

6. Übung zur Vorlesung “Algorithmische Phylogenetik”

Sebastian Böcker, Markus Fleischauer

Aufgabe 1 (iid Modell) (5 Punkte)

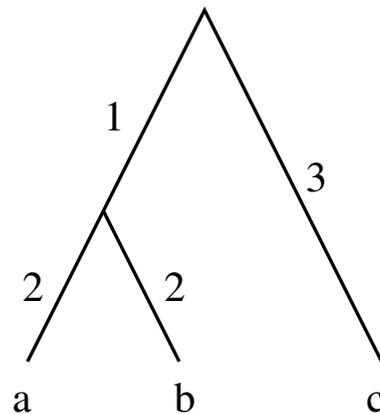
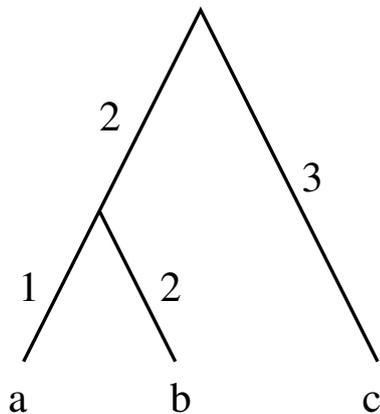
Wir betrachten zwei unabhängige zufällige Sequenzen der Länge n über einem Alphabet der Größe σ . Jeder Buchstabe, darunter A, ist mit gleicher Wahrscheinlichkeit an jeder Position anzutreffen, unabhängig von den anderen Positionen (iid Modell). Wie groß ist die Wahrscheinlichkeit, dass die erste Sequenz genauso viele As enthält wie die zweite?

Aufgabe 2 (Likelihood) (10 Punkte)

Gegeben sei das Alphabet $\Sigma = \{a, b, c, d\}$, die Anfangsverteilung $\pi^{(0)} = (0.3 \ 0.2 \ 0.3 \ 0.2)$ und die folgende Übergangsmatrix

$$P = \begin{pmatrix} 0.7 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.7 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.7 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.7 \end{pmatrix}.$$

Berechnen Sie die Likelihood der folgenden Bäume. Beachten Sie, dass die Wahrscheinlichkeiten von d durchaus eine Rolle spielen, auch wenn d nicht als Label vorkommt.



Aufgabe 3 (Jukes-Cantor) (5 Punkte)

Gegeben sei das multiple Alignment

$$\begin{pmatrix} A & C & G & T & A & T & T & C \\ A & G & G & T & A & A & T & C \\ A & C & C & A & T & A & A & C \\ T & A & A & A & T & G & C & C \end{pmatrix}$$

Berechnen Sie die Jukes-Cantor korrigierten Distanzen der paarweisen Alignments und diskutieren Sie die Ergebnisse.

Aufgabe 4 (Generalised time reversible, GTR) (2 Punkte)

Warum ist das GTR-Modell das "beste" Modell der Sequenzevolution? Und warum ist es oftmals sinnvoll, simple Modelle zu benutzen?

Aufgabe 5 (Bootstrapping) (3 Punkte)

Gegeben sei das Alignment

$$A = \begin{pmatrix} A & G & T & G \\ C & G & T & T \\ A & G & C & G \end{pmatrix}$$

Welche der folgenden Alignments wurden durch Bootstrapping aus A erstellt? Warum?

$$B = \begin{pmatrix} A & T & G \\ C & T & T \\ A & C & G \end{pmatrix} \quad C = \begin{pmatrix} A & T & G & A \\ C & T & G & C \\ A & G & G & A \end{pmatrix} \quad D = \begin{pmatrix} G & G & G & G \\ G & G & G & G \\ G & G & G & G \end{pmatrix}$$

Aufgabe 6 (MC MC) (Bonusaufgabe)

Warum lässt man bei Monte Carlo Markov Chain (MC MC) am Anfang den Prozess einbrennen?