Problems for the Hidden Markov Model

The HMM has \( n \) hidden random variables \( Y_1, \ldots, Y_n \) and \( n \) observed random variables \( X_1, \ldots, X_n \). There is a \( k \times k \)-transition matrix \( A = (a_{ij}) \) for each vertical transition \( Y_r \rightarrow Y_{r+1} \) and a \( k \times l \)-transition matrix \( B = (b_{ij}) \) for each horizontal transition \( Y_r \rightarrow X_r \). The model is defined by the formula

\[
p_{j_1j_2\cdots j_n} = \sum_{i_1i_2\cdots i_n} b_{i_1j_1}a_{i_1i_2}b_{i_2j_2}a_{i_2i_3}b_{i_3j_3}\cdots a_{i_{n-1}i_n}b_{i_nj_n}. \tag{1}
\]

1. Write the sum in (1) explicitly for \( n = 4, k = l = 2 \).
2. How many arithmetic operations are needed to evaluate this sum?
3. Fix \( k = l = 2 \) but vary \( n \). How many terms does the polynomial in (1) have? How many of them are leading terms in some term order?
4. The HMM an algebraic subvariety of dimension [???] in [???]. Fill in.
5. Fix \( n = 2, k = 2 \) and \( l = 3 \). Compute the prime ideal of the HMM.
6. Fix \( n = 3, k = 2 \) and \( l = 2 \). Compute the prime ideal of the HMM.
7. Type “Hidden Markov Model Biology” into google and discuss the results you see. Which biological applications does this model have?
8. Type “algebraic statistics” and “algebraic statistic” into google. Discuss the results you see. Describe a Hidden Markov Model which could be used to implement google’s popular “Did you mean”-feature.
Problems for the Genetic Disease Model

1. Determine the image of the monomial map

\[ \mathbb{R}^2 \to \mathbb{R}^5, \quad (p,q) \mapsto (p^4, p^3 q, p^2 q^2, pq^3, q^4). \]

What happens if we restrict \((p,q)\) to satisfy \(p, q \geq 0\) and \(p + q = 1\)?

2. Describe the set of all zeros of the ideal

\[ \langle x_1^2 - x_0 x_2, x_2^2 - x_1 x_3, x_3^2 - x_2 x_4 \rangle \]

in the probability simplex \(\Delta_4 = \{ x \in \mathbb{R}^5_{\geq 0} : x_1 + x_2 + x_3 + x_4 + x_5 = 1 \} \).

3. Pick your favorite 3 \times 5-matrix \(F\). Compute the image of the polynomial map \(\mathbb{R}^2 \to \mathbb{R}^3\) gotten by composing the monomial map in Problem 1 with your linear map \(F : \mathbb{R}^5 \to \mathbb{R}^3\). Draw a picture of the image.

4. The genetic disease model has three parameters \(f_0, f_1, f_2\). It can be defined as the image of a polynomial map \(\mathbb{R}^2 \to \mathbb{R}^3\) as above, where the 3 \times 5-matrix \(F\) has the specific form

\[
\begin{pmatrix}
4f_2^2 & 16f_1f_2 & 8f_0f_2 + 16f_1^2 & 16f_0f_1 & 4f_0^2 \\
8f_2^2 & 8(f_2 + f_1)^2 & 16f_1^2 + 16f_1f_2 + 16f_0f_1 & 8(f_1 + f_0)^2 & 8f_0^2 \\
4f_2^2 & 8f_2^2 + 8f_1^2 & 4f_2^2 + 16f_1^2 + 4f_0^2 & 8f_1^2 + 8f_0^2 & 4f_0^2
\end{pmatrix}
\]

Compute the image of this map with indeterminate parameters.

5. The strictly dominant genetic disease model has parameters \(f_0 = 0, f_1 = f_2 = 1\). Draw this model in the probability triangle \(\Delta_2 = \{ (z_0, z_1, z_2) \in \mathbb{R}^3_{\geq 0} : z_0 + z_1 + z_2 = 1 \} \).

6. For which values of the parameters \(f_0, f_1, f_2\) does the genetic disease model give a curve of degree less than four in \(\Delta_2\)? Give an example where this curve is a line, and give an example where it is a quadric.

7. The computer algebra system \texttt{Singular} claims an application to “Medicine” on its homepage (click on “Overview/Examples, then on “Applications”). What do you think about this application?