

Math 127, **Homework # 1**

due: Tuesday, January 30

- (1) Write a short essay (\leq one page) about the software **BLAST**. Who developed **BLAST** and why? What does this software do and who uses it? Illustrate the use of **BLAST** for one example chosen by you.
- (2) Consider the statistical model for $m = 3$ events with one parameter θ where the probabilities of the events are

$$f_1(\theta) = 0.2 + \theta, \quad f_2(\theta) = 0.3 + \theta, \quad f_3(\theta) = 0.5 - 2\theta.$$

Determine an explicit formula for the maximum likelihood estimate $\hat{\theta} = \hat{\theta}(u_1, u_2, u_3)$ as a function of the observed data u_1, u_2, u_3 .

- (3) Prove Varchenko's formula (Theorem 1.5) in the case $d = 2$ and $m = 4$.
- (4) Compute the maximum likelihood estimate \hat{p} for $u = (2, 3, 5, 7, 11, 13)$ in the toric model which is specified by the 3×6 -matrix

$$A = \begin{pmatrix} 2 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 2 \end{pmatrix}.$$

- (5) In a certain game of chance, a gambler tosses the same coin four times in a row, and the number of times heads come up are recorded. Hence the possible outcomes are 0, 1, 2, 3, or 4. We observe 1000 rounds of this game, and we record the outcomes in the data vector $u = (u_0, u_1, u_2, u_3, u_4)$, where u_i is the number of trials that had i heads. Hence $u_0 + u_1 + u_2 + u_3 + u_4 = 1000$. Suppose we are led to suspect that the gambler uses **two biased coins**, one in each of his sleeves, and he picks the coin to be used at random (with probabilities θ and $1 - \theta$). We wish to test this hypothesis using the data u . Formulate the corresponding algebraic statistics model, and explain how maximum likelihood estimation can be used to test our hypothesis for some u .
- (6) Formulate and implement the EM algorithm for the two-biased-coins model in (5). Pick your favorite positive integers u_0, u_1, u_2, u_3, u_4 and run the EM algorithm for different starting values (as in Example 1.16).