

Le simmetrie del DNA n. 3

Piano Nazionale Lauree Scientifiche.
Progetto Math-en-Jeans.
Dipartimento di Matematica
Università di Padova.

Mario Puppi, Valentina Novello

1 febbraio 2019

3. Costruzioni di osservabili.

3.1 Costruzioni di pattern. Fissiamo l'attenzione sull'insieme \mathcal{B}^2 delle 16 parole formate da due lettere dell'alfabeto genetico $\mathcal{B} = \{A, C, T, G\}$.

A partire dagli osservabili di \mathcal{B}^2 costruiremo dei nuovi pattern, usando una costruzione standard.

Prendiamo due osservabili $X, Y \in \mathcal{B}^2$ e consideriamo il pattern che inizia con X , finisce con Y , con un separatore di lunghezza l caratteri tra i due pattern X, Y . Il risultato sarà indicato con $X : l : Y$. Assumeremo che la distanza l di separazione possa variare da 1 a 10^7 .

Per confrontare gli osservabili così costruiti, definiamo una frequenza normalizzata:

$$z_{[X,Y]}(l) := \frac{P(X : l : Y)}{P(X)P(Y)}$$

Esercizio 4. Al variare delle parole $X, Y \in \mathcal{B}^2$ e della distanza $l \in \{2^n : n = 0, 1, \dots, 24\}$, calcolare:

- le frequenze $P(X), P(Y)$
- le frequenze delle correlazioni $P(X : l : Y)$
- le frequenze normalizzate $z_{[X,Y]}(l)$ delle stesse correlazioni

Esercizio 5. Studiare l'effetto dell'operatore R di scambio definito da

$$(X; l; Y) \xrightarrow{R} (Y; l; X)$$

Esistono dei pattern X, Y e distanze l tali che $(X; l; Y)$ abbia la stessa frequenza di $(Y; l; X)$?

Esercizio 6. Studiare l'effetto sulle frequenze dell'operatore C definito da

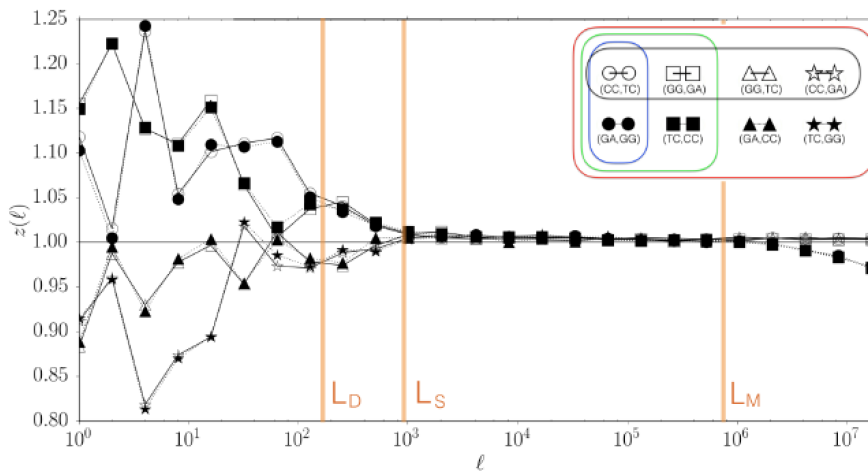
$$(X; l; Y) \xrightarrow{C} (\hat{X}; l; Y)$$

dove \hat{X} è l'inverso complementare di X .

Osserviamo che

- $RC \neq CR$ cioè R e C non commutano
- $RR = CC = Id$, cioè R, C sono involutorie.
- CRC ha lo stesso effetto della simmetria inverso-complementare dato che il suo effetto sul pattern $(X; l; Y)$ è di trasformarlo in $(\hat{Y}; l; \hat{X})$

Esercizio 7. Dati i due dinucleotidi $X_A = CC, X_B = TC$ ottenere i pattern $(X_A; l; X_B)$ e i loro simmetrici, usando gli operatori R, C in tutti i modi possibili. Verificare che le correlazioni normalizzate, $z_{[X_A, X_B]}(l)$, in funzione della scala l , presentano il grafico seguente (Crisadoro G. Degli Esposti M. Altmann E. G. *The common origin of symmetry and structure in genetic sequences*, 31 ottobre 2018).



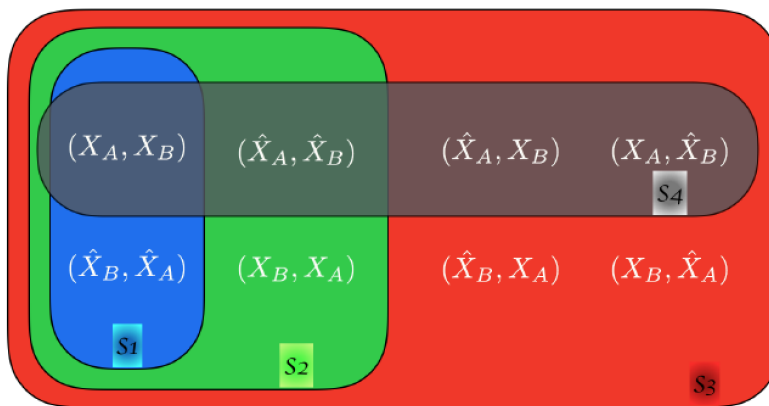
Siano S un insieme finito di simmetrie composte da R e C e Y un osservabile. Consideriamo l'insieme $\mathcal{S}_S(Y)$ degli osservabili ottenuti applicando a Y tutti gli operatori dell'insieme S .

Per esempio, se $S = \{CRC\}$ allora

$$\mathcal{S}_S(X : l : Y) = \{X : l : Y, \hat{Y} : l : \hat{X}\}.$$

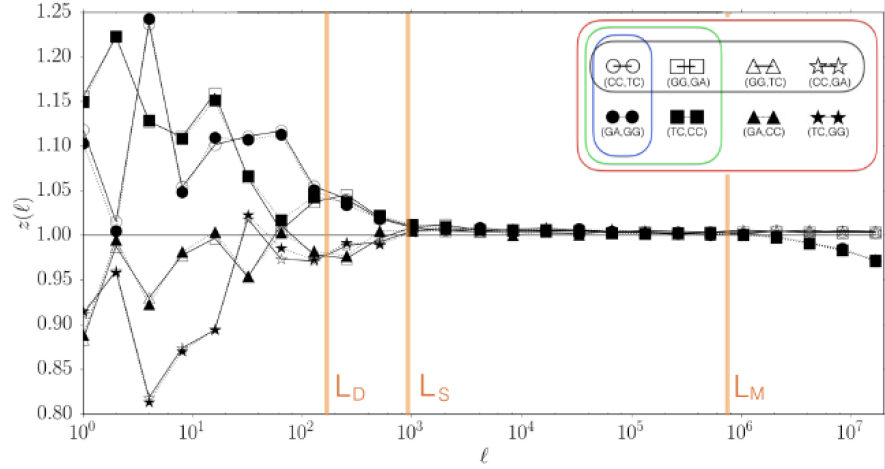
Se $S = \{CRC, R\}$ allora

$$\mathcal{S}_S(X : l : Y) = \{X : l : Y, Y : l : X, \hat{Y} : l : \hat{X}, \hat{X} : l : \hat{Y}\}$$



Esercizio 8. Dati due pattern X_A, X_B , si consideri l'osservabile $Y = X_A : l : X_B$. Applichiamo a Y quattro diversi insiemi di simmetrie $S1, S2, S3, S4$. $S1 = \{CRC\}$, colore blu $S2 = \{CRC, R\}$, colore verde, $S3 = \{R, C\}$, colore rosso, e infine $S4 = \{RCR, C\}$, colore nero. Fare una verifica sperimentale con $X_A = CC, X_B = TC$.

La figura mostra come variano le frequenze di correlazione a scale che vanno da $l = 1$ a $l = 10^7$. A scale diverse si presentano simmetrie diverse.



A una scala $l < l_D$ le curve sono significativamente diverse da $z = 1$ e appaiono a coppie (stesso simbolo, la simmetria S_1). Si ha quasi coincidenza per fluttuazioni casuali. Ad una scala $l = 2^i, i \in \{0, 1, 2, \dots, 24\}$ si presenta la simmetria S_2 in cui due coppie si fondono formando due gruppi di 4 curve ciascuno.

Ad una scala dell'ordine $l_S = 10^3$ tutte le curve coincidono (simmetria S_3).

Un altro valore critico è la soglia $L_M = 10^6$, alla quale compaiono nuove simmetrie.

La dipendenza dei risultati dalla scala ci induce a quantificare la validità delle simmetrie a scale differenti. Calcoliamo per ogni simmetria a una scala l un indice $I_S(l)$ che misura la distanza tra le curve $z(l)$ di coppie di pattern (X_A, X_B) legati da simmetria e confrontiamo questa distanza con quelle che non sono collegate con S . Più precisamente, per una coppia di riferimento $Y = (X_A, X_B)$ e una simmetria $S \in \{S_1, S_2, S_3, S_4\}$, consideriamo la seguente distanza di Y dall'insieme \mathcal{S}_S degli osservabili ottenuti dalla simmetria S .

$$d_\ell(Y_{\text{ref}}, S) := \frac{1}{|\mathcal{S}_S|} \sum_{Y \in \mathcal{S}_S(Y_{\text{ref}})} \frac{[z_Y(\ell) - z_{Y_{\text{ref}}}(\ell)]^2}{\sigma^2(\ell)}$$

dove $\sigma(l)$ denota la deviazione standard di $z(l)$ sopra tutti gli Y . Facciamo poi la media su tutti gli Y , cioè tutte le coppie $Y = (X_A, X_B)$ per ottenere una misura della simmetria S a una scala l con l'espressione:

$$I_S(\ell) := \frac{1}{2|\mathcal{A}|} \sum_{Y_{\text{ref}} \in \mathcal{A}} d_\ell(Y_{\text{ref}}, S)$$

$I_S(l) = 0$ indica la piena validità della simmetria S alla scala l (z è lo stesso per tutti gli Y in S) and $I_S(l) = 1$ indica che la simmetria S non è valida alla scala l (z varia in S alla stessa maniera in cui varia su tutto l'insieme).

Esercizio 9. Validità di una simmetria a scale diverse.
Calcolare l'indice di validità $I_S(l)$ della simmetria S a scale diverse l e verificare che, nel cromosoma 1, si ha la situazione data dal grafico seguente:

