

Models de Markov i Matrius de Substitució

Roderic Guigó i Serra, UPF, gener 2005.

Suposem seqüències només de dos símbols $\{A, B\}$, i suposem que es produeixen mutacions en aquestes seqüències (intercanvis entre A i B) al llarg del temps.

Suposem que després d'un interval de temps t , les probabilitats de mutació són:

$$\begin{aligned}P(A \rightarrow B) &= 0.2 \quad (\text{i } P(A \rightarrow A) = 0.8) \\P(B \rightarrow A) &= 0.1 \quad (\text{i } P(B \rightarrow B) = 0.9)\end{aligned}$$

Suposem una seqüència (o unes seqüències) en aquest alfabet, $\{A, B\}$, i suposem que en un moment determinat—el moment inicial—, la freqüència dels símbols A i B és:

$$\begin{aligned}P_0(A) &= 0.5 \\P_0(B) &= 0.5\end{aligned}$$

Després d'un interval de temps t , les probabilitats d' A i B es poden calcular de la manera següent:

$$\begin{aligned}P_1(A) &= P_0(A)P(A \rightarrow A) + P_0(B)P(B \rightarrow A) \\P_1(B) &= P_0(A)P(A \rightarrow B) + P_0(B)P(B \rightarrow B)\end{aligned}$$

És a dir,

$$\begin{aligned}P_1(A) &= (0.5 \times 0.8) + (0.5 \times 0.1) = 0.45 \\P_1(B) &= (0.5 \times 0.9) + (0.5 \times 0.2) = 0.55\end{aligned}$$

Aquestes equacions es poden representar de manera matricial.

Si $P_0 = (0.5 \ 0.5)$ i $M = \begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix}$, aleshores tenim que

$$P_1 = P_0 M = (0.5 \ 0.5) \times \begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix} = (0.45 \ 0.55)$$

Podem utilitzar el mateix raonament per calcular les probabilitats d' A i B després d'un nou interval de temps t :

$$P_2 = P_1 M = (0.45 \ 0.55) \times \begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix} = (0.415 \ 0.585)$$

(visiteu <http://wims.unice.fr> si voleu multiplicar les matrius)

Fixeu-vos, però, que $P_2 = P_1 M = (P_0 M) M = P_0 M^2$

Donat que $M^2 = \begin{pmatrix} 0.66 & 0.34 \\ 0.17 & 0.83 \end{pmatrix}$, aleshores també haguessim pogut calcular P_2 com

$$P_2 = P_0 M^2 = (0.5 \ 0.5) \times \begin{pmatrix} 0.66 & 0.34 \\ 0.17 & 0.83 \end{pmatrix} = (0.415 \ 0.585)$$

i, en general,

$$P_N = P_0 M^N$$

M és l'anomenada matriu de transició d'una cadena de Markov. Podem interpretar la matriu de Dayhoff com una matriu de transició d'una cadena de Markov que descriu el procés evolutiu mitjançant el qual van canviar les freqüències dels aminoàcids en una (o unes) seqüències de proteïnes. Donades unes determinades freqüències d'ocurrència dels 20 aminoàcids, la matriu de Dayhoff ens permet calcular les freqüències esperades després de l'interval de temps en que s'ha produït una mutació per cada 100 aminoàcids.