

# 5. Übung zur Vorlesung “Algorithmische Phylogenetik”

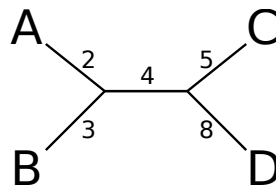
Sebastian Böcker, Markus Fleischauer

Ausgabe: 23.06.2014

Abgabe: 07.07.2014

## Aufgabe 1 (SplitsTree) (5 Punkte)

Gegeben sei folgender Baum:



Berechnen Sie die Isolationsindizes der zugehörigen additiven Metrik.

**Hinweis:** Das geht auch sehr einfach, aber berechnen Sie mindestens einen Isolationsindex explizit!

## Aufgabe 2 (SplitsTree, NeighborNet) (5 Punkte)

Gegeben sei folgende Distanzmatrix:

	A	B	C	D	E
A	0	6	8	9	9
B		0	5	8	10
C			0	4	8
D				0	7
E					0

Testen Sie `SplitsTree` mit der angegebenen Distanzmatrix. Eine Java-Version zum download ist unter <http://www.splitstree.org/> erhältlich.

Vergleichen sie die Ergebnisse von Neighbor Joining, Split Decomposition und NeighborNet. Auch Neighbor Joining und NeighborNet sind in `SplitsTree` enthalten.

## Aufgabe 3 (Übergangsmatrix) (5 Punkte)

Gegeben sei die folgende (1-Schritt-)Übergangsmatrix

$$P = \begin{pmatrix} 8/10 & 1/10 & 1/10 \\ 2/10 & 7/10 & 1/10 \\ 2/10 & 2/10 & 6/10 \end{pmatrix}.$$

Berechnen Sie die 8-Schritt-Übergangsmatrix  $P^8$ .

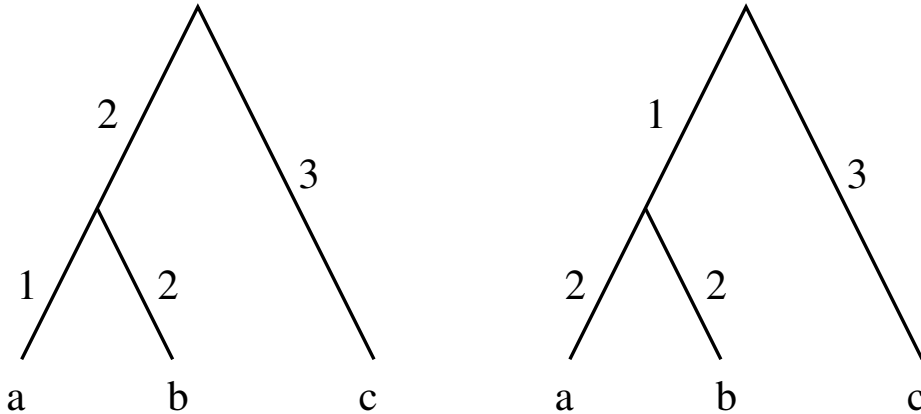
**Hinweis:** Sie können zur Vereinfachung  $\frac{1}{10}$  rausziehen.

**Aufgabe 4 (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 5 (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 6 (Generalised time reversible, GTR)** (3 Punkte)

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

**Aufgabe 7 (Bootstrapping)** (4 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 8 (MC MC)** (3 Punkte)

Warum lässt man bei Monte Carlo Markov Chain (MC MC) am Anfang den Prozess einbrennen? Warum ist bei MCMC Methoden keine zusätzliche Bootstrap-Analyse notwendig um eine Aussage über die Robustheit von Splits/Kladen treffen zu können.