

# 7. Übung zur Vorlesung “Bioinformatische Methoden in der Genomforschung”

Sebastian Böcker, Markus Fleischauer, Sascha Winter  
Zur Verfügung gestellt von Prof. Dr. Ivo Große, Alexander Gabel

Ausgabe: 19.12.2013

Abgabe: 16.01.2014

## Aufgabe 1 (5 Punkte)

1. Leiten Sie die Identität

$$\prod_{n=1}^N \prod_{\ell=1}^{L_n} \theta_{x_{n\ell}} = \prod_{a \in \{A,C,G,T\}} \theta_a^{N_a}$$

her, wobei  $N_a = \sum_{n=1}^N \sum_{\ell=1}^{L_n} \delta_{x_{n\ell},a}$  die absolute Häufigkeit von Base  $a$  in Datensatz  $\underline{x}$  bezeichnet.

2. Leiten Sie die Identität

$$\prod_{n=1}^N \prod_{\ell=1}^W \theta_{x_{n\ell}}^{(\ell)} = \prod_{\ell=1}^W \prod_{a \in \{A,C,G,T\}} (\theta_a^{(\ell)})^{N_a^{(\ell)}}$$

her, wobei  $N_a^{(\ell)} = \sum_{n=1}^N \delta_{x_{n\ell},a}$  die absolute Häufigkeit von Base  $a$  an Position  $\ell$  in Datensatz  $\underline{x}$  bezeichnet.

## Aufgabe 2 (5 Punkte)

Leiten Sie den Maximum-Likelihood-Schätzer der Parameter eines (a) homogenen und (b) inhomogenen Markov-Modells 0. Ordnung her.

## Aufgabe 3 (5 Punkte)

Bestimmen Sie den Maximum-Likelihood-Schätzer

1. für ein homogenes Markov-Modell 0. Ordnung basierend auf dem Datensatz `seq_hMM_1` und berechnen Sie die Log-Likelihoods der Datensätze `seq_hMM_1` und `seq_hMM_2`.
2. für ein inhomogenes Markov-Modell 0. Ordnung basierend auf dem Datensatz `seq_iMM_1` und berechnen Sie die Log-Likelihoods der Datensätze `seq_iMM_1` und `seq_iMM_2`.

**Aufgabe 4** (5 Punkte)

Versuchen Sie, den Maximum-Likelihood-Schätzer der Parameter eines OOPS-Modells analytisch zu bestimmen. Wo bzw. warum scheitert die analytische Maximierung der Log-Likelihood?

**Aufgabe 5** (5 Punkte)

Beweisen Sie, dass aus der Identität

$$P(x_1, \dots, x_N, u_1, \dots, u_N | \theta) = \prod_{i=1}^N P(x_i, u_i | \theta)$$

die Identitäten

$$P(x_1, \dots, x_N | \theta) = \prod_{i=1}^N P(x_i | \theta)$$

$$P(u_1, \dots, u_N | \theta) = \prod_{i=1}^N P(u_i | \theta)$$

$$P(x_1, \dots, x_N | u_1, \dots, u_N, \theta) = \prod_{i=1}^N P(x_i | u_i, \theta)$$

$$P(u_1, \dots, u_N | x_1, \dots, x_N, \theta) = \prod_{i=1}^N P(u_i | x_i, \theta)$$

folgen.

**Aufgabe 6** (5 Punkte)

Beweisen Sie die Identität:

$$\begin{aligned} & \sum_{u_1=1}^K \cdots \sum_{u_N=1}^K \left( \log \prod_{i=1}^N P(U_i = u_i, X_i = x_i | \theta) \left( \prod_{j=1}^N P(U_j = u_j | X_j = x_j, \theta^{(t)}) \right) \right) \\ &= \sum_{i=1}^N \sum_{k=1}^K \log P(U_i = k, X_i = x_i | \theta) P(U_i = k | X_i = x_i, \theta^{(t)}) \end{aligned}$$