

# Algorithms in Bioinformatics II, SS2003

## Assignment sheet # 2

Daniel Huson

May 5, 2003

To complete the following assignments, please download the file [www-ab.informatik.uni-tuebingen.de/teaching/ss03/abi2/java/assign02.zip](http://www-ab.informatik.uni-tuebingen.de/teaching/ss03/abi2/java/assign02.zip). There you will find four java files:

- `HMMBase.java` - basic functionality of an HMM class including reading and writing, please do not modify,
- `HMM.java` - extends `HMMBase`,
- `SimulateData.java` - program for simulating data from an HMM, and
- `DecodeData.java` - program for decoding a sequence of symbols, using an HMM

### 1 Simulating data using an HMM (4 points)

In `HMM.java`, please implement a method `simulateData` that generates a string of symbols using the given HMM. The symbols should be returned by the method `getSimulatedSymbols()`. For comparison purposes, please also implement the related method `getSimulatedStates()`.

Based on this, please complete the program `SimulateData.java` and then run it five times on each of the given HMMs `casino.hmm`, `cpg.hmm` and `urns.hmm` generating files `casino1.symbols` and `casino1.states`, `casino2.symbols` etc.

### 2 Decoding data using the Viterbi algorithm (4 points)

In `HMM.java`, please implement a method `runViterbi` that applies the Viterbi algorithm to a given sequence of symbols and returns the decoded sequence of states.

Based on this, please complete the program `DecodeStates.java` and then run it on each of the data sets generated in the previous exercise, producing an output in the format:

```
Symbols: ACTGTGACGTGT...      (original symbols, 60 char per line)
States  : actGTGACGtgt...      (original states, 60 char per line)
Viterbi: acTGTGAcgtgt...      (decoded states, 60 char per line)
```

### 3 The most probable path (2 points)

The Viterbi algorithm determines the most probable path, i.e. the path that maximizes  $P(x, \pi)$ . Please show that the same path also maximizes  $P(\pi | x)$ , in other words, we have:

$$\arg \max_{\pi} P(x, \pi) = \arg \max_{\pi} P(\pi | x).$$

Assignments due: **Monday, May 12, 10am**