

CAPITOLO 1

Partizioni su insiemi

1.1. Generalità

Introduciamo il formalismo e risultati generali per spazi di partizioni e metriche di Rohlin, seguendo l'approccio in [billingsley, casartelli-vivo]. Sia $(\mathbf{M}, \mathcal{M}, \mu)$ uno spazio di probabilità, ovvero un insieme \mathbf{M} , una σ -algebra \mathcal{M} di sottoinsiemi di \mathbf{M} , una misura normalizzata μ su \mathbf{M} . Nei casi trattati \mathbf{M} può essere una sequenza di simboli, un reticolo bidimensionale, un grafo arbitrario.

Una partizione di \mathbf{M} è una collezione finita $\alpha \equiv (A_1, A_2, \dots, A_N)$ di sottoinsiemi disgiunti misurabili che ricoprono \mathbf{M} , cioè $A_i \cap A_j = \emptyset$ se $i \neq j$ e $\cup_k A_k = \mathbf{M}$. Gli $\{A_k\}$ sono chiamati *atomi* di α . L'insieme di tutte le partizioni misurabili è denotato con $\mathcal{Z} \equiv \mathcal{Z}(\mathbf{M})$. La partizione unitaria ν consiste del singolo atomo coincidente con \mathbf{M} . È possibile introdurre un ordine parziale su \mathcal{Z} , con la relazione $\alpha \leq \beta$ quando β è un raffinamento di α : questo accade quando ogni atomo A_k è esattamente composto da atomi di β , cioè $A_k = \{\cup_j B_j \mid B_j \in \beta\}$. In questo caso, si dice che α è un *fattore* di β . La partizione banale $\nu \leq \alpha$, $\forall \alpha$.

I termini come *unità* e *fattore* dipendono dalla definizione di uno pseudo-prodotto commutativo ed associativo, la composizione $\gamma = \alpha \vee \beta$ (o anche $\gamma = \alpha\beta$ ove non vi sia ambiguità). Il *prodotto* è la partizione meno fine di tutte le partizioni con $\gamma \geq \alpha$ e $\gamma \geq \beta$, i cui atomi sono le intersezioni non vuote di tutti gli atomi di α e β . Chiaramente il prodotto con l'unità si comporta come l'identità del prodotto, con $\alpha\nu = \alpha$ per ogni α , mentre $\alpha\eta = \alpha$ quando $\eta \leq \alpha$. Queste proprietà rendono il risultato della composizione una specie di “minimo comune multiplo”.

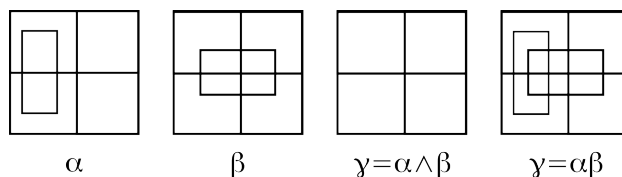


FIGURA 1.1.1. Esempio di prodotto e intersezione tra due partizioni

Un altro modo di calcolare e visualizzare le operazioni è in termini di *bordi* di atomi della partizione, come si vede in figura 1.1.2 nella pagina seguente. Il prodotto $\alpha \vee \beta$ corrisponde alla partizione avente come bordi l'unione dei bordi di α e β , mentre l'intersezione $\alpha \wedge \beta$ ha come bordi l'intersezione di quelli di α e β . Poiché la partizione banale non ha bordi tra atomi, si ricavano immediatamente le sue proprietà nel prodotto e nell'intersezione.

Una partizione può rappresentare un esperimento probabilistico con risultati disgiunti A_1, \dots, A_N , dove l'evento *atomico* A_k ha probabilità $\mu(A_k)$. Un *fattore* è quindi un sottoesperimento dell'esperimento più fine, che raggruppa diversi risultati come equivalenti: ad esempio, “pari o dispari” è un sottoesperimento con due atomi, dell'esperimento $\{1, 2, 3, 4, 5, 6\}$ del lancio di un dado.

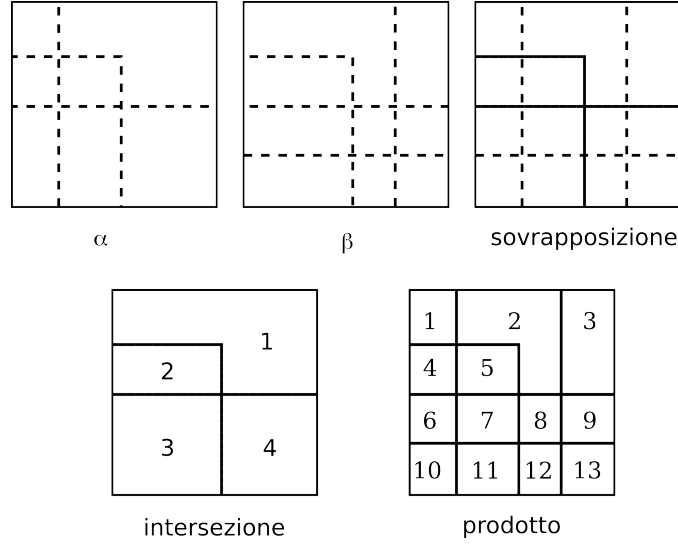


FIGURA 1.1.2. Intersezione e prodotto come operazioni sui bordi. Le partizioni α e β hanno bordi tratteggiati, con fase complementare, che danno l'impressione di linea continua quando sovrapposti. È evidente in questo modo il tratto di bordi in comune. Gli atomi sono stati numerati per rendere evidente la differenza tra le due operazioni.

Sullo spazio \mathcal{Z} possiamo definire dei funzionali *entropia* $H : \mathcal{Z} \rightarrow \mathbb{R}^+$, definiti su ogni partizione. In particolare l'*entropia di Shannon* e l'*entropia topologica*

$$(1.1.1) \quad \begin{aligned} H_S(\alpha) &= - \sum_{i=1}^n \mu(A_i) \ln \mu(A_i) \\ H_T(\alpha) &= \ln \sum_{i=1}^n 1 = \ln(n) \end{aligned}$$

L'entropia di Shannon è una misura dell'informazione media ottenuta dall'esperimento. Si vede immediatamente che la partizione banale ν , non codificando alcuna informazione ha entropia nulla in entrambi i casi. Se $\beta = (B_1, \dots, B_n)$ è un'altra partizione, l'entropia condizionale di α rispetto a β è

$$(1.1.2) \quad H_S(\alpha|\beta) = - \sum_{i=1}^n \sum_{k=1}^m \mu(A_i \cap B_k) \ln \frac{\mu(A_i \cap B_k)}{\mu(B_k)} = H_S(\alpha \vee \beta) - H_S(\beta)$$

dove si prende per convenzione che $x \ln x = 0$ se $x = 0$. L'entropia condizionale è l'informazione media residua ottenuta da α quando il risultato per β è noto. Si noti che l'entropia di Shannon dipende solo dalla distribuzione delle misure degli atomi, non dalla loro natura o "forma", che potrebbe non avere significato in spazi astratti. Le mutue relazioni tra atomi (e possibilmente le loro forme) al contrario influenzano direttamente l'entropia condizionale. L'entropia topologica non permette di definire in maniera ovvia l'entropia condizionale, ma vi sono altre quantità interessanti che non presentano questo problema.

Definiamo una metrica sullo spazio delle partizioni $\mathcal{Z}(\mathbf{M})$ tramite la distanza di Rohlin d_R

$$d_R = H(\alpha|\beta) + H(\beta|\alpha)$$

che misura la complessiva non-similarità tra le partizioni α e β . È possibile dare una definizione alternativa di questa distanza, sfruttando la seconda scrittura della probabilità condizionale, riscrivendo d_R come

$$(1.1.3) \quad d_R = 2H(\alpha \vee \beta) - H(\alpha) - H(\beta)$$

La simmetria $d_R(\alpha, \beta) = d_R(\beta, \alpha)$ e la condizione $d_R(\alpha, \alpha) = 0$ sono manifeste, mentre la disuguaglianza triangolare è soddisfatta se H soddisfa alle condizioni di un funzionale entropia. Si può usare sia l'entropia di Shannon che quella topologica per il calcolo, essendo l'entropia topologica del prodotto di partizioni perfettamente definita.

Se \mathbf{M} è finito, una *configurazione* o *stato* \mathbf{a} su \mathbf{M} è una funzione che assegna ad ogni punto $x_i \in \mathbf{M}$ un valore $a_i = f(x_i)$ nell'alfabeto \mathbb{K} . Tutte le possibili configurazioni formano uno spazio $\mathcal{C} \equiv \mathcal{C}(\mathbf{M})$. Su \mathcal{C} la distanza di Hamming è definita come

$$d_H(\mathbf{a}, \mathbf{b}) = \mathcal{N} \sum_i \rho(a_i, b_i)$$

dove $\rho(a_i, b_i)$ è una distanza su \mathbb{K} e \mathcal{N} una possibile costante di rinormalizzazione, che noi porremo uguale a 1.

1.2. Riduzione

L'essenziale dissimilarità tra due partizioni potrebbe essere confusa ed indebolita dalla presenza di un fattore comune dominante, come ad esempio accade se gli atomi della partizione hanno lunghezza media molto breve, nel qual caso la maggioranza dei confini risulta essere la stessa. Si cerca quindi di eliminare fattori comuni il più possibile, con una RIDUZIONE che ci si aspetta aumenti la distanza relativa. Tuttavia, questa operazione analoga alla riduzione in minimi termini per frazioni, non è unicamente definita, in quanto le partizioni, a differenza degli interi, non ammettono una univoca fattorizzazione in fattori primi. Il ruolo dei fattori primi (ovvero fattori irriducibili) è giocato dalle sottopartizioni *dicotomiche*, che sono tuttavia ancora estremamente ridondanti ($2^{n-1} - 1$ per partizioni con n atomi).

A partire dalla partizione $\alpha \equiv (A_1, \dots, A_n)$, definiamo quindi una famiglia ristretta $\mathbf{E}(\alpha)$ di *fattori dicotomici elementari* $\tilde{\alpha}_1, \tilde{\alpha}_2, \dots, \tilde{\alpha}_n$ con le seguenti caratteristiche:

- (1) $\mathbf{E}(\alpha)$ deve essere ben definita per ogni $\alpha \in \mathcal{Z}$
- (2) $\mathbf{E}(\alpha)$ non deve contenere più di n (il numero di atomi in α) fattori elementari
- (3) $\bigvee_{k=1}^n \tilde{\alpha}_k = \alpha$

Una scelta universale consiste nel prendere come fattori dicotomici $\tilde{\alpha}_k \equiv (A_k, A_k^c)$ le partizioni formate dai singoli atomi e dai loro complementi in \mathbf{M} . Fattori di questo tipo sono chiamati *semplici*.

Una volta che per due partizioni α, β le famiglie di fattori dicotomici $\mathbf{E}(\alpha)$ e $\mathbf{E}(\beta)$ sono state definite, abbiamo diversi possibili processi di riduzione.

DEFINIZIONE 1. Riduzione tramite fattore comune

- (1) Si definisce il massimo fattore comune $\sigma = \alpha \wedge \beta$
- (2) Si tralasciano da $\mathbf{E}(\alpha)$ e $\mathbf{E}(\beta)$ i fattori che non sono relativamente primi con σ , e indichiamo i fattori rimanenti come $\hat{\alpha}_k$ e $\hat{\beta}_k$ rispettivamente. Questo vuol dire che $\hat{\alpha}_k \wedge \sigma = \hat{\beta}_j \wedge \sigma = \nu$.

DEFINIZIONE 2. Riduzione con eliminazione atomi in comune

- (1) Si tralasciano da $\mathbf{E}(\alpha)$ e $\mathbf{E}(\beta)$ i fattori che compaiono in entrambe le partizioni. Se indichiamo i fattori rimanenti come $\hat{\alpha}_k$ e $\hat{\beta}_k$ rispettivamente, questo vuol dire che $\forall \hat{\alpha}_k, \nexists \hat{\beta}_j \in \beta | \hat{\alpha}_k = \hat{\beta}_j$ e viceversa.

DEFINIZIONE 3. Riduzione con eliminazione fattori simili

- (1) Si tralasciano da $\mathbf{E}(\alpha)$ e $\mathbf{E}(\beta)$ i fattori che hanno un corrispondente “simile” nell'altra partizione. Questo vuol dire che scartiamo il fattore

$$\alpha_k \text{ se } \exists \beta_j \in \beta, \text{ tale che } \mu(\alpha_k \triangle \beta_j) \leq \epsilon$$

e viceversa il fattore

$$\beta_k \text{ se } \exists \alpha_j \in \alpha, \text{ tale che } \mu(\beta_k \triangle \alpha_j) \leq \epsilon$$

Il simbolo $\alpha_k \triangle \beta_j$ indica la differenza simmetrica tra i due atomi, ovvero i siti che appartengono ad un atomo ma non all'altro.

Alla fine, per tutti i tipi di riduzione, definiamo le partizioni ridotte come $\hat{\alpha} = \vee_k \hat{\alpha}_k$ e $\hat{\beta} = \vee_k \hat{\beta}_k$, ovvero il prodotto dei fattori dicotomici “sopravvissuti”. Nel capitolo sugli algoritmi presenteremo metodi ottimali per il calcolo dei fattori dicotomici per ogni criterio presentato, che presentano una notevole complessità se eseguiti nel modo naive.

Per il resto della sezione concentreremo la nostra attenzione sulla riduzione tramite fattore comune massimo.

Si motiva la scelta del confronto con il fattore comune poichè vi sono casi in cui le partizioni non hanno atomi in comune, ma ciononostante si ha che $\sigma \neq \nu$. Questo accade, per esempio, quando $\alpha < \beta$ strettamente e non vi sono fattori comuni elementari. In questo caso allora $\sigma = \alpha$ e $\hat{\alpha} = \nu$ con questo metodo di riduzione, mentre $\hat{\alpha} = \alpha$ tralasciando i fattori comuni. Può capitare inoltre che anche se le partizioni sono già ridotte, non sono prime tra di loro.

1.2.1. Amplificazione. Il processo di riduzione porta alla definizione di partizioni con complessità possibilmente inferiore, ovvero $H(\hat{\alpha}) \leq H(\alpha)$. Questo va nel verso opposto quando si considera l'effetto sulla distanza, che invece aumenta.

Il rapporto di *amplificazione* misura quanto la riduzione ha messo in risalto la differenza tra partizioni e ne dimostriamo la proprietà principale

$$R = \frac{d_R(\hat{\alpha}, \hat{\beta})}{d_R(\alpha, \beta)} \geq 1$$

PROPOSIZIONE. $d_R(\hat{\alpha}, \hat{\beta}) \geq d_R(\alpha, \beta)$

DIMOSTRAZIONE. Ricordando che $\sigma = \alpha \wedge \beta$ possiamo scrivere $\alpha = \sigma \hat{\alpha}$ e $\beta = \sigma \hat{\beta}$: infatti σ contiene tutti i fattori tralasciati durante la riduzione. Utilizzando ora l'idempotenza del prodotto, $\sigma = \sigma \sigma$, possiamo riscrivere la tesi utilizzando l'equazione (1.1.3)

$$2H(\sigma \hat{\alpha} \hat{\beta}) - H(\sigma \hat{\alpha}) - H(\sigma \hat{\beta}) \leq 2H(\hat{\alpha} \hat{\beta}) - H(\hat{\alpha}) - H(\hat{\beta})$$

scambiando l'ordine dei termini si ottiene

$$2H(\sigma \hat{\alpha} \hat{\beta}) - 2H(\hat{\alpha} \hat{\beta}) \leq H(\sigma \hat{\alpha}) - H(\hat{\alpha}) + H(\sigma \hat{\beta}) - H(\hat{\beta})$$

e sfruttando la formula (1.1.2) per l'entropia condizionata, la tesi si riduce a

$$2H(\sigma | \hat{\alpha} \hat{\beta}) \leq H(\sigma | \hat{\alpha}) + H(\sigma | \hat{\beta})$$

ma questo è chiaramente vero, in quanto

$$H(\sigma | \hat{\alpha} \hat{\beta}) \leq H(\sigma | \hat{\alpha}) \quad \text{e} \quad H(\sigma | \hat{\alpha} \hat{\beta}) \leq H(\sigma | \hat{\beta})$$

per le proprietà dell'entropia, poichè il termine condizionante è sicuramente maggiore, ovvero $\hat{\alpha} \hat{\beta} \geq \hat{\alpha}$ e $\hat{\alpha} \hat{\beta} \geq \hat{\beta}$. \square

Da notare inoltre, che tutto quanto scritto finora, vale per la distanza di Rohlin definita sia tramite H_S che H_T .

Risulta importante la scelta della famiglia di fattori dicotomici $\mathbf{E}(\alpha)$, che è dettata dalla topologia e geometria dello spazio delle configurazioni. La scelta della famiglia di fattori dicotomici universali semplici è sempre possibile, poichè la determinazione di A_k^c a partire da A_k è un'operazione ben definita in qualunque spazio topologico. Prendere come fattori elementari la parte interna di contorni di cluster ad esempio, richiede un concetto di orientabilità e la possibilità di definire contorni, ovvero insiemi con codimensione 1 su varietà – mentre vorremmo estendere l'analisi anche a grafi generici, privi di strutture geometriche predefinite. Già nel caso lineare è possibile prendere fattori dicotomici diversi e algoritmicamente più performanti, a patto di restringere lo studio alle partizioni con atomi formati da cluster connessi.

1.3. Partizionamento dello spazio delle configurazioni

L'applicabilità dei metodi discussi è assolutamente generica, estendibile a qualunque spazio di probabilità finito si voglia considerare, vediamo dunque di dare esempi dei possibili spazi \mathbf{M} su cui abbiamo lavorato, con i relativi fattori dicotomici e conseguenze computazionali.

Essendo lo studio svolto su calcolatore, lo spazio \mathbf{M} e la sua σ -algebra sono finiti e discreti. I siti appartenenti ad \mathbf{M} possono essere sempre numerati ordinati in modo opportuno x_i , $i \in \{1, \dots, L\}$, dove L è il numero totale di siti, che si tratti di un reticolo o di un grafo.

Per partizionare i siti in atomi disgiunti, richiediamo che la *configurazione* (o *stato*) \mathcal{C} , associ ad ogni sito una lettera dell'alfabeto \mathbb{K} , considerato finito, $|\mathbb{K}| < \infty$. Nel caso in cui lo stato del sistema è descritto con variabili continue (o vi sia un numero enorme di possibili lettere nell'alfabeto, si pensi ad una variabile a 64 bit rappresentante un numero reale), si può sempre ridurre l'alfabeto raggruppando valori "vicini" con $|f(x_i) - k_j| \leq \epsilon$.

1.3.1. Sequenze lineari connesse. Consideriamo sequenze lunghe L , provenienti da due casi:

- Problemi di meccanica statistica, in cui la configurazione è una variabile aleatoria, generata algoritmicamente con metodo Montecarlo su modello di Ising monodimensionale, nel qual caso l'alfabeto corrisponde a $\{-1, +1\}$.
- Sequenze di origine biologica, in particolare sequenze di amminoacidi (proteine), in cui $|\mathbb{K}| = 22$.

Lo studio delle sequenze è solitamente svolto con la distanza di Hamming d_H che tuttavia è molto sensibile a variazioni puntuali dei valori in \mathcal{C} . I siti che compongono la sequenza non si influenzano, una variazione su un sito può solo variare di $\{-1, 0, +1\}$ la distanza totale.

Ad ogni configurazione possiamo associare una partizione in \mathcal{Z} , in cui gli atomi sono formati dai cluster, cioè sottoinsiemi connessi di \mathbf{M} a valori omogenei in \mathbb{K} . Questo stabilisce una mappa $\Phi : \mathcal{C} \rightarrow \mathcal{Z}$ da ogni configurazione ad una partizione corrispondente, rendendo possibile il confronto tra $d_H(\mathbf{a}, \mathbf{b})$ in \mathcal{C} e $d_R(\alpha, \beta)$ in \mathcal{Z} , dove $\alpha = \Phi(\mathbf{a})$ e $\beta = \Phi(\mathbf{b})$. La relazione è chiaramente del tipo molti-a-uno, infatti assegnando ad un segmento di lettere omogeneo in \mathcal{C} un diverso simbolo, non cambia la partizione corrispondente.

È evidente quindi che variazioni locali, ad esempio il cambiamento di un singolo simbolo, possono non modificare affatto la partizione

$$\{\dots, T, T, C, A, A, \dots\} \stackrel{\Phi}{\equiv} \{\dots, T, T, M, A, A, \dots\}$$

che presenta sì una perdita di informazione, ma permette quindi anche di filtrare molto “rumore” e si è dimostrata una ottima scelta sia nel caso biologico che nello studio di sequenze di Ising.

ESEMPIO. Supponiamo di aver partizionato la sequenza dei numeri $\{1,2,3,4,5,6,7,8,9,10\}$, una stringa come $\{D,D,D,C,C,B,B,K,K,K\}$ o $\{A,A,A,B,B,C,C,D,D,D\}$, aventi tutti la stessa partizione

$$\alpha = \{(1, 2, 3), (4, 5), (6, 7), (8, 9, 10)\}$$

il calcolo esplicito dell’entropia è il seguente:

$$\begin{array}{ll} A_1 = (1, 2, 3) & \mu = \frac{3}{10} \\ A_2 = (4, 5) & \mu = \frac{2}{10} \\ A_3 = (6, 7) & \mu = \frac{2}{10} \\ A_4 = (8, 9, 10) & \mu = \frac{3}{10} \end{array}$$

Ora, $H_S(\alpha) = -2(0.3) \ln(0.3) - 2(0.2) \ln(0.2) \simeq 1.36$, mentre $H_T(\alpha) = \ln(4) \simeq 1.38$. Come si vede i valori sono abbastanza simili.