

Distanza genetica fra le popolazioni

Desiderio Poletto

Centro PRISTEM, Università Bocconi

Settembre 2018

Per ricordare Luigi Luca Cavalli Sforza, il grande genetista scomparso lo scorso 31 agosto all'età di 96 anni, proponiamo un'applicazione riguardante la genetica delle popolazioni in cui, per calcolare la "distanza tra le popolazioni", il celebre scienziato ricorre al prodotto scalare di vettori.

Nell'articolo del 1967 di Luigi Luca Cavalli-Sforza and Anthony W. F. Edwards "Phylogenetic analysis: models and estimation procedures" apparso sulla rivista *American Journal of Human Genetics*, si ricorre al prodotto scalare di vettori per misurare la "distanza genetica" tra le popolazioni.

Dell'interessante articolo riportiamo solo la parte relativa all'uso del prodotto scalare di vettori, rimandando i lettori interessati alla lettura integrale dello stesso per ulteriori approfondimenti.

Per il loro esperimento gli autori hanno scelto elementi delle quattro popolazioni (indicate nella tabella sottostante) e di ogni gruppo hanno determinato le frequenze f_{in} dei geni nei gruppi di sangue A B O.

I dati in tabella riportano la distinzione tra quattro alleli A_1 , A_2 , B, O.

Allele	Eschimese f_{1n}	Bantu f_{2n}	Inglese f_{3n}	Coreano f_{4n}
A1	0,2914	0,1034	0,2090	0,2208
A2	0,0000	0,0866	0,0696	0,0000
B	0,0316	0,1200	0,0612	0,2069
O	0,6770	0,6900	0,6602	0,5723
Totale	1,0000	1,0000	1,0000	1,0000

Distanza genetica fra le popolazioni

Per calcolare la “distanza” hanno fatto ricorso al prodotto di vettori secondo lo schema seguente.

Si costruiscono i vettori normalizzati e, siccome stiamo lavorando con le frequenze, è sufficiente eseguire le radici quadrate dei valori presenti nella tabella.

$$x_1 = \begin{bmatrix} x_{11} = \sqrt{f_{11}} \\ x_{12} = \sqrt{f_{12}} \\ x_{13} = \sqrt{f_{13}} \\ x_{14} = \sqrt{f_{14}} \end{bmatrix}; \quad x_2 = \begin{bmatrix} x_{21} = \sqrt{f_{21}} \\ x_{22} = \sqrt{f_{22}} \\ x_{23} = \sqrt{f_{23}} \\ x_{24} = \sqrt{f_{24}} \end{bmatrix}$$

$$x_3 = \begin{bmatrix} x_{31} = \sqrt{f_{31}} \\ x_{32} = \sqrt{f_{32}} \\ x_{33} = \sqrt{f_{33}} \\ x_{34} = \sqrt{f_{34}} \end{bmatrix}; \quad x_4 = \begin{bmatrix} x_{41} = \sqrt{f_{41}} \\ x_{42} = \sqrt{f_{42}} \\ x_{43} = \sqrt{f_{43}} \\ x_{44} = \sqrt{f_{44}} \end{bmatrix}$$

Per poterli confrontare fra loro li disponiamo su una sfera (in uno spazio quadrimensionale) di raggio unitario (l'immagine successiva simula il procedimento da seguire nel caso gli alleli fossero tre).

CAVALLI-SFORZA AND EDWARDS

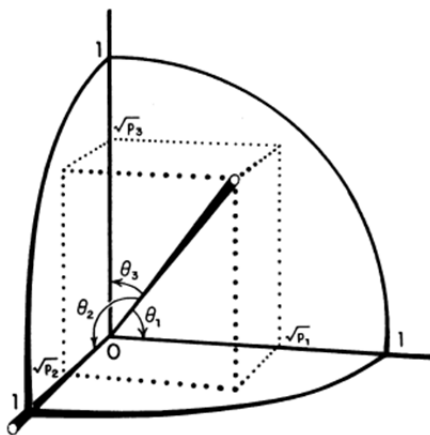


FIG. 4. Representation of a population, with gene frequencies p_1, p_2, p_3 at a single tri-allelic locus, on the octant of a sphere.

Figura tratta dall'articolo citato, visualizza una situazione in uno spazio a tre dimensioni

Se ora moltiplichiamo i vettori

$$x_1^T \cdot x_2 = [0,5398 \quad 0,0000 \quad 0,1778 \quad 0,8228] \cdot \begin{bmatrix} 0,3216 \\ 0,2943 \\ 0,3464 \\ 0,8307 \end{bmatrix} = 0,9187$$

otteniamo il valore coseno dell'angolo che si forma tra i due vettori considerati:

$$\cos(\alpha_{12}) = 0,9187 \text{ da cui } \alpha_{12} = \alpha_{21} = \arccos(0,9187) = 0,4062 \text{ rad}$$

Se trasformiamo questo valore in gradi sessagesimali otteniamo: $\alpha_{12} = \alpha_{21} = 23,2733^\circ$.

Distanza genetica fra le popolazioni

Seguendo l'immagine possiamo prendere questo valore come indicatore della "*distanza genetica*" tra la popolazione Eschimese e quella Bantu.

Ora ripetiamo i calcoli per determinare gli altri quattro valori.

Possiamo scrivere la seguente matrice simmetrica:

	Eschimese	Bantu	Inglese	Coreano
Eschimese	0°	23,2733°	16,4011°	16,8699°
Bantu	23,2733°	0°	9,8449°	20,4351°
Inglese	16,4011°	9,8449°	0°	19,6052°
Coreano	16,8699°	20,4351°	19,6052°	0°

Leggendo i valori possiamo osservare che la più grande "*distanza genetica*" si ha tra la popolazione Eschimese e quella Bantu; mentre quella più breve è quella tra le popolazioni Bantu e Inglesi.