

8. November 2006

Abgabe: 15. November 2006

Aufgabe 1: Sei S_0, S_1, S_2, \dots eine “Score-Irrfahrt” und $\mathcal{S} = \{x \mid \exists i : \text{Ws}(S_i = x) > 0\}$. Implementieren Sie ein Programm, das nach Eingabe von “Aminosäurehäufigkeiten” (p_1, \dots, p_{20}) , einer Scorematrix $(s_{ij})_{i,j \leq 20}$ und einem Wert λ folgende Werte numerisch berechnet:

$$\begin{aligned} \delta &:= \min \left\{ |x - y| \mid x, y \in \mathcal{S}, x \neq y \right\} \\ P &:= \exp \left(- \sum_{\ell=1}^{\infty} \frac{1}{\ell} \cdot \text{Ws}(S_\ell \geq 0) \right) \\ Q &:= \exp \left(- \sum_{\ell=1}^{\infty} \frac{1}{\ell} \sum_{z \in \mathcal{S} \cap]-\infty, 0[} e^{\lambda z} \cdot \text{Ws}(S_\ell = z) \right) \\ k_1 &:= \frac{\delta \cdot P^2 \cdot Q^2}{(e^{\lambda \delta} - 1) \cdot \mathbb{E}(S_1 e^{\lambda S_1})} \\ k_2 &:= k_1 \cdot e^{\lambda \delta} \end{aligned}$$

Aufgabe 2: Simulieren Sie für verschiedene Kombinationen von Aminosäurehäufigkeiten (p_1, \dots, p_{20}) und Score-Matrizen $(s_{ij})_{i,j \leq 20}$ Paare voneinander unabhängiger, langer Sequenzen. Untersuchen Sie, wie Anzahl der HSPs mit Score b von b abhängt. Finden Sie Werte λ und k , so dass $nmke^{-\lambda b}$ für große b möglichst gut mit den von Ihnen beobachteten Anzahlen an HSPs mit $\text{Score} \geq b$ übereinstimmt, wobei n und m die Längen Ihrer Sequenzen sind.

Aufgabe 3: Simulieren Sie neue Daten mit denselben Aminosäurewahrscheinlichkeiten und derselben Score-Matrix und untersuchen Sie wie gut die Häufigkeit von HSPs vom $\text{Score} \geq b$ jeweils mit $nmke^{-\lambda b}$, $nmk_1 e^{-\lambda^* b}$ und $nmk_2 e^{-\lambda^* b}$ übereinstimmt, wobei k und λ Ihre Schätzwerte aus Aufgabe 2 und k_1 , k_2 und λ^* Ihre Ergebnisse aus Aufgabe 1 dieses und Aufgabe 3 des vorherigen Übungsblattes sind.