

# Probabilità e Processi Stocastici (455AA)

## Lezione 9

Dario Trevisan

23/10/2023

## Section 1

**Covarianza**

# Richiami

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

## Richiami

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

## Richiami

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

# Richiami

- *Varianza*  $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- *Deviazione standard*  $\sigma_X = \sqrt{\text{Var}(X)}$
- Diseguaglianza di Chebychev
- *Covarianza*  
 $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

# Richiami

- *Varianza*  $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- *Deviazione standard*  $\sigma_X = \sqrt{\text{Var}(X)}$
- Diseguaglianza di Chebychev
- *Covarianza*  
 $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
- *Coefficiente di correlazione*

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

# Richiami

- *Varianza*  $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- *Deviazione standard*  $\sigma_X = \sqrt{\text{Var}(X)}$
- Diseguaglianza di Chebychev
- *Covarianza*  
 $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
- *Coefficiente di correlazione*

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

- *Matrice delle covarianze* di  $X = (X_i)_{i=1}^d$

$$(\Sigma_X)_{ij} = \text{Cov}(X_i, X_j).$$

## Richiami

- *Varianza*  $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- *Deviazione standard*  $\sigma_X = \sqrt{\text{Var}(X)}$
- Diseguaglianza di Chebychev
- *Covarianza*  
 $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
- *Coefficiente di correlazione*

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

- *Matrice delle covarianze* di  $X = (X_i)_{i=1}^d$   
 $(\Sigma_X)_{ij} = \text{Cov}(X_i, X_j).$

- $\Sigma_X \in \mathbb{R}^{d \times d}$  è simmetrica e semidefinita positiva.

# Dimostrazione positività

# 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  $\Sigma_X$ ).

- La trasformazione  $UX$ , corrisponde ad un cambio di coordinate e trasforma la covarianza

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

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

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

# Commenti

# Esempio: densità Binomiale

# 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  $X_1$ ?





# Problema

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

- 1 Dire se le variabili  $(\Lambda, T)$  sono tra loro indipendenti.
- 2 Calcolare il coefficiente di correlazione tra  $\Lambda$  e  $T$ .
- 3 Si osserva che  $T \geq 1$ . Come cambia la densità di  $T$ ? e di  $\Lambda$ ? e il coefficiente di correlazione?







## Section 2

# Momenti

# Approssimazione della legge tramite polinomi

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

- Consideriamo uno sviluppo di Taylor per  $g$

$$g(x) \sim a_0 + a_1x + a_2x^2 + \dots + a_kx^k$$

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

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

# Approssimazione della legge tramite polinomi

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

- Consideriamo uno sviluppo di Taylor per  $g$

$$g(x) \sim a_0 + a_1x + a_2x^2 + \dots + a_kx^k$$

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

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

- Problemi:

# Approssimazione della legge tramite polinomi

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

- Consideriamo uno sviluppo di Taylor per  $g$

$$g(x) \sim a_0 + a_1x + a_2x^2 + \dots + a_kx^k$$

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

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

- Problemi:
  - 1 determinare un polinomio approssimante per  $g$

# Approssimazione della legge tramite polinomi

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

- Consideriamo uno sviluppo di Taylor per  $g$

$$g(x) \sim a_0 + a_1x + a_2x^2 + \dots + a_kx^k$$

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

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

- Problemi:

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

# 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

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

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

- se  $X$  ha densità (discreta o continua) vale

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

## Momento secondo e varianza, terzo

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

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

## Momento secondo e varianza, terzo

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

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

- 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:

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

- 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:

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

- 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  $\text{MGF}_X : \mathbb{R} \rightarrow [0, \infty]$ , che associa

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

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

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

## 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  $\text{MGF}_X : \mathbb{R} \rightarrow [0, \infty]$ , che associa

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

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

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

- 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  $\text{MGF}_X(t) = \infty$ .

## 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  $\text{MGF}_X : \mathbb{R} \rightarrow [0, \infty]$ , che associa

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

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

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

- 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  $\text{MGF}_X(t) = \infty$ .
- Per  $t = 0$ ,

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

# Esempi

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

# Esempi

# Proprietà della MGF

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

①  $MGF_{aX+b}(t) = e^{tb} MGF_X(at)$

# Proprietà della MGF

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

- 1  $MGF_{aX+b}(t) = e^{tb} MGF_X(at)$
- 2 Se  $X, Y$  sono indipendenti, allora  $MGF_{X+Y}(t) = MGF_X(t) MGF_Y(t)$ .

# Dimostrazione

## Dalla MGF ai momenti

Sia  $X \in \mathbb{R}$  tale che  $\text{MGF}_X(t) < \infty$  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

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

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

# Dimostrazione (cenno)

## 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 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

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

## 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

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

- il momento terzo

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

## 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

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

- il momento terzo

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

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

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



## 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

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

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

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

- 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

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

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

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

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

## Notazione matriciale

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

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

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

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

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

# Formula di inversione

Come prima conseguenza otteniamo la formula di inversione

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

che esplicitamente diventa

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

## Identità dell'energia

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

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

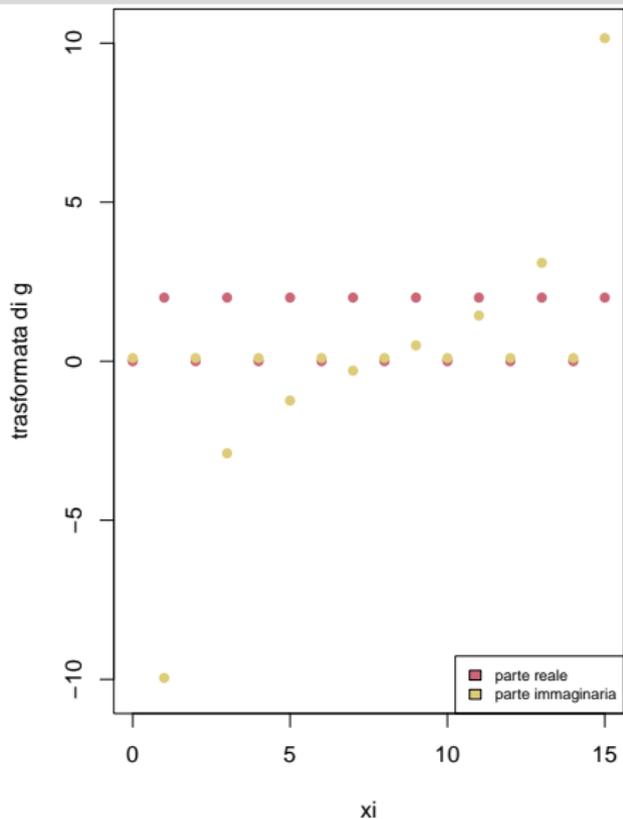
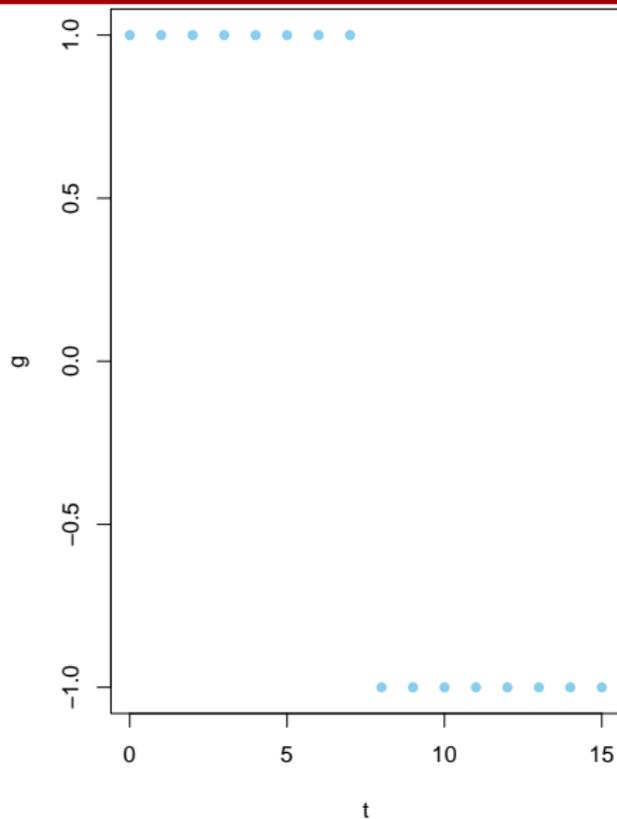
- La norma  $|g|^2$  può essere interpretata come una *energia* del segnale  $g$ , di conseguenza l'identità sopra mostra che la stessa energia può essere ottenuta sommando le energie associate alle singole frequenze, ossia  $|\hat{g}(\xi)|^2$  (e dividendo per  $n$ ).

# Calcolo mediante R della trasformata di Fourier

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

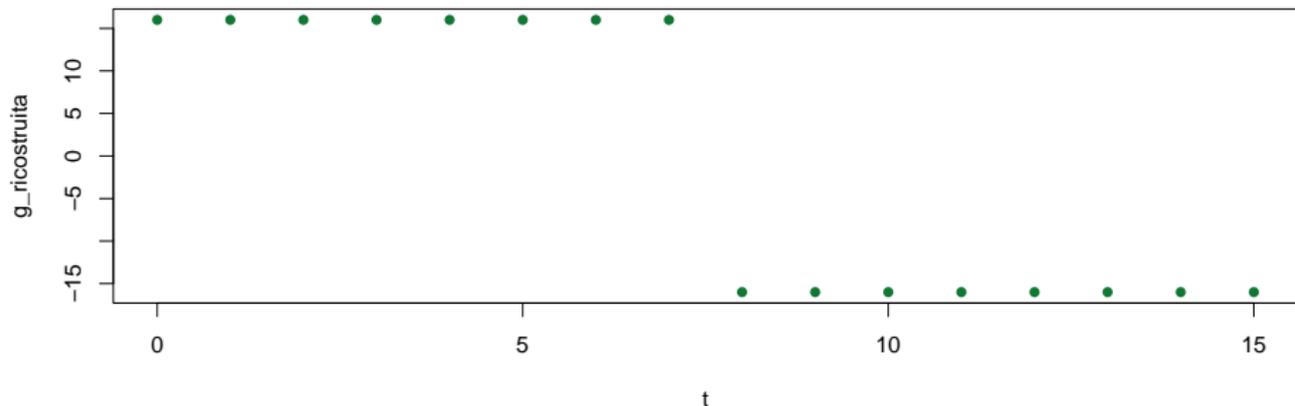
# Calcolo mediante R della trasformata di Fourier

- 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()`.



**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  $g : \mathbb{Z} \rightarrow \mathbb{C}$  (è una situazione ideale ovviamente).

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

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

purché la serie converga, ad esempio se

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

## Caso discreto

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

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

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

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  $\xi$  a  $\xi/n$ , in modo che il dominio sia l'intervallo discreto  $\{0, 1/n, 2/n, \dots, (n-1)/n\}$ . In questo modo, per  $n \rightarrow \infty$  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

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

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

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

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

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

# Convoluzione e trasformata

- Quando si interpreta  $g$  come un segnale, la *convoluzione* di  $g$  con un “filtro”  $f : \mathbb{Z} \rightarrow \mathbb{C}$  è una operazione naturale:

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

Passando alla trasformata di Fourier, possiamo scrivere

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

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

Passando alla trasformata di Fourier, possiamo scrivere

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

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

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

## Caso continuo

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

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

la trasformata è definita come

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

- 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

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

la trasformata è definita come

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

- 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

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

(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

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

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

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

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

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

## Section 4

**Funzione caratteristica**

# Motivazione

- Per il calcolo  $\mathbb{E}[g(X)]$  possiamo approssimare  $g(x)$  usando la *trasformata di Fourier*

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

dove  $\hat{g}(\xi)$  è la trasformata (diretta) di Fourier,

$$\hat{g}(\xi) = \int_{-\infty}^{\infty} g(x) e^{-2\pi i \xi x} dx.$$

# Motivazione

- Per il calcolo  $\mathbb{E}[g(X)]$  possiamo approssimare  $g(x)$  usando la *trasformata di Fourier*

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

dove  $\hat{g}(\xi)$  è la trasformata (diretta) di Fourier,

$$\hat{g}(\xi) = \int_{-\infty}^{\infty} g(x) e^{-2\pi i \xi x} dx.$$

- Approssimiamo (passiamo alla frequenza angolare)

$$g(x) \sim \sum_{\omega} a_{\omega} e^{i\omega x}$$

# Motivazione

- Per il calcolo  $\mathbb{E}[g(X)]$  possiamo approssimare  $g(x)$  usando la *trasformata di Fourier*

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

dove  $\hat{g}(\xi)$  è la trasformata (diretta) di Fourier,

$$\hat{g}(\xi) = \int_{-\infty}^{\infty} g(x) e^{-2\pi i \xi x} dx.$$

- Approssimiamo (passiamo alla frequenza angolare)

$$g(x) \sim \sum_{\omega} a_{\omega} e^{i\omega x}$$

- Passando al valor medio:

$$\mathbb{E}[g(X)] \sim \sum_{\omega} a_{\omega} \mathbb{E}\left[e^{i\omega X}\right].$$

Abbiamo ridotto il problema al calcolo di

$$\mathbb{E} \left[ e^{i\omega X} \right] = \mathbb{E} [\cos(\omega X)] + i\mathbb{E} [\sin(\omega X)].$$

- Data una variabile aleatoria  $X \in \mathbb{R}$ , si definisce la sua **funzione caratteristica**  $\varphi_X : \mathbb{R} \rightarrow \mathbb{C}$ ,

$$\omega \mapsto \varphi_X(\omega) = \mathbb{E} \left[ e^{i\omega X} \right].$$

Abbiamo ridotto il problema al calcolo di

$$\mathbb{E} \left[ e^{i\omega X} \right] = \mathbb{E} [\cos(\omega X)] + i\mathbb{E} [\sin(\omega X)].$$

- Data una variabile aleatoria  $X \in \mathbb{R}$ , si definisce la sua **funzione caratteristica**  $\varphi_X : \mathbb{R} \rightarrow \mathbb{C}$ ,

$$\omega \mapsto \varphi_X(\omega) = \mathbb{E} \left[ e^{i\omega X} \right].$$

- $\varphi_X(\omega)$  è sempre ben definita (ma complessa):

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

# Proprietà

Siano  $X, Y \in \mathbb{R}$  variabili aleatorie e  $a, b \in \mathbb{R}$  costanti (rispetto all'informazione nota  $I$ ). Allora

$$\textcircled{1} \varphi_{aX+b}(\omega) = e^{i\omega b} \varphi_X(a\omega)$$

# Proprietà

Siano  $X, Y \in \mathbb{R}$  variabili aleatorie e  $a, b \in \mathbb{R}$  costanti (rispetto all'informazione nota  $I$ ). Allora

- 1  $\varphi_{aX+b}(\omega) = e^{i\omega b} \varphi_X(a\omega)$
- 2 Se  $X, Y$  sono indipendenti, allora  $\varphi_{X+Y}(\omega) = \varphi_X(\omega)\varphi_Y(\omega)$ .

# Proprietà

Siano  $X, Y \in \mathbb{R}$  variabili aleatorie e  $a, b \in \mathbb{R}$  costanti (rispetto all'informazione nota  $I$ ). Allora

- ①  $\varphi_{aX+b}(\omega) = e^{i\omega b} \varphi_X(a\omega)$
- ② Se  $X, Y$  sono indipendenti, allora  $\varphi_{X+Y}(\omega) = \varphi_X(\omega)\varphi_Y(\omega)$ .
- ③ Se  $X$  ha momento  $k$ -esimo finito, allora

$$\frac{d^k}{d^k \omega} \varphi_X(0) = i^k \mathbb{E} [X^k].$$

# Proprietà

Siano  $X, Y \in \mathbb{R}$  variabili aleatorie e  $a, b \in \mathbb{R}$  costanti (rispetto all'informazione nota  $I$ ). Allora

- ①  $\varphi_{aX+b}(\omega) = e^{i\omega b} \varphi_X(a\omega)$
- ② Se  $X, Y$  sono indipendenti, allora  $\varphi_{X+Y}(\omega) = \varphi_X(\omega)\varphi_Y(\omega)$ .
- ③ Se  $X$  ha momento  $k$ -esimo finito, allora

$$\frac{d^k}{d^k \omega} \varphi_X(0) = i^k \mathbb{E} [X^k].$$

- ④  $\varphi_X(\omega) = \varphi_Y(\omega)$  per ogni  $\omega \in \mathbb{R}$  se e solo se  $X$  e  $Y$  hanno la stessa legge (ossia la stessa densità discreta o continua, quando esistono).

# Esempi

## Il caso vettoriale

Se  $X \in \mathbb{R}^d$ , la funzione generatrice (come anche la trasformata di Fourier) è funzione di  $d$  variabili  $(\omega_1, \omega_2, \dots, \omega_d) \in \mathbb{R}^d$ ,

$$\varphi_X(\omega) = \mathbb{E} \left[ e^{it \cdot \omega} \right] = \mathbb{E} \left[ \exp \left( \sum_{i=1}^d \omega_i X_i \right) \right].$$

- Vale

$$\varphi_{AX+b}(\omega) = e^{ib \cdot \omega} \varphi_X(A^T \omega)$$

## Il caso vettoriale

Se  $X \in \mathbb{R}^d$ , la funzione generatrice (come anche la trasformata di Fourier) è funzione di  $d$  variabili  $(\omega_1, \omega_2, \dots, \omega_d) \in \mathbb{R}^d$ ,

$$\varphi_X(\omega) = \mathbb{E} \left[ e^{it \cdot \omega} \right] = \mathbb{E} \left[ \exp \left( \sum_{i=1}^d \omega_i X_i \right) \right].$$

- Vale

$$\varphi_{AX+b}(\omega) = e^{ib \cdot \omega} \varphi_X(A^T \omega)$$

- e come nel caso reale

$$\varphi_X(\omega) = \varphi_Y(\omega) \quad \text{per ogni } \omega \in \mathbb{R}^d$$

se e solo se  $X$  e  $Y$  hanno la stessa legge (ossia la stessa densità discreta o continua, quando esistono).

## Section 5

**Entropia**

# Una misura dell'informazione (o della sua assenza)

- Vogliamo introdurre una misura del grado di “ignoranza” (o dell'assenza di informazione)

$$H(X)$$

riguardo a quale alternativa sia vera in un dato sistema (associato ad una variabile  $X$ ) e sulla base dell'informazione nota  $I$ .

# Una misura dell'informazione (o della sua assenza)

- Vogliamo introdurre una misura del grado di “ignoranza” (o dell'assenza di informazione)

$$H(X)$$

riguardo a quale alternativa sia vera in un dato sistema (associato ad una variabile  $X$ ) e sulla base dell'informazione nota  $I$ .

- Tanto maggiore è l'ignoranza, maggiore sarà  $H(X)$ .

# Una misura dell'informazione (o della sua assenza)

- Vogliamo introdurre una misura del grado di “ignoranza” (o dell'assenza di informazione)

$$H(X)$$

riguardo a quale alternativa sia vera in un dato sistema (associato ad una variabile  $X$ ) e sulla base dell'informazione nota  $I$ .

- Tanto maggiore è l'ignoranza, maggiore sarà  $H(X)$ .
- Più precisa invece è l'informazione, più piccola sarà  $H(X)$ .

## Definizione di entropia

La scelta più utile (ha migliori proprietà di calcolo) è l'entropia di Shannon

$$H(X) = \begin{cases} -\sum_{x \in E} P(X = x) \log(P(X = x)) & \text{se } X \in E \text{ ha densità discreta,} \\ -\int_{\mathbb{R}^d} p(X = x) \log(p(X = x)) dx & \text{se } X \in \mathbb{R}^d \text{ ha densità continua} \end{cases}$$

- La scelta di base del logaritmo dipende dai vari ambiti (in alcuni casi è preferibile la base 2).

## Definizione di entropia

La scelta più utile (ha migliori proprietà di calcolo) è l'entropia di Shannon

$$H(X) = \begin{cases} -\sum_{x \in E} P(X = x) \log(P(X = x)) & \text{se } X \in E \text{ ha densità discreta,} \\ -\int_{\mathbb{R}^d} p(X = x) \log(p(X = x)) dx & \text{se } X \in \mathbb{R}^d \text{ ha densità continua} \end{cases}$$

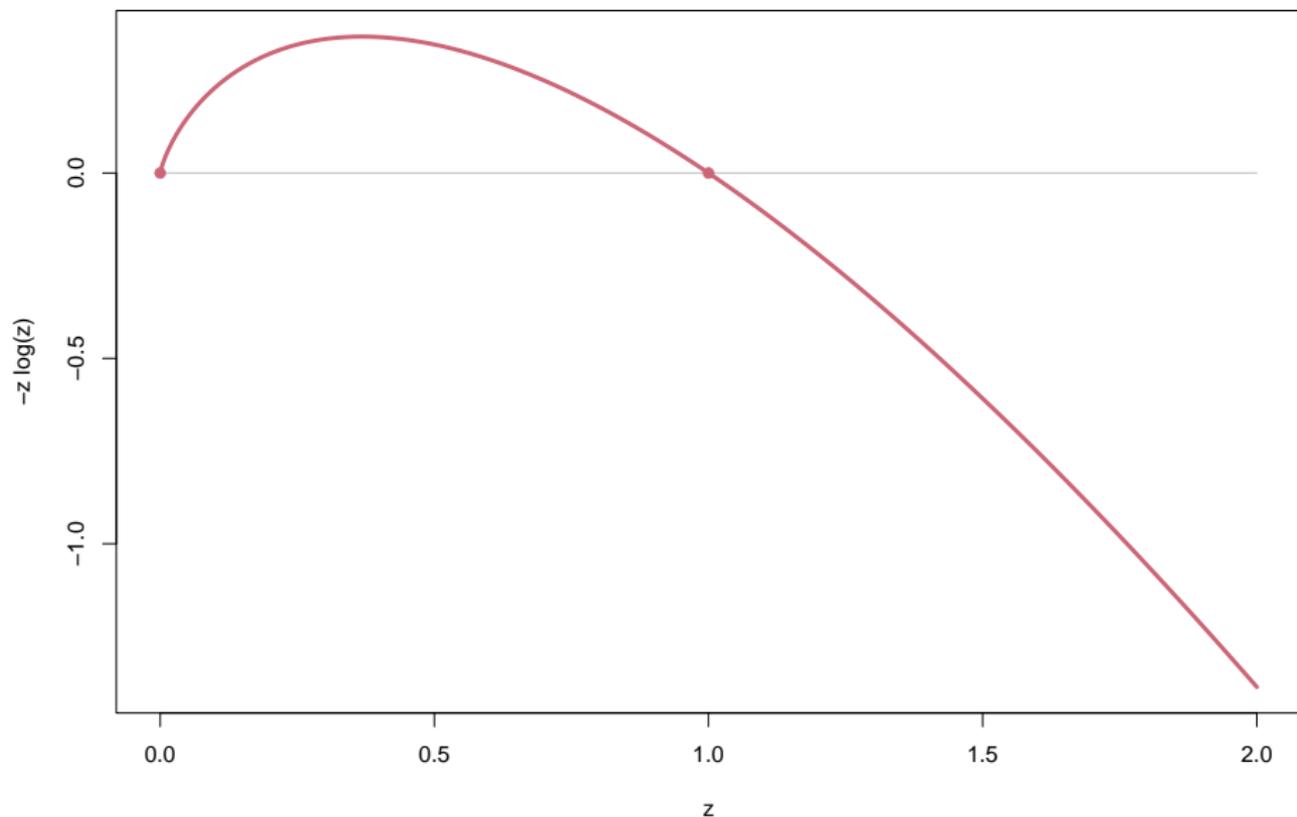
- La scelta di base del logaritmo dipende dai vari ambiti (in alcuni casi è preferibile la base 2).
- Nel caso discreto,  $H(X) \geq 0$  ed è nulla solo se  $X$  è costante (rispetto all'informazione di cui si dispone)

## Definizione di entropia

La scelta più utile (ha migliori proprietà di calcolo) è l'entropia di Shannon

$$H(X) = \begin{cases} -\sum_{x \in E} P(X = x) \log(P(X = x)) & \text{se } X \in E \text{ ha densità discreta,} \\ -\int_{\mathbb{R}^d} p(X = x) \log(p(X = x)) dx & \text{se } X \in \mathbb{R}^d \text{ ha densità continua} \end{cases}$$

- La scelta di base del logaritmo dipende dai vari ambiti (in alcuni casi è preferibile la base 2).
- Nel caso discreto,  $H(X) \geq 0$  ed è nulla solo se  $X$  è costante (rispetto all'informazione di cui si dispone)
- Nel caso continuo invece l'entropia può anche essere negativa.



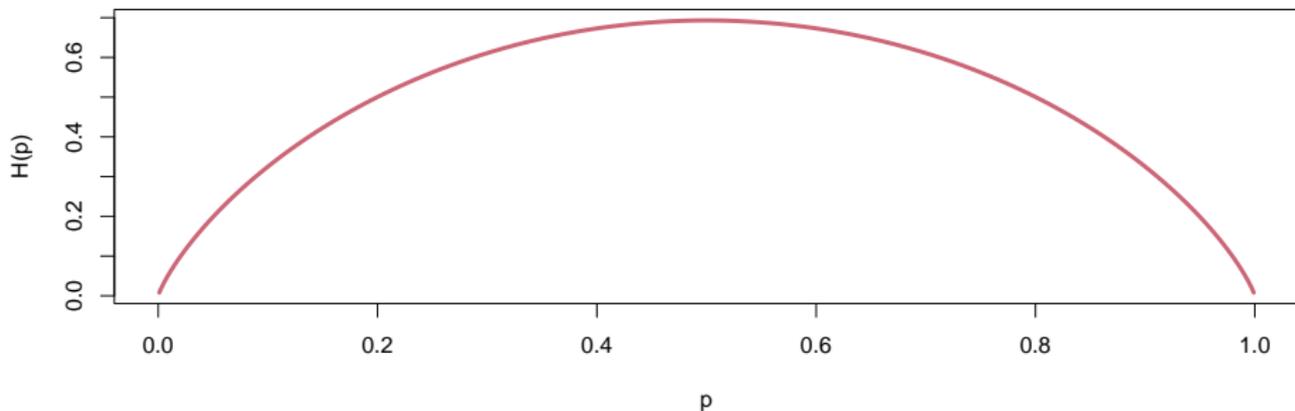
**Figure 2:** grafico della funzione  $-z \log(z)$ .

## Esempi

Sia  $X \in \{0, 1\}$  con legge Bernoulli di parametro  $p \in [0, 1]$ . L'entropia è

$$H(X) = -(1 - p) \log(1 - p) - p \log(p).$$

- Detta anche entropia binaria e indicata solo  $H(p)$ .



# 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

$$\int_0^{\infty} xp(X = x)dx = m.$$

- 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

$$\int_0^{\infty} xp(X = x)dx = m.$$

- 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

$$P(X = k) = \frac{1}{m + 1} \left( \frac{m}{m + 1} \right)^k.$$