Algorithms in Genome Research
## Winter 2012/2013

Exercises

Number 8, Discussion: 14. December 2012

1. Given a genomic DNA-Fragment `"GCTGAAGATCATTGTCAAGT"`
   and the multiple alignment shown below, do the following:

   ```
   G A T T - - A T
   C A T A G C A T
   C A A G G C T A
   G A A G - T C T
   G A A - - C - T
   ```

   (a) Use the alignment to construct a PSSM without Pseudocounts.
   (b) Same as in (a), but add Pseudocounts of 0.5 for each base.
   (c) Find the most probable matches to the genomic sequence using your PSSMs from (a) and (b).

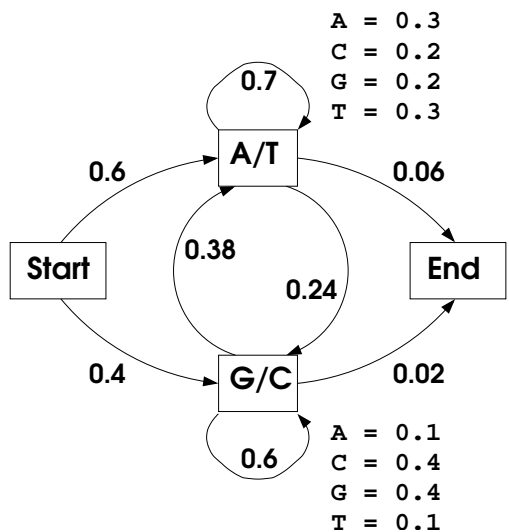2. Given the following Markov chain.
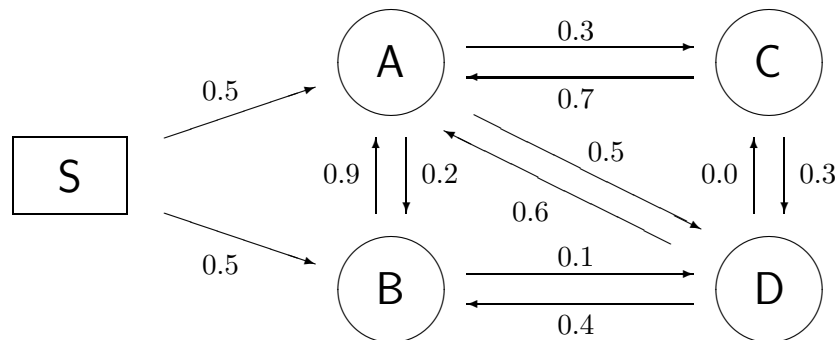
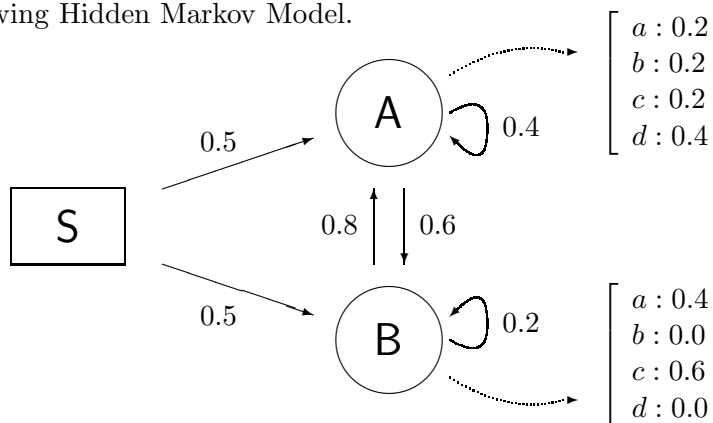   What are the probabilities of the following paths through the chain:

   (a) SABDACDB
   (b) SBADBABA
   (c) SACDBDCA

3. Given the following HMM and an emitted sequence `GATACC`, compute

   (a) the overall probability to observe the emitted sequence.
   (b) the most probable sequence of hidden states.

```
                                    A  =  0.3
                                    C  =  0.2
                  0.7               G  =  0.2
                                    T  =  0.3

                   A/T
         0.6                  0.06

                   0.38
  Start                        End
                        0.24

         0.4       G/C        0.02

                  0.6         A  =  0.1
                              C  =  0.4
                              G  =  0.4
                              T  =  0.1
```

4. Given the following Hidden Markov Model.

```
                                    [ a : 0.2
                     A                b : 0.2
        0.5                 0.4       c : 0.2
                                      d : 0.4 ]

   S              0.8    0.6

        0.5                           [ a : 0.4
                     B      0.2         b : 0.0
                                        c : 0.6
                                        d : 0.0 ]
```

What are the probabilities of the following emissions:

   (a) a
   (b) aa
   (c) aaa
   (d) $abcd$