ è una funzione regolare.

- Consideriamo uno sviluppo di Taylor per g

$$g(x) \sim a_0 + a_1 x + a_2 x^2 + \ldots + a_k x^k$$

dove $a_i \in \mathbb{R}$ sono costanti, e per la linearità del valor medio

$$\mathbb{E}[g(X)] \sim a_0 + a_1 \mathbb{E}[X] + a_2 \mathbb{E}\left[X^2\right] + \ldots + a_k \mathbb{E}\left[X^k\right].$$

- Problemi:

1. determinare un polinomio approssimante per g
2. calcolare i valor medi $\mathbb{E}[X]$, $\mathbb{E}\left[X^2\right]$, $\ldots$, $\mathbb{E}\left[X^k\right]$

I momenti

Sia $X \in \mathbb{R}$ una variabile aleatoria. Per ogni $k \in \mathbb{N}$, si dice **momento di ordine** k (o momento k-esimo) di X la quantità

$$\mathbb{E}\left[X^k\right],$$

se è ben definita (ricordiamo che si richiede che la serie o l'integrale che definisce $\mathbb{E}\left[X^k\right]$ debba convergere).

- se X ha densità (discreta o continua) vale

$$\mathbb{E}\left[X^k\right] = \begin{cases} \sum_{x \in \mathbb{R}} x^k P(X = x) & \text{se } X \text{ ha densità discreta,} \\ \int_{x \in \mathbb{R}} x^k p(X = x) dx & \text{se } X \text{ ha densità continua.} \end{cases}$$

Momento secondo e varianza, terzo

- La varianza è scrivibile come combinazione del momento secondo e dal momento primo:

$$\mathbb{E}\left[X^2\right] = \mathrm{Var}\,(X) + (\mathbb{E}\,[X])^2,$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

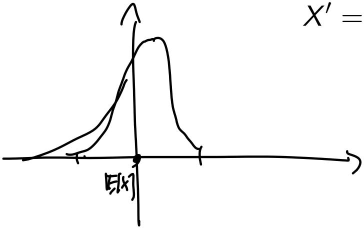
Momento secondo e varianza, terzo

- La varianza è scrivibile come combinazione del momento secondo e dal momento primo:

$$\mathbb{E}\left[X^2\right] = \mathrm{Var}\,(X) + (\mathbb{E}\,[X])^2,$$

- Per gli ordini successivi si considerano anche i momenti della variabile standardizzata

$$X' = (X - \mathbb{E}\,[X])/\sigma_X.$$



Momento secondo e varianza, terzo

- La varianza è scrivibile come combinazione del momento secondo e dal momento primo:

$$\mathbb{E}\left[X^2\right] = \mathrm{Var}\,(X) + (\mathbb{E}\,[X])^2,$$

- Per gli ordini successivi si considerano anche i momenti della variabile standardizzata

$$X' = (X - \mathbb{E}\,[X])/\sigma_X.$$

- il suo momento terzo è detto **skewness** di X (e indica eventuale asimmetria della densità rispetto alla media)

Momento secondo e varianza, terzo

- La varianza è scrivibile come combinazione del momento secondo e dal momento primo:

$$\mathbb{E}\left[X^2\right] = \mathrm{Var}\,(X) + (\mathbb{E}\,[X])^2,$$

- Per gli ordini successivi si considerano anche i momenti della variabile standardizzata

$$X' = (X - \mathbb{E}\,[X])/\sigma_X.$$

- il suo momento terzo è detto **skewness** di X (e indica eventuale asimmetria della densità rispetto alla media)
- il momento quarto è detto **kurtosi**.

La funzione generatrice dei momenti

Data $X \in \mathbb{R}$ una variabile aleatoria reale, si definisce la sua **funzione generatrice dei momenti** (in inglese *moment generating function*, **MGF**) la funzione $\mathrm{MGF}_X : \mathbb{R} \to [0, \infty]$, che associa

$$t \mapsto \mathrm{MGF}_X(t) = \mathbb{E}\left[e^{tX}\right].$$

- Per ciascun $t \in \mathbb{R}$, il valor medio si calcola quindi come

$$\mathrm{MGF}_X(t) = \begin{cases} \sum_{x \in \mathbb{R}} e^{tx} P(X = x) & \text{se } X \text{ ha densità discreta,} \\ \int_{x \in \mathbb{R}} e^{tx} p(X = x) dx & \text{se } X \text{ ha densità continua.} \end{cases}$$

La funzione generatrice dei momenti

Data $X \in \mathbb{R}$ una variabile aleatoria reale, si definisce la sua **funzione generatrice dei momenti** (in inglese *moment generating function*, **MGF**) la funzione $\mathrm{MGF}_X : \mathbb{R} \to [0, \infty]$, che associa

$$t \mapsto \mathrm{MGF}_X(t) = \mathbb{E}\left[e^{tX}\right].$$

- Per ciascun $t \in \mathbb{R}$, il valor medio si calcola quindi come

$$\mathrm{MGF}_X(t) = \begin{cases} \sum_{x \in \mathbb{R}} e^{tx} P(X = x) & \text{se } X \text{ ha densità discreta,} \\ \boxed{\int_{x \in \mathbb{R}} e^{tx} p(X = x) dx} & \text{se } X \text{ ha densità continua.} \end{cases}$$

- Se per qualche $t \in \mathbb{R}$ l'integrale o la serie che definiscono il valor medio di $\mathbb{E}\left[e^{tX}\right]$ non convergono, si pone $\mathrm{MGF}_X(t) = \infty$.

· $f(x)$ trasformata di Laplace $\boxed{\int_{-\infty}^{+\infty} e^{tx} f(x) dx}$

La funzione generatrice dei momenti

Data $X \in \mathbb{R}$ una variabile aleatoria reale, si definisce la sua **funzione generatrice dei momenti** (in inglese *moment generating function*, **MGF**) la funzione $\mathrm{MGF}_X : \mathbb{R} \to [0, \infty]$, che associa

$$t \mapsto \mathrm{MGF}_X(t) = \mathbb{E}\left[e^{tX}\right].$$

- Per ciascun $t \in \mathbb{R}$, il valor medio si calcola quindi come

$$\mathrm{MGF}_X(t) = \begin{cases} \sum_{x\in\mathbb{R}} e^{tx} P(X = x) & \text{se } X \text{ ha densità discreta,} \\ \int_{x\in\mathbb{R}} e^{tx} p(X = x) dx & \text{se } X \text{ ha densità continua.} \end{cases}$$

- Se per qualche $t \in \mathbb{R}$ l'integrale o la serie che definiscono il valor medio di $\mathbb{E}\left[e^{tX}\right]$ non convergono, si pone $\mathrm{MGF}_X(t) = \infty$.
- Per $t = 0$,

$$\mathrm{MGF}_X(0) = \mathbb{E}\left[e^{0\cdot X}\right] = \mathbb{E}[1] = 1.$$

Esempi

$$p(X=x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$MGF_X(t) = \int_{-\infty}^{+\infty} e^{tx} p(X=x)\, dx = \int_0^{+\infty} \underline{e^{tx}} \lambda \underline{e^{-\lambda x}} dx$$

Sia $X \in \mathbb{R}$ con densità continua esponenziale di parametro $\lambda > 0$. Allora

$$-\frac{\lambda}{(t-\lambda)} \int_0^{+\infty} -(t-\lambda) e^{x(t-\lambda)} dx = \begin{cases} +\infty & \text{se } t \geq \lambda \\ \frac{\lambda}{\lambda - t} & \text{se } t < \lambda \end{cases}$$

Esempi discreti

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

$$MGF(t) = \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (e^t p)^k ((1-p))^{n-k}$$

$$\begin{cases} x = e^t p \\ y = 1-p \\ \sum \binom{n}{k} x^k y^{n-k} = (x+y)^n \end{cases}$$

$$= \left(e^t p + (1-p)\right)^n$$

Se $n=1$ $\quad \text{Binomiale}(1,p) = \text{Bernoulli}(p) \Rightarrow MGF(t) = 1-p+e^t p$

Proprietà della MGF

Siano X, $Y \in \mathbb{R}$ variabili aleatorie e a, $b \in \mathbb{R}$ costanti. Allora

1. $\mathrm{MGF}_{aX+b}(t) = e^{tb}\,\mathrm{MGF}_X(at)$

$$\mathbb{E}\left[\exp((aX+b)t)\right] = \mathbb{E}\left[\exp(bt)\exp((at)X)\right]$$

Proprietà della MGF

Siano X, $Y \in \mathbb{R}$ variabili aleatorie e a, $b \in \mathbb{R}$ costanti. Allora

1. $\mathrm{MGF}_{aX+b}(t) = e^{tb}\, \mathrm{MGF}_X(at)$
2. Se X, Y sono indipendenti, allora $\mathrm{MGF}_{X+Y}(t) = \mathrm{MGF}_X(t)\, \mathrm{MGF}_Y(t)$.

$$\mathbb{E}[\exp((X+Y)t)] = \mathbb{E}[\underbrace{\exp(Xt)}_{h(X)}\ \underbrace{\exp(Yt)}_{g(Y)}] = \mathbb{E}[\exp(Xt)] \cdot \mathbb{E}[\exp(Yt)]$$

Dimostrazione

Dalla MGF ai momenti

Sia $X \in \mathbb{R}$ tale che $\mathrm{MGF}_X(t) < \infty$ per ogni $t \in (-\varepsilon, \varepsilon)$, per qualche $\varepsilon > 0$. Allora, per ogni $k \in \mathbb{N}$, X ha momento di ordine k ben definito e vale

$$\boxed{\frac{d^k}{d^k t} \mathrm{MGF}_X(0) = \mathbb{E}\left[X^k\right].}$$

- Per calcolare il momento di ordine k è quindi sufficiente derivare k volte la $\mathrm{MGF}_X(t)$ e successivamente porre $t = 0$.

Dimostrazione (cenno)

$$\mathbb{E}[e^{tX}] = \mathbb{E}\left[\sum_{k=0}^{+\infty} \frac{(tX)^k}{k!}\right] = \sum_{k=0}^{+\infty} \frac{\mathbb{E}[(tX)^k]}{k!}$$

$$\boxed{MGF_X(t) = \sum_{k=0}^{+\infty} \frac{t^k}{k!} \mathbb{E}[X^k]}$$

Coeff. dello svil. prod. Taylor in $t=0$

$$\frac{d}{dt}\mathbb{E}[e^{tX}]\Big|_{t=0} = \lim_{h\to 0} \frac{\mathbb{E}[e^{(t+h)X}] - \mathbb{E}[e^{tX}]}{h}$$

$$= \lim_{h\to 0} \mathbb{E}\left[\frac{e^{(t+h)X} - e^{tX}}{h}\right] \text{“}=\text{”} \mathbb{E}\left[\frac{d}{dt} e^{tX}\right]\Big|_{t=0} = \mathbb{E}[X]$$

Il caso vettoriale

Si può estendere il concetto di momento a variabili vettoriali $X \in \mathbb{R}^d$, considerando prodotti delle marginali.

- Il momento primo corrisponde al vettore dei valor medi,

$$\left(\mathbb{E}[X_i] \right)_{i=1 \ldots d}$$

Il caso vettoriale

Si può estendere il concetto di momento a variabili vettoriali $X \in \mathbb{R}^d$, considerando prodotti delle marginali.

- Il momento primo corrisponde al vettore dei valor medi,
- il vettore secondo alla collezione

$$\mathbb{E}[X_i X_j] \quad \text{per } i, j \in \{1, \ldots, d\} \text{ (anche } i = j\text{)}$$

Il caso vettoriale

Si può estendere il concetto di momento a variabili vettoriali $X \in \mathbb{R}^d$, considerando prodotti delle marginali.

- Il momento primo corrisponde al vettore dei valor medi,
- il vettore secondo alla collezione

$$\mathbb{E}[X_i X_j] \quad \text{per } i, j \in \{1, \ldots, d\} \text{ (anche } i = j)$$

- il momento terzo

$$\mathbb{E}[X_i X_j X_{\cancel{i}k}] \quad \text{per } i, j, k \in \{1, \ldots, d\}.$$

Il caso vettoriale

Si può estendere il concetto di momento a variabili vettoriali $X \in \mathbb{R}^d$, considerando prodotti delle marginali.

- Il momento primo corrisponde al vettore dei valor medi,
- il vettore secondo alla collezione

$$\mathbb{E}[X_i X_j] \quad \text{per } i, j \in \{1, \ldots, d\} \text{ (anche } i = j)$$

- il momento terzo

$$\mathbb{E}[X_i X_j X_j] \quad \text{per } i, j, k \in \{1, \ldots, d\}.$$

- La MGF di X vettoriale è funzione di $t = (t_1, t_2, \ldots, t_d) \in \mathbb{R}^d$,

$$\mathrm{MGF}_X(t) = \mathbb{E}\left[\exp\left(\sum_{i=1}^d t_i X_i\right)\right] = \mathbb{E}[\exp(t \cdot X)]$$

- $$MGF_{AX+b}(t) = e^{t\cdot b}\, MGF_X(A^T t)$$

$$t\cdot(AX) = t^T A X = (A^T t)^T X$$

- $MGF_{X+Y}(t) = MGF_X(t)\, MGF_Y(t)$ se X, Y <u>indip</u>

- $$\mathbb{E}[X_1^{\alpha_1} X_2^{\alpha_2} \cdots X_d^{\alpha_d}] = \frac{\partial^{\alpha_1+\dots+\alpha_d}}{(\partial t_1)^{\alpha_1}(\partial t_2)^{\alpha_2}\cdots(\partial t_d)^{\alpha_d}} MGF(t)\Big|_{t=0}$$

- $$\mathbb{E}[X_1 X_2] = \frac{\partial^2}{\partial t_1 \partial t_2} MGF_X(t)\Big|_{t_1=t_2=0}$$

Section 3

Richiami sulla trasformata di Fourier

Caso finito

Fissano $n \in \mathbb{N}$, si consideri un *segnale* definito (o misurato) su un intervallo discreto di n valori

$$g : \{0, 1, \ldots, (n-1)\} \to \mathbb{C}, \quad t \mapsto g(t).$$

Si definisce la sua *trasformata di Fourier* come la funzione

$\frac{2\pi\xi}{n} = \omega$

$$\hat{g} : \{0, 1, \ldots, (n-1)\} \to \mathbb{C}, \quad \xi \mapsto \hat{g}(\xi) := \sum_{t=0}^{n-1} g(t) e^{-2\pi i t\xi/n}.$$

- Il dominio della g può essere pensato come un insieme di tempi

Caso finito

Fissano $n \in \mathbb{N}$, si consideri un *segnale* definito (o misurato) su un intervallo discreto di n valori

$$g : \{0, 1, \ldots, (n-1)\} \to \mathbb{C}, \quad t \mapsto g(t).$$

Si definisce la sua *trasformata di Fourier* come la funzione

$$\hat{g} : \{0, 1, \ldots, (n-1)\} \to \mathbb{C}, \quad \xi \mapsto \hat{g}(\xi) := \sum_{t=0}^{n-1} g(t) e^{-2\pi i t \xi / n}.$$

- Il dominio della g può essere pensato come un insieme di tempi
- il dominio di $\hat{g}$ come *frequenze* ξ (precisamente, le frequenze sarebbero ξ/n, mentre le frequenze angolari $2\pi\xi/n$)

Notazione matriciale

Interpetiamo $g = (g(t))_{t=0}^{n-1}$ e $(\hat{g}(\xi))_{\xi=0}^{n-1}$ come vettori in $\mathbb{C}^n$: allora $\hat{g} = Fg$, dove $F \in \mathbb{C}^{n \times n}$

$$\boxed{F_{\xi t} = e^{-2\pi i t\xi/n}.}$$

La matrice F è *unitaria*, ossia l'inversa di F è la sua trasposta coniugata (a meno di una costante $1/n$). Per s, $t \in \{0, 1, \ldots, n-1\}$,

$$(\bar{F}^T F)_{st} = \sum_{\xi=0}^{n-1} \underbrace{e^{2\pi i s\xi/n}}_{\bar{F}^T_{s\xi}} \underbrace{e^{-2\pi i t\xi/n}}_{F_{\xi t}} = \begin{cases} n & \text{se } s = t \\ 0 & \text{altrimenti.} \end{cases}$$

Per dimostrarlo basta ricordare la somma geometrica $\sum_{j=0}^{k-1} z^j = (z^k - 1)/(z-1)$ e il fatto che $e^{2\pi i} = 1$.

$$\boxed{z = e^{2\pi i (s-t)}}$$

Formula di inversione

Come prima conseguenza otteniamo la formula di inversione

$$g = \frac{1}{n} \bar{F}^t F g,$$

che esplicitamente diventa

$$\boxed{g(t) = \frac{1}{n} \sum_{\xi=0}^{n-1} \hat{g}(\xi) e^{2\pi i \xi t/n}.}$$

Identità dell'energia

$$\sum_{\xi=0}^{n-1} |\hat{g}(\xi)|^2 = n \sum_{s=0}^{n-1} |g(s)|^2$$

Una seconda conseguenza è il fatto che la norma (Euclidea) del vettore g coincide (a meno di un fattore $1/n$) con quella del vettore $\hat{g}$, perché

$$|\hat{g}|^2 = \bar{F}\bar{g}^T F g = \bar{g}^T \bar{F}^T F g = n\bar{g}^T g = n|g|^2.$$

- La norma $|g|^2$ può essere intepretata come una *energia* del segnale g, di conseguenza l'identità sopra mostra che la stessa energia può essere ottenuta sommando le energie associate alle singole frequenze, ossia $|\hat{g}(\xi)|^2$ (e dividendo per n).

Calcolo mediante R della trasformata di Fourier

- Tutte le trasformate di Fourier che si calcolano numericamente sono ridotte al caso di tempi finiti.

Calcolo mediante R della trasformata di Fourier

- Tutte le trasformate di Fourier che si calcolano numericamente sono ridotte al caso di tempi finiti.
- Per questo vi sono algoritmi particolarmente veloci, che in R si possono usare mediante la funzione `fft()`.

g_hat = fft(g)

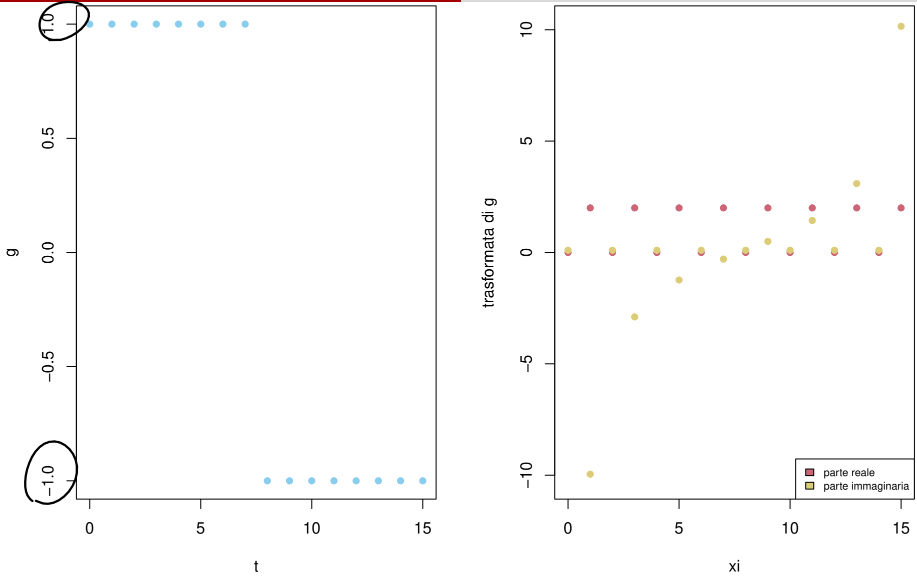
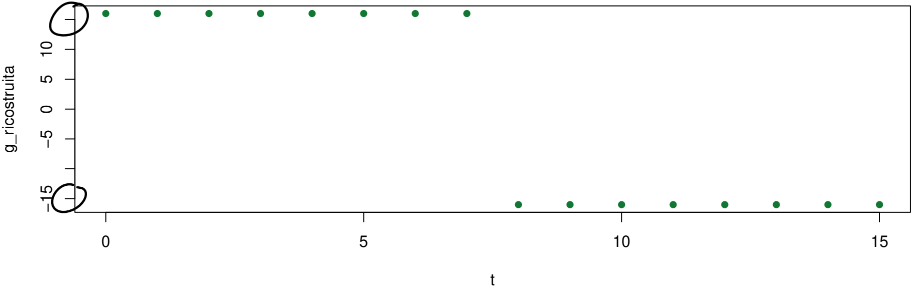


Figure 1: esempio di un segnale finito (onda quadra) e della sua trasformata di Fourier

Possiamo anche verificare la formula di inversione, usando l'opzione `inverse =TRUE` nella stessa funzione `fft()`. Bisogna tuttavia ricordare il fattore n.



Caso discreto

Supponiamo ora di osservare un segnale definito su un tempo infinito discreto $g : \mathbb{Z} \to \mathbb{C}$ (è una situazione ideale ovviamente).

- L'analoga trasformazione stavolta definisce la trasformata di Fourier a tempi discreti

$$\hat{g} : [0,1] \to \mathbb{C}, \quad \xi \mapsto \hat{g}(\xi) := \sum_{t \in \mathbb{Z}} g(t) e^{-2\pi i t \xi},$$

  purché la serie converga, ad esempio se

$$\boxed{\sum_{t \in \mathbb{Z}} |g(t)| < \infty.}$$

Caso discreto

Supponiamo ora di osservare un segnale definito su un tempo infinito discreto $g : \mathbb{Z} \to \mathbb{C}$ (è una situazione ideale ovviamente).

- L'analoga trasformazione stavolta definisce la trasformata di Fourier a tempi discreti

$$\hat{g} : [0,1] \to \mathbb{C}, \quad \xi \mapsto \hat{g}(\xi) := \sum_{t \in \mathbb{Z}} g(t) e^{2\pi i t \xi},$$

  purché la serie converga, ad esempio se

$$\sum_{t \in \mathbb{Z}} |g(t)| < \infty.$$

- Per passare dal finito al discreto l'idea è di cambiare la variabile frequenza nel caso discreto, ossia di passare da ξ a ξ/n, in modo che il dominio sia l'intervallo discreto $\{0, 1/n, 2/n, \ldots, (n-1)/n\}$. In questo modo, per $n \to \infty$ si ottiene una funzione definita sull'intervallo continuo di frequenze $[0,1]$.

Anche senza ricorrere al limite dal caso finito, si hanno la formula di inversione

$$g(t) = \int_0^1 \hat{g}(\xi) e^{2\pi i t \xi} d\xi,$$

e l'identità dell'energia

$$\sum_{t \in \mathbb{Z}} |g(t)|^2 = \int_0^1 |\hat{g}(\xi)|^2 d\xi.$$

- Il punto chiave è la relazione di ortogonalità

$$\boxed{\int_0^1 e^{2\pi i s \xi} e^{-2\pi i t \xi} d\xi = \begin{cases} 1 & \text{se } s = t, \\ 0 & \text{altrimenti,} \end{cases}}$$

che si dimostra ad esempio integrando per parti. Le due relazioni sopra seguono ripercorrendo la dimostrazione del caso finito sfruttando questa ortogonalità.

Operatore ritardo (lag) e trasformata

Definiamo l'operatore di *ritardo* L (in inglese *lag*), che trasforma g nel segnale

$$t \mapsto (Lg)(t) = g(t-1),$$

allora

$$\widehat{Lg}(\xi) = \sum_{t \in \mathbb{Z}} g(t-1) e^{-2\pi i t \xi} = e^{-2\pi i \xi} \hat{g}(\xi).$$

In termini fisici, la traslazione (o ritardo) fa acquisire una fase alla trasformata.

- Iterando l'operazione, la fase si accumula: posta $L^s g(t) = g(t-s)$, ossia L applicata s-volte a g, si ha

$$\widehat{L^s g}(\xi) = e^{-2\pi i s \xi} \hat{g}(\xi).$$

Convoluzione e trasformata

- Quando si intepreta g come un segnale, la *convoluzione* di g con un "filtro" $f : \mathbb{Z} \to \mathbb{C}$ è una operazione naturale:

$$(g * f)(t) = \sum_{s \in \mathbb{Z}} g(t-s) f(s).$$

Passando alla trasformata di Fourier, possiamo scrivere

$$
\begin{aligned}
\widehat{g * f}(\xi) &= \sum_{s\in\mathbb{Z}} \widehat{g(\cdot - s)f(s)}(\xi) \\
&= \sum_{s\in\mathbb{Z}} \widehat{g(\cdot - s)}(\xi) f(s) \\
&= \sum_{s\in\mathbb{Z}} \widehat{L^s g}(\xi) f(s) \\
&= \sum_{s\in\mathbb{Z}} e^{-2\pi i s} \hat{g}(\xi) f(s) \\
&= \hat{g}(\xi) \sum_{s\in\mathbb{Z}} e^{-2\pi i s} \hat{g}(\xi) f(s) = \hat{g}(\xi)\hat{f}(\xi).
\end{aligned}
$$

- Nelle frequenze la convoluzione con un f diventa il prodotto con $\hat{f}$.

Passando alla trasformata di Fourier, possiamo scrivere

$$\begin{aligned}
\widehat{g * f}(\xi) &= \sum_{s\in\mathbb{Z}} \widehat{g(\cdot - s) f(s)}(\xi) \\
&= \sum_{s\in\mathbb{Z}} \widehat{g(\cdot - s)}(\xi) f(s) \\
&= \sum_{s\in\mathbb{Z}} \widehat{L^s g}(\xi) f(s) \\
&= \sum_{s\in\mathbb{Z}} e^{-2\pi i s} \hat{g}(\xi) f(s) \\
&= \hat{g}(\xi) \sum_{s\in\mathbb{Z}} e^{-2\pi i s} \hat{g}(\xi) f(s) = \hat{g}(\xi) \hat{f}(\xi).
\end{aligned}$$

- Nelle frequenze la convoluzione con un f diventa il prodotto con $\hat{f}$.
- Identità dell'energia:

$$\sum_{t\in\mathbb{Z}} |g * f|^2(t) = \int_0^1 |\hat{g}|^2(\xi) |\hat{f}|^2(\xi) d\xi.$$

Caso continuo

Pasando da tempi da discreti a continui, si trova che per un segnale

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

la trasformata è definita come

$$\hat{g} : \mathbb{R} \to \mathbb{C}, \quad \hat{g}(\xi) := \int_{-\infty}^{\infty} g(t) e^{-2\pi i t \xi} dt,$$

- purché l'integrale converga, ad esempio se

$$\int_{-\infty}^{\infty} |g(t)| dt < \infty.$$

Caso continuo

Pasando da tempi da discreti a continui, si trova che per un segnale

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

la trasformata è definita come

$$\hat{g} : \mathbb{R} \to \mathbb{C}, \quad \hat{g}(\xi) := \int_{-\infty}^{\infty} g(t) e^{-2\pi i t \xi} dt,$$

- purché l'integrale converga, ad esempio se

$$\int_{-\infty}^{\infty} |g(t)| dt < \infty.$$

- Si può anche usare come variabile della trasformata la frequenza angolare $\omega = 2\pi\xi$

- Formula di inversione

$$\boxed{g(t) = \int_{-\infty}^{\infty} \hat{g}(\xi) e^{2\pi i t \xi} d\xi,}$$

(purché l'integrale abbia senso)

- Formula di inversione

$$g(t) = \int_{-\infty}^{\infty} \hat{g}(\xi) e^{2\pi i t \xi} d\xi,$$

(purché l'integrale abbia senso)

- Identità dell'energia

$$\boxed{\int_{-\infty}^{\infty} |g(t)|^2 dt = \int_{-\infty}^{\infty} |\hat{g}(\xi)|^2 d\xi.}$$

- Convoluzione con $f : \mathbb{R} \to \mathbb{C}$,

$$\boxed{g * f(t) = \int_{-\infty}^{\infty} g(t-s) f(s) ds,}$$

e nella base delle frequenze l'operazione si riduce ad un prodotto:

$$\boxed{\widehat{g * f}(\xi) = \hat{g}(\xi) \hat{f}(\xi).}$$

OSS Se X e Y sono continue e indipendenti $X+Y=Z$ ha densità data dalla convoluzione delle p_X, p_Y

$$p(X+Y=z) = \int_{\mathbb{R}} p(Y=z-x \mid \cancel{X=x}) \, p(X=x) \, dx$$

eliminato per indip.

Entropia di una densità uniforme

- caso discreto

$$H(X \text{ uniforme su } n \text{ valori}) = -\sum_{i=1}^{n} \log\left(\frac{1}{n}\right)\frac{1}{n} = \log(n),$$

Entropia di una densità uniforme

- caso discreto

$$H(X \text{ uniforme su } n \text{ valori}) = -\sum_{i=1}^{n} \log\left(\frac{1}{n}\right)\frac{1}{n} = \log(n),$$

- caso continuo

$$\begin{aligned} H(X \text{ uniforme continua su } [a,b]) &= -\int_a^b \log\left(\frac{1}{b-a}\right)\frac{1}{b-a}dx \\ &= \log(b-a). \end{aligned}$$

Entropia di una densità uniforme

- caso discreto

$$H(X \text{ uniforme su } n \text{ valori}) = -\sum_{i=1}^{n} \log\left(\frac{1}{n}\right)\frac{1}{n} = \log(n),$$

- caso continuo

$$\begin{aligned} H(X \text{ uniforme continua su } [a,b]) &= -\int_a^b \log\left(\frac{1}{b-a}\right)\frac{1}{b-a}dx \\ &= \log(b-a). \end{aligned}$$

- In particolare, più grande è tale insieme, maggiore è l'entropia (c'è meno informazione).

Il principio di massima entropia

L'entropia ha un ruolo importante nel determinare densità (discrete o continue) per variabili aleatorie X.

- Si estende il principio di Laplace (per la probabilità uniforme) al **principio di massima entropia**.

Il principio di massima entropia

L'entropia ha un ruolo importante nel determinare densità (discrete o continue) per variabili aleatorie X.

- Si estende il principio di Laplace (per la probabilità uniforme) al **principio di massima entropia**.
- Qualora l'informazione disponibile permetta solo di indentificare una classe $\mathcal{D}$ di densità ammissibili, allora *si sceglierà la densità per cui $H(X)$ sia massima tra quelle in $\mathcal{D}$*.

Il principio di massima entropia

L'entropia ha un ruolo importante nel determinare densità (discrete o continue) per variabili aleatorie X.

- Si estende il principio di Laplace (per la probabilità uniforme) al **principio di massima entropia**.
- Qualora l'informazione disponibile permetta solo di indentificare una classe $\mathcal{D}$ di densità ammissibili, allora *si sceglierà la densità per cui* $H(X)$ *sia massima tra quelle in* $\mathcal{D}$.
- Molte densità notevoli sono di **massima entropia** in una opportuna classe, che ne giustifica l'uso nella pratica.

Esempi

- La densità uniforme (discreta) su un insieme E con n elementi massimizza l'entropia tra tutte le densità discrete su E.

Esempi

- La densità uniforme (discreta) su un insieme E con n elementi massimizza l'entropia tra tutte le densità discrete su E.
- La densità uniforme continua su $E = [a, b]$ massimizza l'entropia tra le densità continue nulle fuori da $[a, b]$.

- La densità esponenziale di parametro $\lambda = 1/m$ massimizza l'entropia tra le densità continue $p(X = x)$ nulle fuori da $[0, \infty)$ e di valor medio fissato
$$\int_0^\infty x p(X = x) dx = m.$$

- La densità esponenziale di parametro $\lambda = 1/m$ massimizza l'entropia tra le densità continue $p(X = x)$ nulle fuori da $[0, \infty)$ e di valor medio fissato
$$\int_0^\infty x p(X = x) dx = m.$$
- Tra le densità discrete a valori in $\mathbb{N}$ con valor medio m, l'entropia è massima per una variabile con densità **geometrica**, ossia
$$P(X = k) \propto (1 - p)^k$$
Si calcola che
$$\mathbb{E}[X] = \frac{1 - p}{p},$$
da cui $p = 1/(m + 1)$ e quindi si può anche scrivere
$$P(X = k) = \frac{1}{m+1} \left(\frac{m}{m+1} \right)^k.$$