

Probabilità e Processi Stocastici (455AA)

Lezione 9

Dario Trevisan – <https://web.dm.unipi.it/trevisan>

20/10/2025

Prova scritta 24 Novembre Iscrivetevi su esami.unipi.it

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

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

Richiami

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

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

- ▶ Covarianza $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
- ▶ Positivamente / Negativamente correlate

Matrice delle covarianze

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

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

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

Proprietà

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

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

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

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

$$\text{Var}(v \cdot X) = \Sigma_{v \cdot X} = \underbrace{v^T}_{\underline{\underline{v}}} \Sigma_X v \geq 0$$

ossia Σ_X è (semi-)definita positiva.

Dimostrazione $v^T \sum_x v = \text{Var}(v \cdot x)$

$$\begin{aligned} \text{Var}(v \cdot x) &= \text{Var}\left(\sum_{i=1}^d v_i x_i\right) = \\ &= \text{Cov}\left(\sum_{i=1}^d v_i x_i, \sum_{j=1}^d v_j x_j\right) \\ &= \sum_{j=1}^d v_j \text{Cov}\left(\sum_{i=1}^d v_i x_i, x_j\right) \\ &= \sum_{j=1}^d v_j \sum_{i=1}^d v_i \underbrace{\text{Cov}(x_i, x_j)}_{\Sigma_{ij}} \\ &= \sum_{i,j=1}^d v_i v_j \Sigma_{ij} = v^T \Sigma v \end{aligned}$$

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

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

$$0 \leq \det \Sigma_{(X,Y)} = \text{Var}(x)\text{Var}(y) - (\text{Cov}(x,y))^2$$

$$(\text{Cov}(x,y))^2 \leq \text{Var}(x)\text{Var}(y)$$

$$\left| \frac{\text{Cov}(x,y)}{\sigma_x \sigma_y} \right| \leq \frac{\sigma_x \sigma_y}{\sigma_x \sigma_y} = 1$$

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

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

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

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

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

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

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

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

- ▶ ossia, dopo alcune operazioni elementari

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

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

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

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

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

- ▶ ossia, dopo alcune operazioni elementari

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

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

Standardizzazione nel caso vettoriale

Il teorema spettrale permette di decomporre

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

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

- ▶ 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 .

$$\mathbb{E}[\hat{X}] = 0 \quad \underline{\text{Var}(\hat{X}) = \text{Id}}$$

- ▶ 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

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

- ▶ 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

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

- ▶ Attenzione! quando si passa alle osservazioni di un campione, la *standardizzazione* si riferisce all'operazione effettuata sulle marginali (comando `scale()` in R).

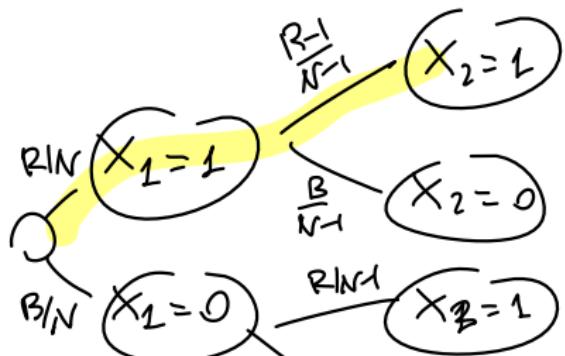
Problema

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 ?

①

x_1, x_2



$(x_2) \rightarrow (x_2)$

$$\text{Cov}(x_1, x_2) < 0 ?$$

$$\mathbb{E}[x_1 x_2] - \mathbb{E}[x_1] \mathbb{E}[x_2]$$

!!

\mathbb{P}_{x_1, x_2}

!!

$$\text{Cov}(x_1, x_2) = -\frac{R}{N} \frac{B}{N} \frac{1}{N-1} = -\frac{1}{N-1}$$

$$\sqrt{\frac{R}{N} \frac{B}{N} \frac{R}{N} \frac{B}{N}}$$

$$\begin{aligned}
 & \left(\frac{R}{N} \cdot \frac{R-1}{N-1} - \frac{R}{N} \cdot \frac{R}{N} \right) = \\
 & = \frac{R}{N} \frac{(R-1)N - R(N-1)}{N(N-1)} \\
 & = \frac{R}{N} \frac{R-N}{N(N-1)} = -\frac{RB}{N^2(N-1)} \\
 & = -\frac{R}{N} \cdot \frac{B}{N} \cdot \frac{1}{N-1}
 \end{aligned}$$

② Matrice delle covariante

Usiamo il seguente fatto la legge empirica di due estrazioni diverse (x_i, x_j) è la stessa delle estrazioni (x_1, x_2) (esempio (x_2, x_3))

$$\Rightarrow \text{Cov}(x_i, x_j) = [E(x_i x_j) - E[x_i] E[x_j]] \quad (x_1, x_2)$$

| ↑ congiunta ↑ marginali
 $= \text{Cov}(x_1, x_2)$

$$\sum_X = \begin{pmatrix} \dots & -\frac{R_B}{N} \frac{1}{N} \\ \dots & \dots \\ -\frac{R_B}{N} \frac{1}{N} & \dots \end{pmatrix}$$

diagonale $\rightarrow \frac{R_B}{N} \frac{1}{N} = \text{Var}(x_i)$

$$\textcircled{3} \quad Y = \text{numero di "successi" in n esperimenti} \\ = X_1 + X_2 + \dots + X_n$$

$$E[Y] = E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] = n \cdot \frac{R}{N}$$

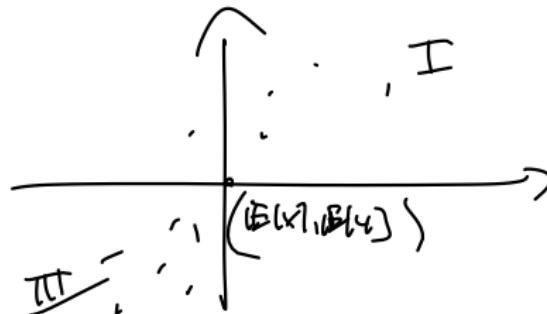
$$\text{Var}(Y) = \text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j)$$
$$= n \underbrace{\frac{R}{N} \frac{B}{N}}_{-} - n(n-1) \cdot \frac{R}{N} \frac{B}{N} \cdot \frac{1}{N-1}$$

$$\textcircled{4} \quad \text{Cov}(X_1, Y) = \text{Cov}(X_1, \sum_{i=1}^n X_i)$$

$$= \sum_{i=1}^n \text{Cov}(X_1, X_i) = \frac{R}{N} \frac{B}{N} - (n-1) \frac{R}{N} \frac{B}{N} \frac{1}{n-1}$$

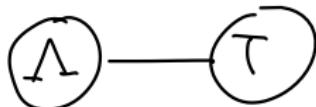
$$= \frac{R}{N} \frac{B}{N} \left(1 - \frac{n-1}{n-1} \right)$$

$$\text{Se } n \leq N \rightarrow n-1 \leq N-1 \rightarrow \frac{n-1}{N-1} \leq 1$$



$$1 - \frac{n-1}{N-1} \geq 1 - 1 \geq 0$$

Problema



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.
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?

* Cambiare densità $\left\{ p(\lambda = \lambda) = \frac{\lambda^2}{2} \exp(-\lambda) \quad \lambda > 0 \right.$

$$\textcircled{1} \quad \Lambda \longrightarrow T \quad p(\Lambda, T) = p(\Lambda) p(T|\Lambda)$$

$$= \exp(1) \cdot \exp(-\Lambda)$$

$$p(\Lambda=\lambda, T=t) = \exp(-\lambda) \cdot \lambda \exp(-\lambda t) \quad \boxed{\begin{array}{l} \lambda > 0 \\ t > 0 \end{array}}$$

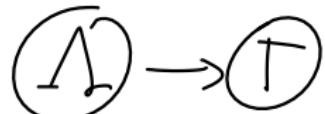
$p(T=t | \Lambda=\lambda)$ dipende da $\lambda \Rightarrow$ non sono indipendenti

$$\textcircled{2} \quad \begin{aligned} & E[\Lambda], E[T], E[\Lambda T] \\ & E[\Lambda^2], E[T^2] \end{aligned}$$

$$E[\Lambda] = E[\exp(1)] = 1$$

$$E[\Lambda^2] = E[\exp(1)^2] = \int_0^{+\infty} \exp(-\lambda) d\lambda = 2$$

$$\mathbb{E}[T] = \int_0^{+\infty} \mathbb{E}[T | \Lambda = \lambda] \rho(\Lambda = \lambda) d\lambda$$

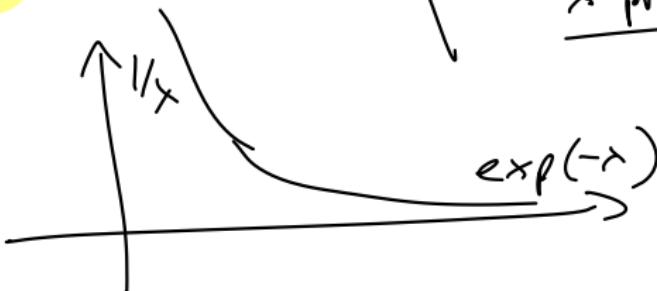


disintegrazione

$$\exp(-\lambda)$$

$$= \int_0^{+\infty} \frac{1}{\lambda} \exp(-\lambda) d\lambda = +\infty$$

$\Delta \approx$ priori
troppo pesante
su i valori
 λ priori



$$\begin{aligned}
 \textcircled{1*} \quad \mathbb{E}[\Lambda] &= \int_0^{+\infty} \lambda \frac{\lambda^2}{2} e^{-\lambda} d\lambda = \frac{1}{2} \int_0^{+\infty} \lambda^3 e^{-\lambda} d\lambda \\
 &= + \frac{1}{2} \int_0^{+\infty} \left(\frac{d}{dx} \lambda^3 \right) e^{-\lambda} d\lambda \quad \frac{d}{dx} (-e^{-x}) \\
 &= \frac{3}{2} \int_0^{+\infty} \lambda^2 e^{-\lambda} d\lambda = 3
 \end{aligned}$$

$$\mathbb{E}[\Lambda^2] = \int_0^{+\infty} \lambda^2 \frac{\lambda^2}{2} e^{-\lambda} d\lambda \stackrel{IBP}{=} 4 \int_0^{+\infty} \lambda^3 \frac{e^{-\lambda}}{2} d\lambda = 4 \cdot 3$$

$$\text{Var}(\Lambda) = 4 \cdot 3 - 3^2 = \underline{\underline{3}}$$

$$2^* \quad \mathbb{E}[T] = \int_0^{+\infty} \mathbb{E}[T | \lambda=\lambda] \frac{\lambda^2}{2} e^{-\lambda} d\lambda$$

$$= \int_0^{+\infty} \frac{1}{\lambda} \cdot \frac{\lambda^2}{2} e^{-\lambda} d\lambda = \frac{1}{2}$$

$$\mathbb{E}[T^2] = \int_0^{+\infty} \mathbb{E}[T^2 | \lambda=\lambda] \frac{\lambda^3}{2} e^{-\lambda} d\lambda$$

$$= \int_0^{+\infty} \frac{2}{\lambda} \cdot \frac{\lambda^3}{2} e^{-\lambda} d\lambda = 1$$

$$\text{Var}(z) = \mathbb{E}(z^2) - \mathbb{E}(z)^2 \quad \text{Var}(T) = 1 - \left(\frac{1}{2}\right)^2 = \frac{3}{4}$$

$$\mathbb{E}[\Delta T] = \int_0^{+\infty} \mathbb{E}[\Delta T | \Delta = \lambda] \frac{\lambda^2}{2} e^{-\lambda} d\lambda$$

\uparrow
 λ

$$= \int_0^{+\infty} \lambda \underbrace{\mathbb{E}[T | \Delta = \lambda]}_{\frac{1}{\lambda}} \frac{\lambda^2}{2} e^{-\lambda} d\lambda$$

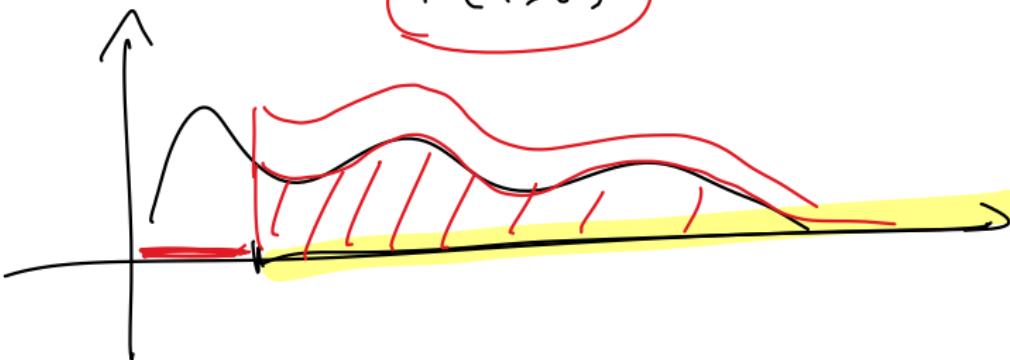
$$= \int_0^{+\infty} \frac{\lambda^3}{2} e^{-\lambda} d\lambda = 1$$

$$\overline{\text{Cov}(\Delta, T)} = \mathbb{E}[\Delta T] - \mathbb{E}[\Delta] \mathbb{E}[T]$$

$$= 1 - 3 \cdot \frac{1}{2} = -\frac{1}{2} \leq 0$$

$$\overline{\text{Cor}(\Delta, T)} = \frac{-\frac{1}{2}}{\sqrt{3 \cdot 3/4}} = -\frac{1}{2} \cdot \frac{2}{3} = \boxed{-\frac{1}{3}}$$

$$\begin{aligned}
 (3^*) \quad p(T=t | T \geq 1) &= \frac{p(T=t, T \geq 1)}{P(T \geq 1)} \\
 &= \begin{cases} 0 & \text{se } t < 1 \\ \frac{p(T=t)}{P(T \geq 1)} & \text{se } t \geq 1 \end{cases}
 \end{aligned}$$



$$P(T=t, \Lambda=\lambda | T \geq 1) = \begin{cases} 0 & \text{se } t < 1 \\ \frac{P(T=t, \Lambda=\lambda)}{P(T \geq 1)} & \text{altrimenti} \end{cases}$$

per conoscere le proprietà di Λ dobbiamo integrale

$$\begin{aligned} P(\Lambda \geq \lambda | T \geq 1) &= \int_0^{+\infty} P(T=t, \Lambda \geq \lambda | T \geq 1) dt \\ &= \int_1^{+\infty} P\left(\frac{T=t, \Lambda=\lambda}{P(T \geq 1)}\right) dt \end{aligned}$$

Momenti

$$\mathbb{E}[x], \quad \mathbb{E}[x^2], \dots$$

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

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

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

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

misuramenti

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

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

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

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

- ▶ 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

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

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

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

- ▶ 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

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

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

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

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

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

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

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

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

\downarrow
trasf di Laplace della $f(x) = p(X=x)$

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

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

- ▶ 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 $\mathbb{E}[e^{tX}]$ 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

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

- ▶ 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 $\mathbb{E}[e^{tX}]$ non convergono, si pone $\text{MGF}_X(t) = \infty$.

- ▶ Per $t = 0$,

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

Esempi

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

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

Allora

$$\text{MGF}_X(t) = \begin{cases} \frac{1}{1-t/\lambda} & \text{se } t < \lambda \\ +\infty & \text{altrimenti} \end{cases}$$

Esempi

Bernoulli(p) $\mathbb{E}[e^{tX}] = e^{t \cdot 1} \cdot p(X=1) +$

$$+ e^{t \cdot 0} p(X=0)$$

$$= e^t p + (1-p)$$

$$= 1 + p(e^t - 1)$$

Poisson(λ) $P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$

$$\mathbb{E}[e^{tX}] = \sum_{k=0}^{+\infty} e^{tk} \frac{x^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(e^t \lambda)^k}{k!}$$

Esempi

$$\hookrightarrow MGF_X(t) = e^{-\lambda} e^{\lambda e^t} = \exp(\lambda(e^t - 1))$$

Proprietà della MGF

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

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

$$\begin{aligned}\mathbb{E}[\exp(t(\underline{ax+b}))] &= \exp(tb) \mathbb{E}[\exp(t\underline{a}x)] \\ &= \exp(tb) \mathbb{E}[f_X(at)]\end{aligned}$$

Proprietà della MGF

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

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

$$\text{MGF}_{X+Y}(t) = \text{MGF}_X(t) \text{MGF}_Y(t).$$

$$\mathbb{E}[\exp(t(X+Y))] = \mathbb{E}[\underbrace{\exp(tx)}_{f(x)} \underbrace{\exp(tY)}_{g(Y)}]$$

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)

$$MGF_X(t) := E[e^{tx}] = E\left[\sum_{k=0}^{+\infty} \frac{(tx)^k}{k!}\right]$$

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

$$= \sum_{k=0}^{+\infty} \frac{t^k}{k!} E[x^k]$$

$$MGF_X(t) = \sum_{k=0}^{+\infty} \omega_k \frac{t^k}{k!} = \sum_{k=0}^{+\infty} \left(\frac{1}{(dt)}^k MGF_X(0) \right) \frac{t^k}{k!}$$

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 valori 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 $t = (t_1, t_2, \dots, t_d) \in \mathbb{R}^d$,

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

$$\cdot \text{MGF}_{AX+b}(t) = e^{t^T b} \text{MGF}_X(A^T t)$$

$$\cdot \text{MGF}_{X+Y}(t) = \text{MGF}_X(t) \text{MGF}_Y(t) \quad \text{se } X, Y \text{ indipendenti}$$

$$\cdot \mathbb{E}\left[\prod_{i=1}^d X_i^{k(i)} \right] = \frac{d}{\prod_{i=1}^d (dt_i)^{k(i)}} \text{MGF}_X(\theta)$$

Richiami sulla trasformata di Fourier

Caso finito

Fissano $n \in \mathbb{N}$, si consideri un *segnalet* 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* 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 $F \in \mathbb{C}^{n \times n}$

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

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

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

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 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()`.

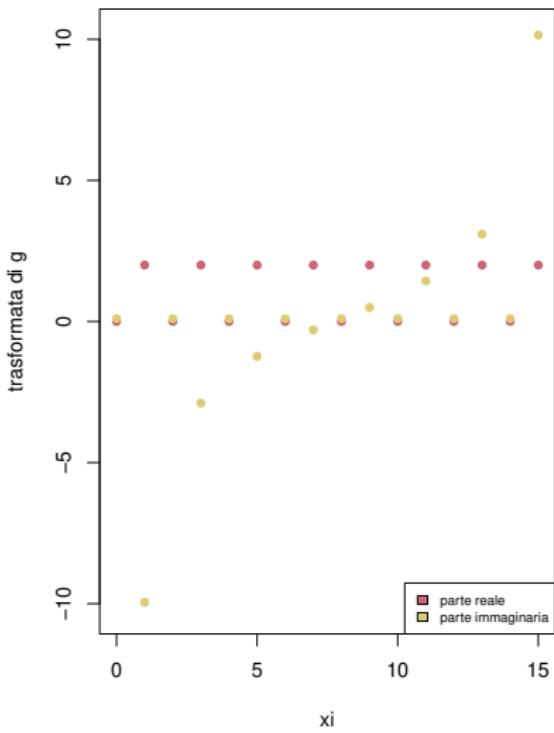
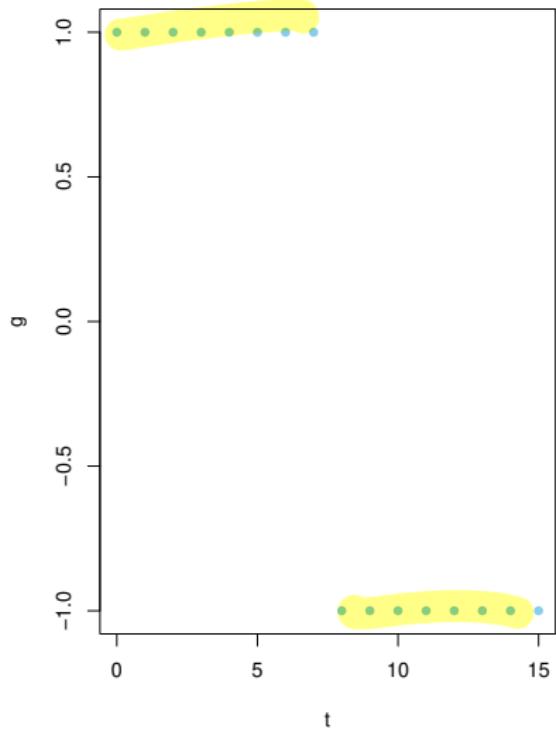
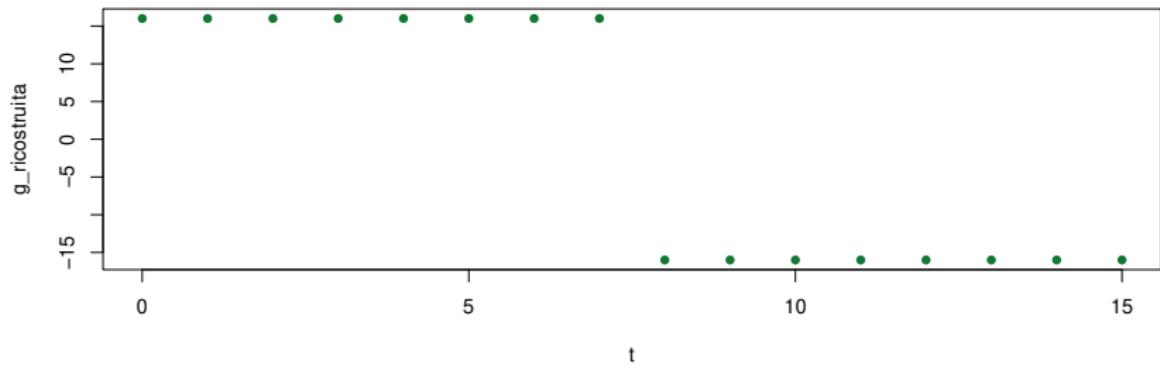


Figure 1: esempio di un segnale finito (onda quadra) e della sua trasformata di Fourier

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 ξ a ξ/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

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

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

Passando alla trasformata di Fourier, possiamo scrivere

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

- ▶ 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:

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

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,

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

$$\begin{aligned} & \mathbb{E} \left[\int_{-\infty}^{+\infty} \hat{g}(\xi) e^{-2\pi i \xi X} d\xi \right] \\ &= \int_{-\infty}^{+\infty} \hat{g}(\xi) \underbrace{\mathbb{E}[e^{-2\pi i \xi X}]}_{\text{ }} d\xi \end{aligned}$$

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:

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

frequenze caratteristiche
di X

Abbiamo ridotto il problema al calcolo di

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

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

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

Abbiamo ridotto il problema al calcolo di

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

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

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

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

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

↑ trasf di Fourier di $p(x=x)$

Proprietà

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

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

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

$$\text{Esempio } f_X(t) = \mathbb{E}[e^{itX}]$$
$$\varphi_X(\omega) = \mathbb{E}[e^{i\omega X}]$$

Bernoulli (p)

$$\varphi_X(\omega) = 1 + p(e^{i\omega} - 1)$$

Binomiale (n, p)

$$X = X_1 + \dots + X_n$$



$$\varphi_{x_1+\dots+x_n}(\omega) = \varphi_{x_1}(\omega) \cdot \varphi_{x_2}(\omega) \cdots \varphi_{x_n}(\omega)$$

$$= \left(1 + p(e^{i\omega} - 1)\right)^n$$

Poisson (λ)

$$\varphi_X(\omega) = \exp(X(e^{i\omega} - 1))$$

Esponentiale

$$\varphi_X(\omega) = \frac{\lambda}{\lambda - i\omega} \quad \forall \omega$$

Il caso vettoriale

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

$$\varphi_X(\omega) = \mathbb{E} \left[e^{\omega \cdot X} \right] = \mathbb{E} \left[\exp \left(i \sum_{j=1}^d \omega_j X_j \right) \right].$$

► 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 $(\omega_1, \omega_2, \dots, \omega_d) \in \mathbb{R}^d$,

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

- ▶ Vale

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

- ▶ e come nel caso reale

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

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

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

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

- ▶ 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.

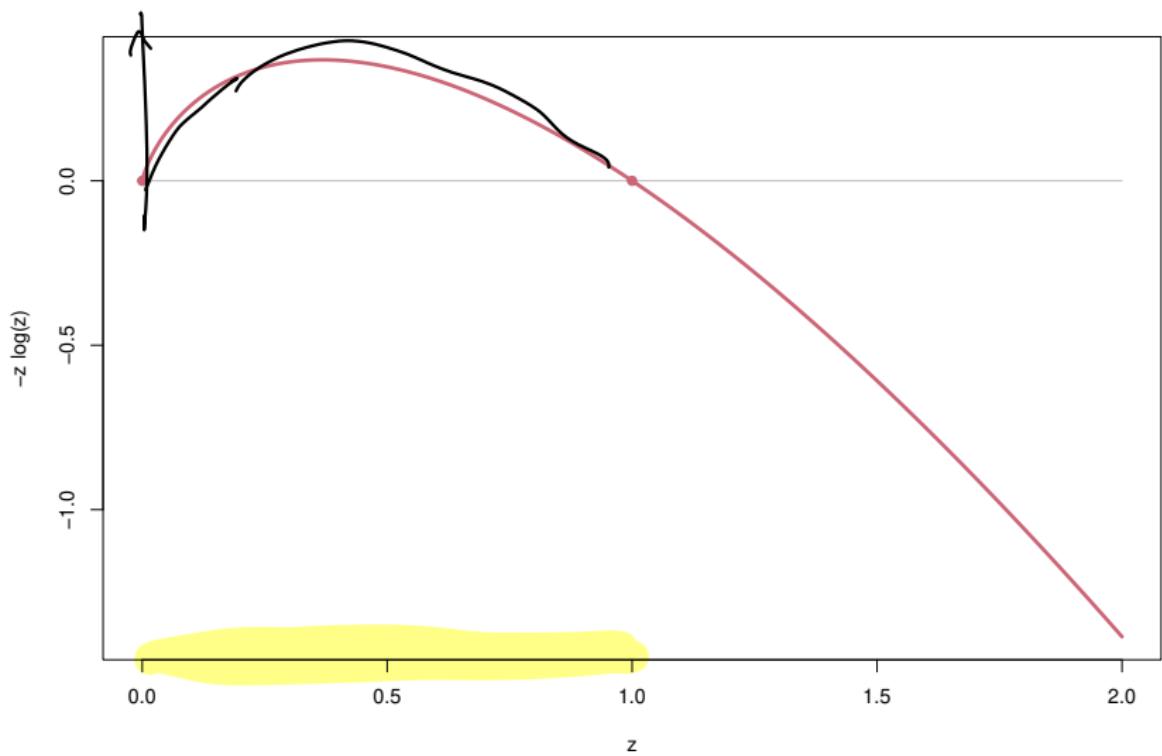


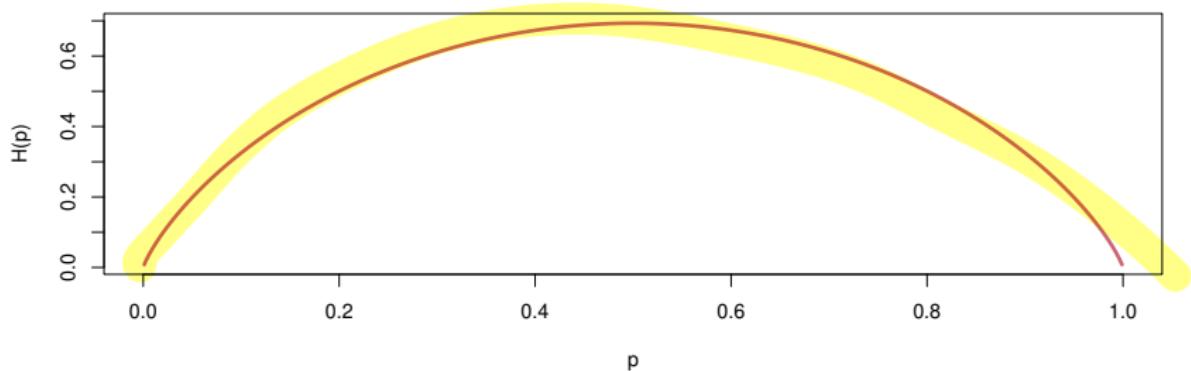
Figure 2: grafico della funzione $-z \log(z)$.

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

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

- ▶ 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

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