Teoria dell'Informazione

Introduzione essenziale

Indice generale

Premessa

Questi appunti contengono una breve introduzione ai concetti fondamentali della Teoria dell'Informazione. Il testo assume che il lettore sia in possesso delle seguenti nozioni propedeutiche:

- Matematica: proprietà dei logaritmi, calcolo combinatorio
- Statistica: probabilità, variabile casuale discreta, valore atteso
- Informatica: concetti di bit e byte, cambio di base di un numero

Misura dell'informazione

La definizione di informazione formulata da Shannon (detta anche *autoinformazione*) è:

$$I(x) = - \log_b(Px)$$

dove **b** è una base arbitraria, **Px** la probabilità che una variabile casuale **X** assuma valore il valore x, dove **x** è un qualsiasi evento "emesso" da una sorgente di informazioni.

Questa definizione non "piove dal cielo", ma deriva da ragionamenti sul concetto di informazione, sorpresa e probabilità. Poiché tali ragionamenti sono già riportati altrove [1] [2], in questa sede eviteremo di giustificare la formula di Shannon: cercheremo piuttosto di fornire alcuni esempi che dovrebbero aiutarne la comprensione.

Iniziamo col capire il significato del logaritmo: siccome il logaritmo rappresenta il **numero di cifre** dell'argomento, la scelta di b è dettata unicamente da criteri pratici. Ad esempio, se l'evento **x** avesse probabilità 1/100 di verificarsi, potremmo scegliere b in modi diversi:

- se b = 10 → x avrebbe informazione I = 2 (infatti 100 ha 2 cifre in base 10)
- se b = 2 → x avrebbe informazione I = 6,6 (infatti 100 ha "poco meno" di 7 cifre in base 2)

Se invece consideriamo un evento **y** con probabilità 1/128, avremmo:

- se b = 10 → y avrebbe informazione I = 2,1 (infatti 128 ha "poco più" di 2 cifre in base 10)
- se b = 2 → y avrebbe informazione I = 7 (infatti 128 ha esattamente 7 cifre in base 2)

Poiché in informatica con 7 bit si possono rappresentare tutti i numeri compresi tra 1 e 128, si capisce che scegliendo b = 2 si ottiene una misura compatibile con le tecnologie informatiche. Inoltre, siccome cambiare base del logaritmo significa moltiplicare per una costante, ovvero

$$\log_2(x) = \log_{10}(x) / \log_{10}(2) \approx \log_{10}(x) \cdot 3{,}32$$

allora tutte le scelte della base b sono equivalenti tra loro, a parte un fattore di scala.

In altre parole, scegliere la base b significa scegliere l'unità di misura: tutti i ragionamenti e le conclusioni che otterremo sono le stesse per qualsiasi scelta di b, a meno di una costante moltiplicativa.

Sorgenti di informazione

La definizione di informazione di Shannon si può applicare a qualsiasi sistema in grado di contenere informazione. Ad esempio, se consideriamo come sorgente di informazione le 21 lettere dell'alfabeto italiano, ovvero:

A,B,C,D,E,F,G,H,I,L,M,N,O,P,Q,R,S,T,U,V,Z

allora "estrarre a caso" un **simbolo** da questa sorgente significa scegliere a caso una di queste 21 lettere. Nella teoria dell'informazione il processo di "estrazione" non modifica la sorgente, come invece avviene nel gioco del lotto quando si estrae una pallina da un'urna. Quando diciamo di "estrarre" informazione dal sistema, intendiamo un'operazione in grado di scegliere uno dei simboli della sorgente, senza doverlo "togliere" dalla sorgente.

L'informazione contenuta dall'estrazione di una delle 21 lettere dell'alfabeto è quindi:

$$I(x) = -\log_b(Px) = -\log_2(1/21) \approx 4{,}39 \text{ bit}$$

In termini informatici questo risultato può essere letto come segue: "per rappresentare una lettera dell'alfabeto italiano servono un po' più di 4 bit, ma meno di 5 bit". Sapendo che con 4 bit possiamo rappresentare un numero da 1 a 16 e con 5 bit possiamo contare fino 32, il risultato è sensato: infatti il numero 21 è compreso tra 16 e 32.

Consideriamo ora una sorgente di informazione contenente sempre 21 caratteri, ma stavolta i simboli non sono tutti diversi, perché l'insieme contiene solamente quattro lettere diverse, ovvero: A,B,C e D. Tale sorgente potrebbe essere qualcosa del genere:

A,A,A,A,A,A,A,B,B,B,B,B,B,B,C,C,C,C,C,C,C,D,D,D,D,D,D,D

dove ogni simbolo è ripetuto sette volte. In questo caso, quando estraggo un simbolo a caso, ho il 25% di probabilità di trovare uno dei quattro simboli disponibili. L'informazione di una singola estrazione è quindi:

$$I(x) = -\log_2(1/4) \approx 2 \text{ bit}$$

Infatti, per rappresentare un generico simbolo scelto tra soli 4 valori, basta una cifra binaria di 2 bit. Questo ci dice che quello che determina l'informazione è il numero di simboli **diversi**, non la quantità di "oggetti" presenti nella sorgente.

Informazione media

Consideriamo una certa sorgente **E** di informazione, ovvero un modo di assegnare un valore ad una variabile casuale X. Se un evento $x \in E$ ha una certa probabilità P_x di verificarsi, quando ciò accade l'osservatore acquisisce l'informazione $I(x) = -\log_b(P_x)$. Se si continuano a ripetere misurazioni, ovvero ad estrarre messaggi dall'insieme, si può calcolare il **valore atteso** dell'informazione di un generico evento, ovvero la quantità di informazione *mediamente* emessa dal sistema E. Indicheremo tale quantità con la lettera **H** maiuscola:

$$H(E) = \langle I(x) \rangle = -\sum P_x \cdot \log_b(P_x)$$

Dove la sommatoria è su tutti gli eventi x che possono essere prodotti (o "estratti") dalla sorgente E, ovvero su tutti i possibili simboli (non per forza equiprobabili). Notiamo inoltre che, per le proprietà dei logaritmi, se due eventi casuali A,B sono **indipendenti** tra loro si ha:

$$I\,(\text{si ottiene sia A che B}) = I(A) + I(B)$$

Utilizzando la definizione di informazione di Shannon possiamo calcolare l’informazione prodotta da alcuni semplici sistemi. Questo ci aiuterà a “toccare con mano” il concetto di informazione. Siccome al termine della discussione sarà utile confrontare questi sistemi daremo un nome ad ogni esempio, in modo da poterli referenziare più chiaramente.

Alcuni sistemi semplici

Alfabeto italiano

Come già visto, l'informazione associata ad un simbolo x scelto a caso tra le 21 lettere dell'alfabeto italiano è:

$$I(x) = -\log_2(1/21) \approx 4{,}39 \text{ bit}$$

In questo caso gli esiti sono tutti equamente probabili, perciò la media coincide con quella di ogni singolo valore:

$$H = \langle I(x) \rangle \approx 4{,}39 \text{ bit}$$

Ciò significa che per rappresentare l'esito di una "misura" del sistema ci servono più di 4 bit di informazione. Supponiamo di voler sviluppare un'applicazione informatica che utilizzi l'informazione così ottenuta per produrre qualcosa. Ad esempio, l'applicazione potrebbe estrarre 8 simboli a caso da questo alfabeto per generare una password casuale. Siccome ogni estrazione è un evento **indipendente** dall'altro, la probabilità di estrarre una certa sequenza di 8 simboli è pari al prodotto delle singole probabilità:

$$P(y = \text{una sequenza di 8 simboli}) = (1/21)\cdot(1/21)\cdot \ldots \cdot(1/21) \qquad \text{(8 volte)}$$

se x rappresenta l'estrazione di un unico simbolo, per le proprietà dei logaritmi abbiamo:

$$I(y) = -\log_b(Py) = I(x) + I(x) + \ldots + I(x) = 8 \cdot I(x) \approx 8 \cdot 4{,}39 \text{ bit} \approx 35 \text{ bit}$$

Inoltre, siccome gli eventi sono tutti equiprobabili tra loro, allora l'informazione media è pari a quella di un singolo evento, per cui l'informazione associata ad una password generata in questo modo sarebbe pari a 35 bit. Proviamo a verificare questo numero mediante un approccio diverso: siccome ogni carattere corrisponde ad un valore tra 1 e 21, si hanno 21^8 password possibili, cioè circa $3{,}8\cdot10^{10}$ combinazioni. Per rappresentare questo numero in binario servono esattamente 35 bit, infatti:

$$\text{Numero di bit} = \log_2(3{,}8\cdot10^{10}) \approx 35 \text{ bit}$$

che coincide perfettamente col risultato ottenuto dalla formula di Shannon.

Nota: se invece dell'alfabeto italiano si utilizzasse la sorgente vista nell'esempio precedente, ovvero l'insieme dei soli simboli A,B,C e D, si otterrebbe:

$$I(x) = -\log_2(1/4) \approx 2 \text{ bit}$$

per cui una password di 8 simboli estratti da questo insieme avrebbe informazione:

$$I(y) = 8 \cdot I(x) = 8 \cdot 2 \text{ bit} = 16 \text{ bit}$$

com'è logico aspettarsi, se la password viene scelta tra un numero minore di simboli (ovvero: una sorgente con meno informazione) allora anche l'informazione mediamente associata alla password è minore.

Dizionario italiano

Consideriamo ora la sorgente di informazione data dal "dizionario italiano", ovvero l'insieme di tutte le parole elencate su un vocabolario della lingua italiana. Immaginiamo di avere un vocabolario di 100 pagine e che ogni pagina contenga 100 parole. Allora per scegliere un simbolo a caso si potrebbero estrarre a caso due numeri tra 1 e 100: il primo numero dice a che pagina andare, il secondo quale parola guardare. Infine, una volta ottenuta una parola, scegliamo a caso una lettera. Ad esempio, se abbiamo estratto la parola "gatto", dovremmo scegliere a caso un numero tra 1 e 5. Se il risultato di quest'ultimo processo fosse 2, allora il simbolo estratto sarebbe la vocale "a", ovvero "x = a".

Di primo acchito si potrebbe pensare che questa sorgente di informazione produca gli stessi 21 simboli dell'alfabeto italiano, per cui l'informazione di una singola estrazione dovrebbe essere la stessa, cioè 4,39 bit. In realtà, in base alla **analisi delle frequenze**, si scopre che alcune lettere sono usate molto più spesso di altre. Ad esempio la vocale "a" compare in circa il 12% delle parole esistenti, mentre la consonante "b" meno del 1% (nella lingua italiana).

In questo caso perciò l'informazione non è la stessa per tutti i simboli, ma dipende dal simbolo considerato. Ad esempio:

$$I(x = a) \approx -\log_2(12/100) \approx 3{,}1 \text{ bit}$$
$$I(x = b) \approx -\log_2(1/100) \approx 6{,}6 \text{ bit}$$

Per cui il calcolo dell'informazione media della sorgente diventa più complesso:

$$H = -\sum P_x \cdot \log_b(P_x) \approx -\left((12/100)\cdot 3{,}1 + (1/100)\cdot 6{,}6 + \ldots\right) \qquad \text{(21 addendi)}$$

Da cui si ottiene un'informazione media di circa 4 bit per simbolo. Tale valore dipende dalla lingua presa in esame e dal periodo storico: ogni lingua è in continua evoluzione, e il concetto di "dizionario" va inteso in modo dinamico. Ad esempio, negli ultimi decenni l'informazione media della lingua italiana è stata misurata più volte, trovando valori compresi tra 3,96 bit e 4,02 bit [3]. Di seguito, per praticità, assumeremo che l'informazione associata alla lingua italiana sia pari a circa 4 bit. Ne segue che una "parola" composta da 8 lettere scelte a caso dal dizionario contiene 32 bit. Da ciò si può ottenere un altro risultato: sapendo che il dizionario contiene 32 bit, possiamo calcolare il numero di combinazioni prodotte dal dizionario stesso. Si trova così che il numero di "parole" nel dizionario è pari a: $2^{32} \approx 4{,}3\cdot 10^9$. Siccome le combinazioni ottenibili dall'alfabeto sono associate a $3{,}8\cdot 10^{10}$ "parole" diverse, si deduce che in media l'alfabeto contiene più informazione del dizionario.

Alfabeto contro Dizionario

Negli esempi precedenti abbiamo discusso due sorgenti di informazione simili tra loro: l'alfabeto italiano (i 21 simboli possibili) e il dizionario italiano (i simboli utilizzati dall'insieme di tutte le parole della lingua italiana). L'informazione media associata a questi due insiemi è:

H(alfabeto) ≈ 4,39 bit
H(dizionario) ≈ 4 bit

ovvero l'informazione media contenuta nel dizionario è *minore* di quella contenuta nell'alfabeto. Infatti abbiamo visto che dall'alfabeto possiamo estrarre $3{,}8 \cdot 10^{10}$ "parole" di 8 caratteri, mentre dal dizionario ne possiamo estrarre "solo" $4{,}3 \cdot 10^{9}$ (sempre da 8 caratteri). Qualcuno potrebbe pensare che "il dizionario dovrebbe contenere più informazione dell'alfabeto". Dobbiamo però pensare ai due sistemi come dei dispositivi in grado di generare un carattere tipografico scelto a caso: quando si tiene conto della frequenza dei simboli l'alfabeto genera caratteri in modo più imprevedibile, perciò contiene più informazione.

Il paragone tra alfabeto e dizionario è simile all'esempio in cui abbiamo confrontato l'informazione dell'alfabeto (21 simboli diversi) con quella di un insieme contenente solamente 4 simboli distinti (le lettere A,B,C,D). Anche in quel caso, siccome i simboli si ripetevano all'interno dell'insieme, l'informazione dell'insieme di soli 4 simboli era minore dell'insieme con 21 simboli diversi. Lo stesso vale per il dizionario: siccome alcune lettere sono usate più delle altre, l'informazione contenuta dal dizionario è in genere minore dell'alfabeto.

Esempio: proviamo a generare due password casuali di 8 caratteri: la prima estraendo i caratteri dall'alfabeto, la seconda dal dizionario:

Esempio di password prodotta dall'alfabeto: *zfgibmfn* (35 bit)
Esempio di password prodotta dal dizionario: *adilnrot* (32 bit)

siccome nel dizionario alcuni simboli sono più frequenti di altri, è probabile che la password ottenuta dal dizionario contenga alcuni dei simboli "più comuni", come ad esempio le lettere "a, e, i, o, n, r", per cui assomiglia spesso ad una parola comprensibile. Al contrario, la password generata dall'alfabeto sarà quasi sempre illeggibile. Si potrebbe dire che l'alfabeto contiene parole "più strane" di quelle che si ottengono estraendo lettere a caso da un dizionario, che è un altro modo di dire l'alfabeto contiene più informazione del dizionario.

Algoritmi di compressione

Il confronto tra dizionario e alfabeto può aiutare a capire il principio di funzionamento degli algoritmi di compressione. Consideriamo una messaggio costituito da sequenza di 8 caratteri, ottenuta sapendo solo che ciascun carattere assume uno dei 21 simboli dell'alfabeto, ovvero ignorando da quale sorgente è estratta la sequenza. Immaginiamo poi di dover trasmettere tale messaggio di 8 simboli su un canale informatico, come ad esempio una e-mail.

Siccome non sappiamo da quale sorgente è stata generata la sequenza, dobbiamo assumere che la sequenza contenga una qualsiasi combinazione dei 21 simboli ammessi, ovvero sia una delle 21^8 combinazioni possibili. Come visto sopra, per rappresentare tale informazione servono circa 35 bit. Supponiamo però di scoprire, in un secondo tempo, che la sequenza è prodotta da una sorgente di tipo dizionario, ovvero che ogni messaggio contiene solo 32 bit di informazione. Allora, prima della trasmissione, potrei “tagliare” il messaggio da 35 bit (“perdendo” 3 bit) e trasmettere lo stesso contenuto con soli 32 bit, senza perdere informazione. Un'operazione di questo tipo significa applicare un procedura, ovvero un algoritmo, che “comprime” il messaggio.

Vediamo un rozzo esempio di algoritmo di questo tipo:

- Assegno a tutte le possibili sequenze di 8 caratteri producibili dal dizionario un numero. Ogni “parola” del dizionario sarà perciò associata ad un numero compreso tra 1 e $4{,}3\cdot10^9$ (un numero a 32 bit)
- Quando ricevo il messaggio controllo se la sequenza appartiene al dizionario. Se la risposta è affermativa, allora la sequenza è associabile ad un numero a 32 bit. In questo caso trasmetto solamente 32 bit sul canale (anziché 35)
- In caso contrario, cioè se la sequenza *non* appartiene al dizionario, allora trasmetto tutti e 35 i bit della sequenza originale

Ovviamente un algoritmo del genere non è realistico, perché in realtà servirebbero dei codici di controllo, ad esempio per dire al ricevente se sto per trasmettere 32 o 35 bit. L'esempio spiega comunque il principio di funzionamento degli algoritmi di compressione: siccome *in media* il messaggio è più "scontato" del previsto, ovvero "poco imprevedibile", allora possiamo trasmettere *mediamente* meno dati sul canale, senza perdere informazione.

Prima di pensare che un tale algoritmo sia una specie di magia, ricordiamo che si tratta di una compressione **in media**: per alcuni messaggi serviranno 35 bit, per altri 32, per altri ancora forse addirittura meno. Ma se facciamo la media su tutti i messaggi inviati dopo la compressione, troveremo che *mediamente* servono solamente 32 bit per messaggio.

Compressione di immagini

Accenniamo infine al funzionamento degli algoritmi di compressione di immagini. Prendiamo un'ipotetica immagine quadrata avente dimensione di 10 pixel di lato (pari a 100 pixel totali).

Se ogni pixel contiene un colore, assumendo di usare una tavolozza di soli 256 colori, possiamo associare il contenuto di ogni pixel ad un numero tra 1 e 256. Ne segue che la nostra immagine è una griglia di 10x10 caselle, dove ogni casella contiene un simbolo numerico compreso tra 1 e 256. Questo formato è simile al formato *bitmap* (o "raster") perché di fatto stiamo descrivendo l'immagine attraverso la mappa di tutti i pixel che la costituiscono.

Altri formati, quali ad esempio GIF e PNG, sono invece formati compressi. Se una certa immagine occupa 10 KB di memoria in formato *raster*, potrebbe richiederne solo 1 o 2 KB una volta convertita in formato GIF o PNG. Per capire come ciò sia possibile, calcoliamo l'informazione ottenibile leggendo **1** pixel a caso dall'immagine descritta un formato di tipo *raster*, ovvero memorizzando il colore di tutti i cento pixel che la compongono:

$$I(x) = -\log_2(1/256) = 8 \text{ bit}$$

Siccome abbiamo cento pixel, per descrivere l'intera immagine serviranno 800 bit.

Supponiamo ora che il nostro algoritmo di compressione sia abbastanza intelligente da scoprire se esistono aree **uniformi** all'interno dell'immagine. Ad esempio, supponiamo che un intero quadrato di 4x4 pixel sia completamente arancione, e che il colore arancione sia associato al simbolo numero 42. Siccome l'area di tale quadrato è pari a 16 pixel, se scelgo di rappresentarlo trasmettendo il colore di tutti i suoi 16 pixel ottengo:

$$I(x) = 16 \cdot 8 \text{ bit} = 128 \text{ bit}$$

In alternativa il mio algoritmo potrebbe dire: "quando trovo un quadrato di colore uniforme, invece di trasmettere tutti i pixel, trasmetto solo le informazioni sul quadrato".

Ad esempio, potrei descrivere un area quadrata come segue:

- Coordinate del vertice in alto a sinistra del quadrato (due numeri tra 1 e 10)
- Dimensione del lato del quadrato (un numero tra 1 e 10)
- Colore del quadrato (un numero tra 1 e 256, in questo caso 42)

Andiamo a calcolare l'informazione trasportata da questi tre messaggi:

$I(x)|_{vertice}$ = 2 · I (un numero tra 1 e 10) = 2 · (- $\log_2(1/10)$) ≈ 6,6 bit

$I(x)|_{lato}$ = I (un numero tra 1 e 10) = - $\log_2(1/10)$ ≈ 3,3 bit

$I(x)|_{colore}$ = I (un numero tra 1 e 256) = - $\log_2(1/256)$ = 8 bit

Perciò, la quantità di informazione necessaria per "ricostruire" il quadrato è:

I(compressa) ≈ 6,6 + 3,3 + 8 ≈ 20 bit

di gran lunga inferiore ai 128 bit utilizzati quando si trasmette le informazione del quadrato di 4x4 pixel utilizzando il formato *raster*, ovvero tutte le informazioni dei 16 pixel.

In questo caso l'efficienza di compressione è:

Rapporto di compressione = 128 / 20 = 6,4

Che espresso in termini di "spazio risparmiato", è pari ad una compressione del 84%.

Ricordiamo che tale compressione va intesa **in media**. Ad esempio, se un algoritmo offrisse un rapporto di compressione del 50% e noi lo applicassimo per comprimere migliaia di immagini di 10 pixel per lato, si avrebbe che *mediamente* l'immagine compressa occuperebbe 64 bit anziché 128. Ciò significa che alcune immagini verrebbero compresse a solo un paio di bit (esempio: un'immagine tutta blu o tutta rossa) mentre altre immagini richiederebbero tutti e 128 bit anche dopo la compressione (esempio: la foto di una manciata di coriandoli). In ogni caso, calcolando la *media aritmetica* di tutte le immagini compresse dal nostro algoritmo, si dovrebbe trovare un valore prossimo a 64 bit.

Conclusioni

La Teoria dell'Informazione è molto vasta, ma le nozioni elementari affrontate in questo documento dovrebbero fornire gli strumenti essenziali per comprendere i principi fondamentali della teoria. In particolare, ricordiamo che l'informazione contenuta in un singolo messaggio, evento o misura è:

$I(x) = -\log_b(P_x)$ dove P_x è la probabilità dell'evento stesso

Inoltre, se la possibilità di ricevere un certo simbolo x è indipendente dalla probabilità di ricevere il simbolo y, allora vale:

$$I\ (x\ \text{AND}\ y) = I(x) + I(y)$$

mentre l'informazione ottenuta mediamente da una sorgente di informazioni è:

$$H = \langle I(x) \rangle = - \sum P_x \cdot \log_b(P_x)$$

Nella formulazione originale di Shannon questa quantità è chiamata **Entropia**. Finora abbiamo chiamato tale quantità "informazione media del sistema" per evitare di confondere il lettore, poiché spesso il concetto di entropia è associato alla termodinamica. La relazione tra l'entropia di Shannon e quella termodinamica è molto interessante, ma esula dallo scopo di questi appunti [5]. In ogni caso, per coerenza con la nomenclatura ufficiale, è bene utilizzare il termine *entropia* per identificare l'*informazione media* contenuta da una sorgente.

Abbiamo inoltre accennato a grandi linee il principio di funzionamento di un algoritmo di compressione. A tal riguardo osserviamo che il rapporto di compressione non è un indicatore di qualità dei dati, perché in alcuni casi è bene avere un elevato rapporto di compressione, mentre in altri casi è meglio averlo il più piccolo possibile: tutto dipende da ciò che vogliamo fare con le informazioni che stiamo maneggiando.

Esempio 1: devo sviluppare un nuovo algoritmo di compressione delle immagini e scopro che l'informazione trasmessa sul canale è pari a 100 bit, mentre l'entropia della sorgente (cioè la sua informazione media) è di soli 50 bit . Questa è una buona situazione, perché so per certo di avere un ampio margine di manovra. In altri termini: so che è teoricamente possibile comprimere il messaggio fino al 50% dello "spazio" attualmente utilizzato.

Esempio 2: devo controllare la robustezza di una password di 8 caratteri. Siccome so che l'entropia di una password prodotta da una sorgente di tipo "alfabeto" è pari a 35 bit, se l'entropia prodotta dalla password in esame risultasse di soli 20 bit allora avrei un problema, perché la mia attuale sorgente di password è peggiore di un dizionario (che produrrebbe comunque 32 bit di informazione). In questo caso dovrei sostituire la sorgente di password con una nuova sorgente: quanto più la nuova sorgente avrà entropia vicino a 35 bit - che è il massimo teoricamente ottenibile in questo contesto - tanto più efficiente sarà la nuova sorgente di informazioni.

Bibliografia

[1] Wikipedia, *Teoria dell'Informazione*, 2022
https://it.wikipedia.org/wiki/Teoria_dell%27informazione

[2] Wikipedia, *Entropia (teoria dell'informazione)*, 2022
https://it.wikipedia.org/wiki/Entropia_(teoria_dell%27informazione)

[3] Minnaja, Paccagnella*, Variazioni dell'entropia linguistica dell'italiano scritto e calcolo di un'entropia fonematica*, 1977
http://www.numdam.org/article/RSMUP_1977__57__247_0.pdf

[4] Wikipedia, *Analisi delle frequenze*, 2022
https://it.wikipedia.org/wiki/Analisi_delle_frequenze

[5] Stefano Adriani, *Informazione e Termodinamica*, 2022
http://adriani.altervista.org/author/notes.php