Informazione e Termodinamica

Entropia di Shannon ed Entropia termodinamica

Indice generale

Premessa

Questi appunti contengono esempi sul rapporto tra entropia di Shannon ed entropia termodinamica. Il testo assume che il lettore sia in possesso di alcune nozioni propedeutiche:

- Matematica: proprietà dei logaritmi, calcolo combinatorio
- Statistica: probabilità, variabile casuale discreta, valore atteso
- Fisica: concetto di gradi di libertà, spazio delle fasi, principi della termodinamica

Per un'introduzione alle nozioni fondamentali della teoria dell'informazione rimandiamo agli appunti sulla "Teoria dell'Informazione" [1].

Contesto termodinamico

In questi appunti andremo ad analizzare alcuni sistemi termodinamici dal punto di vista della teoria dell'informazione. Poiché tale teoria può essere applicata a qualsiasi sistema (un libro, un messaggio ecc.), analizzare i sistemi termodinamici significa concentrarsi su un caso **particolare** di sorgente di informazione. Di norma, quando si discute l'applicazione specifica di una teoria generale, è bene precisare il contesto. Quanto segue dovrebbe chiarire il lessico e la notazione utilizzati, senza pretendere di essere una definizione esaustiva.

Gradi di libertà

Quando si ha a che fare con dei sistemi termodinamici, la "griglia" delle configurazioni possibili è rappresentata dallo **spazio delle fasi**, che di solito è multi-dimensionale. Siccome è difficile rappresentare visivamente uno spazio in molte dimensioni, in questi appunti faremo riferimento a sistemi con solo **1 grado di libertà**. Ciò significa che gli elementi del sistema potranno collocarsi su una griglia unidimensionale, in pratica una fila di caselle distinte.

Indicheremo col numero **C** il numero di posizioni possibili sulla griglia (il numero di *caselle*) e con **N** il numero di *simboli* che il sistema è in grado di generare dal punto di vista della teoria dell'informazione (il numero di eventi distinti).

Vediamo alcuni esempi di semplici sistemi con un solo grado di libertà.

Esempio 1: consideriamo un sistema con C=2, ovvero una griglia di due caselle (vicine tra loro, rigide e fissate nello spazio). All'interno delle caselle può trovarsi una pallina nera. Tale sistema può assumere solo una di queste configurazioni:

Se osserviamo il contenuto di una casella possiamo ottenere solo due informazioni: la pallina è presente (1° esito possibile) oppure è assente (2° esito possibile). Perciò, fin tanto che ci limitiamo ad estrarre informazione da una singola casella, diremo che il sistema è in grado di generare 2 simboli (ad esempio "sì" e "no", oppure 1 e 0), ovvero che N = 2.

Per comodità, quando faremo riferimento a questo sistema, scriveremo:

$(C=2)|_{N2P1}$ ovvero "C = 2 caselle, N = 2 simboli e P = 1 pallina"

Esempio 2: consideriamo un sistema che contiene 1 pallina nera all'interno di una griglia di 4 caselle. Un sistema di questo tipo può trovarsi in quattro configurazioni possibili:

Se osserviamo il contenuto di una singola casella possiamo trovare solo due risultati: pallina presente o assente. Perciò, fintanto che otteniamo informazione osservando il contenuto di una *singola* casella, il sistema è equivalente a quello dell'esempio precedente, perché forniscono entrambi solo *due* simboli. In realtà, nonostante questa somiglianza, vedremo che i due sistemi differiscono per la quantità di informazione **totale** in essi contenuta.

Macrostato e microstato

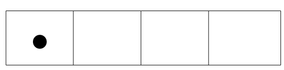
Definiamo macrostato un qualsiasi criterio descrittivo del sistema in termini di misure macroscopiche, ovvero funzioni di stato o indicatori statistici aggregati. In altre parole, qualsiasi descrizione del sistema che ignora la sua "configurazione interna".

Esempio: consideriamo il sistema $(C=8)|_{N2P2}$ (8 caselle, 2 simboli, 2 palline).
Alcuni macrostati di questo sistema potrebbero essere:

- *Macrostato A*: "le due palline nere sono in una posizione qualsiasi"
- *Macrostato B*: "Le due palline nere si trovano nelle prime quattro caselle"
- *Macrostato C*: "Le due palline nere *non* sono vicine tra loro"

La definizione di macrostato potrebbe sembrare un po' fumosa, ma diventa più chiara introducendo il concetto di microstato: un microstato è la conoscenza completa di **tutti** gli elementi del sistema, ovvero il contenuto di tutte le caselle.

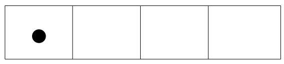
Esempio: se il sistema $(C=4)|_{N2P1}$ (4 caselle, 2 simboli, 1 pallina) si trova in questa configurazione:



allora diciamo che il sistema di trova nel *microstato* "pallina nera in prima casella". Infatti, siccome abbiamo solo una pallina, in questo caso tale informazione determina *completamente* la configurazione del sistema.

Mentre il macrostato fornisce una descrizione di "alto livello" del sistema, il microstato ne rappresenta una descrizione di "basso livello", esaustiva e completa. In alcuni casi le due definizioni sembrano sovrapporsi. Ad esempio, nel caso del sistema appena visto (4 caselle e 1 pallina nera), si potrebbe dire che:

- *Macrostato A*: "*tutte* le palline nere sono in prima posizione"
- *Microstato b*: "*La* pallina nera è in prima posizione"



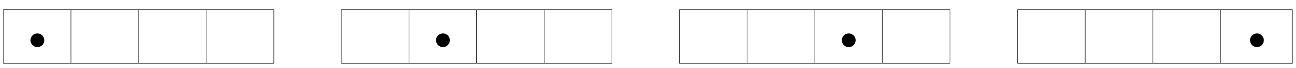
In questo caso è corretto dire che se il sistema di trova nel macrostato A, allora si trova *anche* nel microstato b. Questa distinzione sembra un **sofismo lessicale**, perché quando parliamo di macrostato usiamo termini generici ("*tutte* le palline nere sono ...") mentre quando parliamo di microstato siamo più specifici ("*la* pallina nera è …"). In questo caso, però, macrostato A e microstato b identificano la medesima configurazione del sistema.

Si tratta in realtà di un'eccezione, perché il sistema è talmente semplice che il macrostato A "coincide" con il microstato b. E' comunque importante esprimersi in modo corretto, ovvero dire che "il macrostato A è formalmente diverso dal microstato b, anche se essi identificano la medesima configurazione".

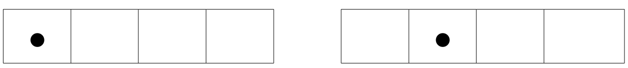
Per comprendere meglio la relazione tra macrostato e microstato consideriamo di nuovo il sistema $(C=4)|_{N2P1}$ (4 caselle, 2 simboli, 1 pallina) e due macrostati:

- *Macrostato Q*: "La pallina nera è in una posizione qualsiasi"
- *Macrostato D*: "La pallina nera è in posizione 1 o 2"

In questo caso il macrostato Q (come "**q**uattro") permette **4** possibili microstati:



Mentre il macrostato D (come "**d**ue") permette solo **2** possibili microstati:



Per definizione, se il sistema si trova in un certo macrostato, l'insieme di tutte le configurazioni di basso livello corrispondenti a quel macrostato sono detti microstati **equivalenti** tra loro. Ciò significa che un sistema può restare nello stesso macrostato pur assumendo configurazioni “interne” diverse: tutte le configurazioni che soddisfano una certa definizione di macrostato sono dette *microstati associati a quel macrostato*.

Se il macrostato è noto, allora conosciamo tutti i possibili microstati associati.
Il viceversa però non è vero: se conosciamo un microstato,
non sappiamo in quale macrostato si trova il sistema

Infatti, per identificare il macrostato occorre osservare il sistema per un tempo sufficiente a misurare la "configurazione media” nel tempo, in modo da ottenere informazioni macroscopiche. Da questo punto di vista il macrostato può essere pensato come la "media" dei suoi microstati: ecco perché, anche se è noto in quale microstato si trova il sistema, in generale non possiamo usare questa conoscenza per dedurre il suo macrostato.

Esempio: se il sistema $(C=4)|_{N2P1}$ (vedi sopra) si trova in questo stato:

●			

non possiamo sapere se ci troviamo nel macrostato Q o D, in quanto entrambi prevedono questo particolare microstato tra quelli ammessi. Se però ripetiamo migliaia di osservazioni e troviamo sempre e solamente questi due tipi di microstato:

●			

	●		

allora possiamo ragionevolmente supporre che il sistema si trovi nel macrostato D.

Probabilità dei microstati

Nella teoria della termodinamica si assume che:

I microstati di un sistema in equilibrio (macrostato fissato) sono equiprobabili tra loro

condizione non richiesta nell'ambito della teoria di Shannon. Assumere che i microstati siano tutti **equiprobabili** serve principalmente a sottolineare che non è possibile identificare (o "indovinare") il microstato a partire dalla conoscenza del macrostato, perché nessun microstato è più frequente degli altri.

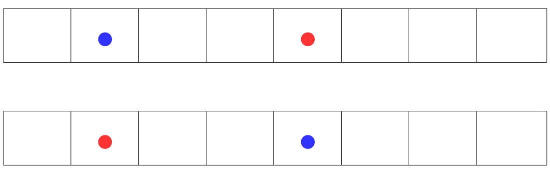
Possiamo pensare ad un sistema termodinamico come ad una sorta di "gif animata" in perenne cambiamento. Se tutte le possibili immagini sono equiprobabili, allora l'animazione della "gif" sarà un guazzabuglio di puntini in continuo mutamento. Se l'animazione è abbastanza veloce, come avviene nei sistemi termodinamici, non riusciremo a scorgere l'accendersi e spegnersi dei singoli pixel della "gif", ma vedremo un'unica tonalità di grigio. La particolare tonalità di grigio percepita rappresenta il macrostato del sistema, mentre l'immagine "fotografata" in un preciso istante rappresenta il microstato. Perciò, anche se ad ogni istante il sistema "salta" da un microstato ad un altro, a livello macroscopico non si vede alcun cambiamento: l'immagine rimane sempre della stessa tonalità di grigio (i.e. lo stesso macrostato).

Affermare che tutti i microstati siano equiprobabili implica che anche i microstati più "strani" siano possibili. Ad esempio, si potrebbe pensare che un gas in equilibrio (valori costanti di pressione, temperatura e volume) possa trovarsi per un breve istante tutto racchiuso in una zona piccolissima del recipiente, oppure ad una temperatura diversa dall'equilibrio. Questo non è un problema, perché le grandezze termodinamiche sono espresse come **media** nello spazio e nel tempo, per cui dal punto di vista macroscopico ogni fluttuazione microscopica "si cancella" quando si estraggono informazioni aggregate sull'insieme, come nel caso della gif animata.

Particelle indistinguibili

Un altro postulato della termodinamica richiede che le particelle di un sistema, ovvero le "palline" che compongono il sistema, siano **indistinguibili** tra loro. Ciò non significa che gli elementi del sistema siano perfettamente identici, ma solamente che essi non siano distinguibili dal punto di vista fisico, ovvero all'osservatore.

Esempio: immaginiamo che le palline del sistema $(C=8)|_{N2P2}$ (8 caselle, 2 simboli, 2 palline) abbiano colore diverso, ma di non essere in grado di percepirne il colore. Se ai nostri occhi le palline appaiono tutte nere, allora le seguenti configurazioni identificano lo *stesso* identico microstato del sistema:



Alcuni chiarimenti sul principio delle "palline indistinguibili":

- Il concetto riguarda solo gli aspetti *fisici* del sistema, non gli ipotetici ragionamenti astratti. Ad esempio, abbiamo appena *immaginato* di poter colorare le due palline, e così facendo abbiamo ipotizzato un modo *fittizio* di poterle distinguere. Le palline restano comunque indistinguibili dal punto di vista *fisico*
- Il principio si applica solo a insiemi di palline omogenee. Ad esempio, se considerassimo un sistema composto da palline fisicamente eterogenee (ad esempio: 10 palline pesanti e 10 palline leggere) allora il concetto di “palline indistinguibili” significherebbe solo che le palline pesanti sono indistinguibili di tra loro (idem per le palline leggere)

Metodo di misura dell'informazione

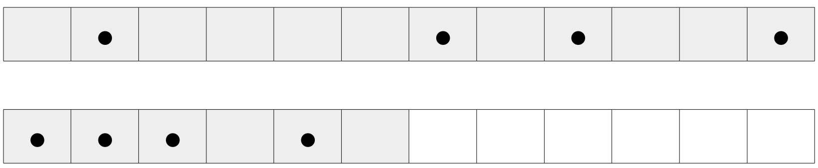
Per trattare un sistema termodinamico come sorgente di informazioni dobbiamo definire un modo **operativo** di estrarre conoscenza da esso. Nella teoria dell'informazione ciò non è necessario, perché l'approccio è astratto: per applicare la formula di Shannon basta sapere che una sorgente emette una serie di simboli casuali (non necessariamente equiprobabili) senza preoccuparsi del metodo di misura.

Le cose sono diverse nel caso dei sistemi termodinamici. Ad esempio, per contestualizzare i nostri esempi, dobbiamo precisare che:

- Misurare un sistema termodinamico non significa estrarre la pallina dal sistema, come avviene nel gioco del lotto, ma semplicemente “guardare” cosa si trova in una casella
- Il sistema “salta” continuamente da un microstato all’altro, per cui, anche guardando due volte di seguito nella stessa casella, in genere si ottiene un risultato diverso

Esempio: consideriamo un sistema del tipo $(C=12)|_{N2P4}$ (12 caselle, 2 simboli, 4 palline) e due sue possibili definizioni di macrostato:

- Tutte le palline sono distribuite a caso all'interno della nostra "griglia" (stato aeriforme)
- Tutte le palline sono su un lato della griglia (stato liquido)



(stato aeriforme)

(stato liquido)

dove le caselle grigie rappresentano le posizioni occupabili dalle palline nere: nel macrostato "aeriforme" le particelle del sistema possono occupare una casella qualsiasi, mentre nel macrostato "liquido" le particelle sono tutte contenute entro le prime sei caselle.

Consideriamo adesso il seguente criterio di misura: "per estrarre informazione, guardo se c'è una pallina nella prima casella". Si ha quindi:

$I(\text{aeriforme}) = -\log_2(4/12) \approx 1{,}6$ bit
$I(\text{liquido}) = -\log_2(4/6) \approx 0{,}58$ bit

che sono le quantità di informazioni ottenibile scegliendo di pescare informazione della *prima* casella della griglia. Se invece scegliessimo di guardare nell'*ultima* casella avremmo:

$I(\text{aeriforme}) = -\log_2(4/12) \approx 1{,}6$ bit
$I(\text{liquido}) = -\log_2(0/12) = ?$

Questo perché nello stato “liquido” l'ultima casella del sistema è sempre vuota, per cui è impossibile trovarvi una pallina. Secondo la definizione di Shannon gli eventi impossibili hanno informazione infinita, quindi vanno ignorati. Perciò, se raccogliamo informazioni dall'ultima casella della griglia, il sistema nello stato liquido **non** è una sorgente di informazione, perché **non** produce informazione. L'esempio suggerisce che – nel caso dei sistemi termodinamici – dobbiamo ragionare come segue:

Come criterio di misurazione è corretto osservare il contenuto di una qualsiasi casella della griglia (un punto nello spazio delle fasi) fintanto che il sistema è uniforme.
Al contrario, se un sistema è caratterizzato da aree omogenee ma diverse tra loro, allora l'informazione ottenuta dipende in genere dal criterio di misura.

Ovviamente esistono infiniti modi di estrarre informazione da un sistema termodinamico. Ad esempio potrei scegliere una particella a caso e considerare come informazione la sua posizione, oppure potrei scegliere una coordinata a caso e osservare di volta in volta il contenuto di una casella diversa. Come vedremo più avanti, se il criterio di misurazione raccoglie **tutta** l'informazione del sistema, allora il risultato è lo stesso per qualsiasi metodo di misura. Al contrario, se il criterio di misura raccoglie solo informazioni parziali, la quantità di informazione ottenuta dipende dal particolare metodo di misura.

Informazione totale di un sistema

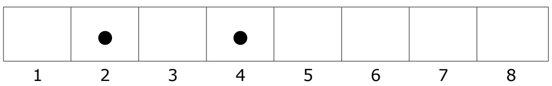
Negli esempi precedenti abbiamo anticipato che, se raccogliamo tutta l'informazione contenuta in un sistema, allora il risultato non dipende dal metodo di misura. Ma quanto vale l'informazione **totale** contenuta nel sistema? Ovvero: quanta informazione devo “estrarre” da un sistema per identificare il suo microstato? Ma soprattutto, siamo sicuri che tale informazione non dipenda dal modo in cui otteniamo informazioni sul sistema?

Per rispondere a queste domande, consideriamo il seguente esempio.

Esempio: prendiamo il sistema $(C=8)|_{N2P2}$ (8 caselle, 2 simboli, 2 palline). In questo caso, per estrarre *tutta* l'informazione del sistema potrei operare in modi diversi:

- Criterio 1: osservo contemporaneamente il contenuto di tutte e 8 le caselle
- Criterio 2: osservo contemporaneamente la posizione delle due palline

Ipotizzando che la misura venga eseguita da un dispositivo elettronico è lecito ipotizzare che l'esito delle osservazioni sia esprimibile mediante un numero, ad esempio in notazione decimale o binaria. Se il sistema di trovasse in questo microstato:



Allora, applicando i criteri definiti sopra otterrei due diverse rappresentazioni del sistema:

- Criterio 1: guardo nelle 8 caselle e dico che il sistema ha fornito il simbolo 01010000
- Criterio 2: guardo dove sono le 2 palline e dico il sistema ha fornito i simboli 2 e 4

Nel primo caso il sistema sembra fornire un numero di 8 bit, ovvero un valore tra 1 e 256. Nel secondo caso il sistema invece sembra produrre due numeri tra 1 e 8, per cui si avrebbero $8 \cdot 8 = 64$ esiti possibili. Quale dei due valori è l'informazione totale contenuta nel sistema?

Nel primo caso l'informazione associata all'osservazione di una singola casella è:

$$I(x) = -\log_2(2/8) = 2 \text{ bit}$$

Per cui, quando guardiamo il contenuto delle otto caselle, si potrebbe pensare di ottenere $8 \cdot 2 = 16$ bit di informazione. Ma questo risultato è sbagliato, perché l'informazione prodotta da eventi distinti si può sommare solo se gli eventi sono **indipendenti** tra loro.

Chiariamo questo punto, perché è fondamentale:

- E' *corretto* dire che l'informazione associata alla misura di una <u>singola</u> casella è 2 bit
- E' *corretto* dire che se misuro 8 volte il contenuto della <u>stessa</u> casella ottengo 16 bit
- E' *sbagliato* dire che se misuro contemporaneamente 8 caselle <u>diverse</u> ottengo 16 bit

L'errore è dovuto al fatto che nel momento in cui osservo il contenuto di una casella, altero la probabilità delle altre misurazioni. Ad esempio, dopo aver esaminato il contenuto delle prime sette caselle il contenuto dell'ottava casella non è affatto casuale, ma è completamente determinato.

Consideriamo il primo criterio e ipotizziamo che il sistema si trova in un preciso microstato (ad esempio "01010000", come in figura precedente): stimiamo la probabilità di tale evento. Per farlo calcoliamo il numero di esiti possibili, ovvero le **permutazioni con ripetizioni** di 8 elementi di cui due si ripetono (la pallina nera si ripete 2 volte, la casella vuota 6 volte):

$$N_{esiti} = n! / (k_1!\cdot k_2!) = 8! / (2!\cdot 6!) = 7\cdot 8 / 2 = 28$$

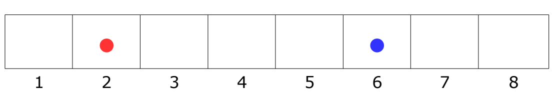
Per cui un evento del tipo "x = 01010000" è solo una delle 28 combinazioni possibili. Perciò, se scegliamo come simbolo un numero da 1 a 28, esso descrive completamente lo stato del sistema. L'informazione associata ad un generico simbolo x di questo tipo è:

$$I(x) = -\log_2(1/28) \approx 4{,}8 \text{ bit}$$

quindi, per conoscere **completamente** lo stato del sistema è necessario estrarre mediamente circa 4,8 bit di informazione. Notiamo che, siccome le 28 combinazioni sono tutte equiprobabili, quando calcoliamo l'informazione media del sistema (ovvero l'entropia di Shannon) otteniamo:

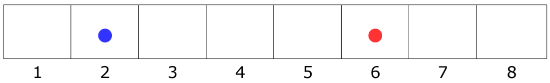
$$H = \langle I(x) \rangle \approx 4{,}8 \text{ bit}$$

Vediamo adesso qual è il modo giusto di ottenere questo risultato dal secondo criterio. Quando misuro la posizione della prima pallina devo rappresentare l'informazione ottenuta come un numero **da 1 ad 8**. Quando invece misuro la posizione della seconda basta un numero **da 1 a 7**, perché le due palline non possono occupare la stessa casella. Una "codifica" di questo tipo è più chiara se indichiamo le due palline con colori diversi:



Allora, invece che descrivere la configurazione con la coppia “2;6” potremmo dire che la pallina rossa si trova in posizione 2 e la pallina blu si trova “4 posti più a destra”. Questo precisando che il termine “più a destra” nel caso dell'ultima casella significa riprendere a contare dalla prima casella, come se le estremità della griglia fossero adiacenti. Secondo tale codifica la configurazione “x=01000010” si potrebbe descrivere come “x=2;4”.

Osserviamo però che tale codifica richiede di distinguere formalmente le due palline, perché dobbiamo sapere a quale pallina fa riferimento il *primo* numero (rossa) e a quale pallina fa riferimento il *secondo* numero (blu). Ad esempio, la misura “x=6;4” descriverebbe la configurazione:



Siccome dal punto di vista fisico le palline sono **indistinguibili** tra loro, questa configurazione è la stessa configurazione della codifica “2;4”, ovvero: abbiamo due codifiche diverse (“6;4” e “2;4”) per la stessa configurazione. Perciò il numero di configurazioni differenti in cui può trovarsi il sistema è in realtà la **metà** di quello fornito dalla codifica, ovvero:

$$N_{esiti} = 7 \cdot 8 \, / \, 2 = 28$$

da cui ottiene nuovamente:

$$I(x) = -\log_2(1/28) \approx 4,8 \text{ bit} \qquad \text{e} \qquad H = \langle I(x) \rangle \approx 4,8 \text{ bit}$$

Possiamo quindi dire che:

L'informazione totale di un sistema termodinamico omogeneo non dipende dal modo in cui viene raccolta l'informazione. E' perciò legittimo scegliere il modo più comodo (o intuitivo) per estrarre informazioni dal sistema (ad esempio per semplificare i calcoli)

In particolare, nel caso dei sistemi termodinamici qui trattati (un solo grado di libertà e nessuna informazione sulla velocità delle particelle) possiamo calcolare l'informazione totale sia guardando il contenuto di tutte le caselle, sia osservando la posizione di tutte le palline, sia contando il numero di microstati che può assumere il sistema. In tutti i casi si dovrebbe ottenere lo stesso risultato, ovvero l'informazione necessaria per conoscere completamente il sistema, cioè:

Nel caso di un sistema termodinamico, se scegliamo un criterio di misura in grado di descrivere completamente il sistema, allora l'entropia di Shannon rappresenta sia la quantità di informazione media che possiamo estrarre da una singola misura, sia la quantità di informazione totale contenuta nel sistema

E' importante sottolineare che questo risultato è valido solo se il criterio di misura descrivere **completamente** il sistema. Se ciò non è vero, allora informazione media e informazione totale sono in genere diverse.

Compressione dell'informazione

Negli esempi precedenti abbiamo trovato risultati apparentemente in contraddizione tra loro. Anche se l'argomento esula dai fini di questa trattazione, è interessante osservare che situazioni di questo tipo hanno spesso a che fare con la *compressione* dei dati.

Esempio: nel caso del sistema $(C=8)|_{N2P2}$ (8 caselle, 2 simboli, 2 palline) abbiamo trovato risultati dipendenti dal criterio di misura. In particolare i due criteri:

- Criterio 1: dopo 8 misure il sistema fornisce il simbolo 01010000
- Criterio 2: dopo 8 misure il sistema fornisce i valori 2 e 4

identificavano entrambi una configurazione del tipo:

	●		●				

Abbiamo poi tentato di quantificare l'informazione totale del sistema in modo "ingenuo", semplicemente contando il numero di bit necessari per rappresentare i due esiti possibili:

- Criterio 1: fornisce una serie di 8 bit → 8 bit di informazione?
- Criterio 2: fornisce due numeri da 1 ad 8 → 64 simboli → 6 bit di informazione?

Ragionando in modo corretto abbiamo però ottenuto che l'informazione totale di questo sistema è pari a soli **4,8 bit**. Questo ci dice che le codifiche associate ai due criteri di misurazione non sono ottimali, ovvero che è teoricamente possibile *comprimere* i dati per ridurre lo "spazio" richiesto dall'informazione. Ma quanto possiamo comprimere?

La risposta emerge proprio dalla conoscenza dell'**informazione totale**: siccome il sistema contiene intrinsecamente 4,8 bit di informazione, nessuna codifica potrà descrivere completamente il sistema con meno di 4,8 bit. Questo permette di quantificare *quanto* i due criteri siano ridondanti:

- Criterio 1: richiede 8 bit invece di 4,8 → ridondante al 166% (pessimo)
- Criterio 2: richiede 6 bit invece di 4,8 → ridondante al 125% (mediocre)

Esempi di sistemi semplici

Avendo chiarito il contesto e le modalità che permettono di trattare un sistema termodinamico come una sorgente di informazione, possiamo calcolare l'informazione totale associata ad alcuni semplici sistemi termodinamici.

Una pallina in quattro caselle

Consideriamo il sistema $C=4|_{N2P1}$ (4 caselle, 2 simboli, 1 pallina), che permette quattro microstati possibili:

Se scegliamo come criterio di misura l'osservazione di una singola casella abbiamo:

I(Nera) = $- \log_2(1/4) = 2$ bit
I(Vuota) = $- \log_2(3/4) \approx 0{,}42$ bit

da cui l'informazione media:

H (prima casella) = $\langle I(x) \rangle = (1/4)\cdot 2 + (3/4)\cdot 0{,}42 \approx 0{,}81$ bit

Calcoliamo adesso l'informazione **totale** contenuta dal sistema, ovvero la quantità di conoscenza necessaria per identificare il microstato del sistema. Siccome abbiamo 4 microstati possibili, per conoscere la configurazione del sistema possiamo associare un numero da 1 a 4 a ciascun microstato. Conoscere la configurazione del sistema significa perciò sapere quale di questi quattro simboli è stato estratto, ovvero (x = 1,2,3 o 4):

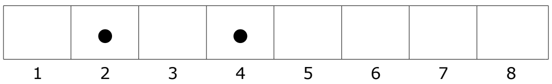
$I(x) = -\log_2(1/4) = 2$ bit

essendo i microstati tutti equiprobabili si ottiene:

$H(totale) = \langle I(x) \rangle = 2$ bit

Due palline in otto caselle

Consideriamo ora il sistema $C=8|_{N2P1}$ (8 caselle, 2 simboli, 2 palline), che permette diverse combinazioni, tra cui ad esempio:



Come visto in precedenza, il modo più semplice di enumerare le configurazioni ammesse consiste nell'osservare che i microstati possibili sono pari al numero di permutazioni con ripetizioni, dove la pallina nera si ripete 2 volte, la casella bianca 6 volte, ovvero:

$N_{microstati} = n! / (k_1!\cdot k_2!) = 8! / (2!\cdot 6!) = 7\cdot 8 / 2 = 28$

Mettiamo un attimo da parte questo risultato e calcoliamo l'informazione media ottenibile misurando il contenuto di una singola casella della griglia:

$I(Nera) = -\log_2(2/8) = 2$ bit
$I(Vuota) = -\log_2(6/8) \approx 0{,}42$ bit

da cui l'informazione media:

H (prima casella) $= \langle I(x) \rangle = (1/4)\cdot 2 + (3/4)\cdot 0{,}42 \approx 0{,}81$ bit

Che è esattamente lo **stesso** valore dell'esempio precedente, ovvero del sistema $C=4|_{N2P1}$ (4 caselle, 2 simboli, 1 pallina), quando guardavano solo il contenuto della prima casella.

Questo risultato sembra strano, perché si potrebbe pensare che ad un maggior numero di microstati dovrebbe corrispondere una maggiore entropia. Ricordiamo però che l'entropia di Shannon dipende da **come** si sceglie di estrarre informazione dal sistema. Ecco perché – a seconda del criterio di misura – possiamo ottenere la stessa quantità di informazione da sistemi diversi tra loro. Il confronto tra sistemi *non* va fatto comparando l'informazione ottenibile attraverso un arbitrario criterio di misura, ma bensì confrontando le quantità di informazione **totali** contenute nei sistemi.

Chiariamo tale affermazione. Nel caso del sistema $C=4|_{N2P1}$ si aveva:

$$H(\text{totale}) = 2 \text{ bit}$$

mentre per il sistema $C=8|_{N2P1}$ abbiamo (se x è un simbolo tra 1 e 28):

$$I(x) = -\log_2(1/28) \approx 4{,}80 \text{ bit} \qquad \rightarrow \qquad H(\text{totale}) = \langle I(x) \rangle \approx 4{,}80 \text{ bit}$$

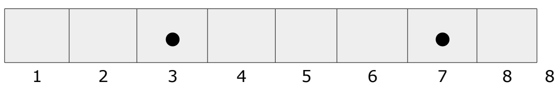
Per cui l'informazione totale contenuta dal sistema con 8 palline è *maggiore* di quella totale contenuta dal sistema di 4 palline (com'era logico aspettarsi). Questo conferma che:

Se il criterio di misura descrive completamente un sistema,
allora l'entropia di Shannon rappresenta l'informazione totale del sistema

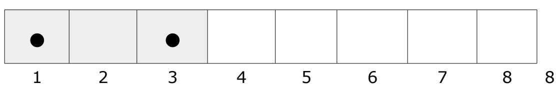
Sistemi non omogenei

Consideriamo il sistema $C=8|_{N2P1}$ (8 caselle, 2 simboli, 2 palline) dell'esempio precedente e due suoi possibili macrostati:

- Aeriforme: le 2 palline sono distribuite a caso tra le 8 caselle
- Liquido: le 2 palline sono confinate entro le prime 3 caselle del sistema



(stato aeriforme)



(stato liquido)

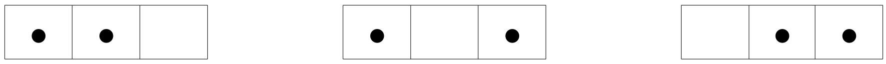
Dove le caselle grigie rappresentano le posizioni ammesse per le palline. Se estraiamo informazioni dallo stato liquido secondo il criterio “guardo nella prima casella” abbiamo:

$$I(\text{Nera}) = -\log_2(2/3) \approx 0{,}58 \text{ bit}$$
$$I(\text{Vuota}) = -\log_2(1/3) \approx 1{,}58 \text{ bit}$$

da cui l'informazione media:

$$H = <I(x)> = (2/3)\cdot 0{,}58 + (1/3)\cdot 1{,}58 \approx 0{,}91 \text{ bit}$$

Se invece volessimo conoscere interamente lo stato del sistema, sapendo che le due palline si trovano nelle prime tre caselle, abbiamo solamente tre configurazioni possibili:



per cui:

$$I(x = \text{una delle tre configurazioni}) = -\log_2(1/3) \approx 1{,}58 \text{ bit}$$
$$H = <I(x)> \approx 1{,}58 \text{ bit}$$

Proviamo adesso a confrontare l'informazione media associata ai due diversi criteri di misura, per i due macrostati presi in esame:

Criterio di misura	**Stato aeriforme**	**Stato liquido**
Guardo nella 1° casella (informazione parziale)	$I_X = -\log_2(2/8) = 2$ bit $H_{prima} = <I> = 2$ bit	$I_N = -\log_2(2/3) \approx 0{,}58$ bit $I_V = -\log_2(1/3) \approx 1{,}58$ bit $H_{prima} = <I> \approx 0{,}91$ bit
Guardo tutte le caselle (informazione totale)	$I_X = -\log_2(1/28) \approx 4{,}80$ bit $H_{TOT} = <I> \approx 4{,}80$ bit	$I_X = -\log_2(1/3) \approx 1{,}58$ bit $H_{TOT} = <I> \approx 1{,}58$ bit

Dove $I_N = I(\text{Nera})$ e $I_V = I(\text{Vuota})$. Si ha quindi che:

- Se guardo nella prima casella → lo stato aeriforme fornisce una variabile casuale con entropia mediamente *maggiore* rispetto allo stato liquido (2 è maggiore di 0,91), ovvero lo stato aeriforme produce circa il "doppio" di informazione di quello liquido
- Se misuro l'informazione totale → si ha un risultato simile (4,80 è maggiore di 1,58), ma stavolta lo stato aeriforme produce circa il "triplo" di informazione di quello liquido

L'esempio suggerisce che "di solito" un sistema con maggiore entropia produce più informazione parziale di uno a bassa entropia, indipendentemente dal metodo di misura.

Le virgolette attorno alla parola "di solito" sono importanti, perché l'informazione parziale dipende in realtà dal metodo di misura. Ad esempio, se decidessimo come criterio di misura "guardo il contenuto dell'ottava casella", allora per il sistema nello stato liquido si avrebbe mediamente I=0 (perché l'ottava casella è sempre vuota) per cui si potrebbe pensare che il sistema nello stato liquido è privo di informazione. Ecco perché, prima di fare confronti sull'informazione *parziale* ottenibile da un sistema, occorre sempre tenere conto del criterio di raccolta dell'informazione stessa.

L'esempio ci fornisce anche un altro risultato: quando misuriamo l'entropia totale del sistema nello stato liquido, le caselle associate alle posizioni 4,5,6,7,8 non forniscono alcuna informazione, per cui sono del tutto ininfluenti ai fini dei calcoli. In altre parole:

$$H_{TOT} = \langle I(\text{liquido}) \rangle = H(\text{guardo le prime 3caselle}) + H(\text{guardo le ultime 5 caselle})$$

perché l'informazione media estraibile dalle ultime 5 caselle è nulla. Questo suggerisce che:

Se un sistema termodinamico non è omogeneo, la sua informazione totale è pari alla somma delle quantità di informazione totale associate alle zone omogenee

che ci permette di calcolare l'informazione totale di un qualsiasi sistema termodinamico, anche nel caso che si trovi in uno stato non omogeneo.

Esempio 1: l'informazione totale estraibile da un sistema costituito da un cubetto di ghiaccio in un bicchiere d'acqua è pari all'informazione totale del cubetto di ghiaccio più l'informazione totale dell'acqua restante.

Esempio 2: se un cubetto di ghiaccio si trova sul fondo di un recipiente sotto vuoto e sigillato, e se trascuriamo la presenza di molecole d'acqua allo stato aeriforme (prodotte per sublimazione), allora l'informazione associata al cubetto di ghiaccio non dipende dalle dimensioni del recipiente. In altre parole: qualsiasi siano le dimensioni della griglia in cui "immaginiamo" trovarsi il sistema, si ottiene sempre la stessa quantità di informazione totale.

Due palline distinguibili

Consideriamo il sistema $C=4|_{N3P2}$ (4 caselle, 3 simboli, 2 palline), dove le due palline (nera e bianca) sono fisicamente distinguibili tra loro. In questo caso si hanno 12 microstati possibili:

●	o				o	●				o		●			o			●
●		o				●	o				o	●				o		●
●			o			●		o				●	o				o	●

Verifichiamo questo risultato osservando che ogni casella può contenere solo tre valori (bianco, nero o casella vuota), per cui la casella vuota può essere considerata il terzo "simbolo". Il numero di configurazioni possibili è perciò pari alle permutazioni di 4 elementi (le 4 caselle) con un elemento che si ripete due volte (la casella vuota):

$N = 4! / (2!) = 24 / 2 = 12$

In questo caso, se scegliamo di raccogliere informazioni osservando il contenuto di una singola casella, abbiamo 3 esiti possibili:

$I(\text{Bianca}) = -\log_2(1/4) = 2$ bit
$I(\text{Nera}) = -\log_2(1/4) = 2$ bit
$I(\text{Vuota}) = -\log_2(1/2) = 1$ bit

Da cui l'informazione media (entropia di Shannon):

$$H(C=4|_{N3P2}) = (1/4)\cdot 2 + (1/4)\cdot 2 + (1/2)\cdot 1 = ½ + ½ + ½ = 2{,}5 \text{ bit}$$

Proviamo adesso a cambiare la definizione operativa di misura: invece di guardare il contenuto della prima casella della griglia, estraiamo un'informazione del tipo "dove si trova la pallina nera?". Siccome la pallina si può trovare in una qualsiasi delle 4 caselle (con la stessa probabilità per ogni casella), per ciascun possibile esito (x = 1,2,3 o 4) si ha:

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

Da cui:

$$H\ (C=4|_{N3P2}) = (1/4)\cdot 2 + (1/4)\cdot 2 + (1/4)\cdot 2 + (1/4)\cdot 2 = ½ + ½ + ½ + ½ = 4\ \text{bit}$$

Adottando questo nuovo criterio di osservazione, ogni singola misura produce *più informazione* del criterio precedente. Ciò è corretto, perché adesso abbiamo 4 simboli (1,2,3 o 4) mentre prima si avevano solo 3 simboli (bianco, nero o niente).

L'esempio conferma quanto detto in precedenza, ovvero che in caso di misura parziale dell'informazione, per i sistemi termodinamici è necessario spiegare **come** viene raccolta l'informazione. E' bene essere sempre consapevoli di questo aspetto, per evitare di giungere a conclusioni errate. Ad esempio sarebbe errato dire che il sistema fornisce in media "tra i 2,5 e i 4 bit" di informazione, perché ciò lascerebbe pensare che l'informazione media cambi nel tempo. In realtà, una volta fissato il criterio per ottenere informazioni sul sistema, l'informazione media ottenuta è sempre la stessa. In generale si sono solo due modi di cambiare l'informazione parziale emessa da un sistema termodinamico:

- Il sistema cambia macrostato
- L'osservatore cambia criterio di misura

Entropia di Shannon e termodinamica

Dopo aver definito il contesto e visto come calcolare l'informazione associata ad alcuni semplici sistemi termodinamici, possiamo finalmente comprendere la relazione tra entropia di Shannon e entropia termodinamica.

Le formule di questi due tipi di entropia sono:

$\mathbf{H} = \langle I(x) \rangle = -\sum P_x \cdot \log_b(P_x)$ Entropia di Shannon (informazione)

$\mathbf{S} = k \cdot \ln(W)$ Entropia di Boltzamn (termodinamica)

Dove la prima sommatoria va intesa sugli **N** simboli possibili, **W** è il numero di microstati ammessi per il sistema (in quel macrostato) e **k** la costante di Boltzamn.

Abbiamo però detto che esiste un particolare tipo di entropia di Shannon: quella che misura la quantità **totale** di informazione contenuta in un sistema, ovvero la quantità di informazione necessaria per descriverlo *completamente*:

$$H_{TOT} = \langle I(x) \rangle = - \sum P_x \cdot \log_b(P_x)$$

dove (in questo caso) P_x rappresenta la probabilità di ottenere un simbolo capace di descrivere completamente il sistema. Abbiamo anche visto che il modo più semplice di calcolare questa quantità consiste nell'immaginare di poter associare un simbolo diverso ad ogni possibile microstato del sistema, ovvero porre N=W. Siccome tutti i microstati sono equiprobabili, ovvero hanno tutti probabilità $\mathbf{P_w}$ di essere osservati, si ottiene:

$$H_{TOT} = \langle I(x) \rangle = - \sum P_w \cdot \log_b(P_w) = - \log_b(P_w)$$

dove la probabilità P_w è pari ad 1/W da cui:

$$H_{TOT} = - \log_b(P_w) = - \log_b(1/W) = \log_b(W)$$

ovvero

$$H_{TOT} = \log_b(W)$$

Sfruttando le proprietà dei logaritmi e ricordando che il logaritmo naturale *ln* è la scrittura alternativa di $\log_e$, possiamo scrivere:

$$H_{TOT} = \log_b(W) = \log_e(W) / \log_e(b) = \ln(W) / \ln(b)$$

da cui si ottiene la relazione tra entropia di Shannon ed entropia termodinamica:

$$\mathbf{S} = k \cdot \ln(W) = k \cdot \ln(b) \cdot (\ln(W) / \ln(b)) = \mathbf{k \cdot \ln(b) \cdot H_{TOT}}$$

dove **b** è la base utilizzata per calcolare l'informazione secondo Shannon.

La teoria dell'informazione si dice che la base b è arbitraria, ovvero che scegliere il valore di b corrisponde a fissare “l'unità di misura” dell'entropia di Shannon (le virgolette ci ricordano che in realtà l'informazione è adimensionale) [3]. Ad esempio, se esprimiamo la relazione tra entropia di Shannon e entropia termodinamica in base alle scelte di b più popolari si trova:

$\mathbf{S} = k \cdot \ln(e) \cdot H_{TOT} = \mathbf{k \cdot H_{TOT}}$ se b = e (H misurata in "nat")

$\mathbf{S} = k \cdot \ln(2) \cdot H_{TOT} \approx \mathbf{0{,}69 \cdot k \cdot H_{TOT}}$ se b = 2 (H misurata in "bit")

Possiamo quindi concludere che l'entropia termodinamica è un **caso particolare** di quella di Shannon, perché le due grandezze coincidono (a meno dell'unità di misura) per i sistemi termodinamici che soddisfano le seguenti proprietà:

- Sistema in equilibrio termodinamico, con microstati equiprobabili
- Le particelle (molecole) del sistema sono indistinguibili tra loro
- L'Entropia di Shannon è riferito all'informazione totale del sistema

Che detto in altri termini:

L'entropia termodinamica rappresenta (a meno dell'unità di misura) la quantità di informazione totale necessaria per conoscere completamente lo stato di un sistema

Poiché nella pratica è noto solo il macrostato del sistema, mentre il microstato rimane "nascosto" all'osservatore, si può pensare che l'entropia termodinamica sia la quantità di informazione **nascosta** nel sistema. In altre parole, se conosciamo solamente il macrostato di un sistema, allora l'entropia termodinamica è la quantità di informazione "che ci manca" per descrivere completamente il sistema.

Esempio: consideriamo 6 penne allineate su una scrivania. Se le penne sono perfettamente parallele, per descrivere il sistema mi basterà dire qualcosa del genere "le penne si trovano tutte a 8 cm dal bordo inferiore del tavolo e 4 cm dal bordo destro. Sono tutte orientate verso Nord-Est e distanti 2 cm l'una dall'altra". Se invece le 6 penne sono mescolate a casaccio sulla scrivania, allora per descrivere il sistema dovrò dire qualcosa del tipo "la prima penna si a 7 cm dal bordo inferiore, a 4 cm dal bordo destro ed è orientata di 42° rispetto al Nord. La seconda penna si trova a 12 cm dal bordo inferiore ecc. ecc.". Come si vede, tanto più disordinato è un sistema (maggiore entropia termodinamica), tanta più informazione è necessaria per descriverlo (maggiore entropia di Shannon).

Bibliografia

[1] Stefano Adriani, *Teoria dell'Informazione*, 2022

http://adriani.altervista.org/author/notes.php

[2] Wikipedia, *Entropia (teoria dell'informazione)*, 2022

https://it.wikipedia.org/wiki/Entropia_(teoria_dell%27informazione)

[3] Wikipedia, Autoinformazione, 2022

https://it.wikipedia.org/wiki/Autoinformazione