Statistics of Big Data, Spring 2015

Homework 4

Due date: 2 July 2015, by email or in my mailbox

1. **Weighted cardinality estimation (from Liran Katzir)**

   Extend the cardinality estimation scheme we saw in class to a weighted version. In this version when we encounter an element e_i, we also get w_i, and the objective is to estimate $w = \sum_i w_i$ (counting every element only once of course). In case all $w_i = 1$, we get the original cardinality estimation problem.

   Derive an explicit formula using m cells of memory as we did in class for the non-weighted case.

   Hint: Carefully review the properties of Beta distributions in Section 3.1 of Liran's writeup and try to figure out how to generalize the use of $U(0,1) = Beta(1,1)$ variables in the non-weighted case to our case. Note the property $Beta(\alpha,\beta) = 1 - Beta(\beta,\alpha)$ is a statement about the cumulative probabilities (hence the densities are mirror images around 0.5).

2. **Inferring rotations in Cryo-EM (from Yoel Shkolnisky)**

   Let $R_1,\ldots,R_n$ be unknown 3×3 rotation matrices. Suppose we are given the products $R_iR_j^T$ for $1 \le i < j \le n$. Let S be a $3n \times 3n$ matrix whose (i,j) block of size 3×3 is equal to $R_iR_j^T$, $i,j = 1,\ldots,n$.

   (a) As only blocks of S above the diagonal are given, show how to fill the blocks below the diagonal, that is, $R_iR_j^T$ for $1 \le j \le i \le n$.

   (b) Denote the three columns of R_i by

   $$R_i = \begin{pmatrix} | & | & | \\ R_i^{(1)} & R_i^{(2)} & R_i^{(3)} \\ | & | & | \end{pmatrix}.$$

   Show that the three vectors of length $3n$ given by

   $$v_j = \left(\left(R_1^{(j)}\right)^T, \ldots \left(R_n^{(j)}\right)^T\right)^T, j = 1,2,3,$$

   (that is, the concatenation of the first, second, and third columns of all the R_is) are eigenvectors of S corresponding to the same eigenvalue. What is the eigenvalue?

   (c) Using the previous result, explain how to recover $R_1,\ldots,R_n$ from S (or more precisely, how to recover $R_1O,\ldots,R_nO$, where O is an arbitrary orthogonal 3×3 matrix).

   (d) Denote the (i,j) block of size 3×3 of S by B_{ij}. Now, we randomly perturb some of the blocks B_{ij}, and assume that we have n^2 variables w_{ij}, $i,j = 1,\ldots,n$, indicating whether B_{ij} was perturbed or not. Namely,

   $$w_{ij} = \begin{cases} 1, & B_{ij} = R_iR_j^T \\ 0, & B_{ij} \text{ was perturbed} \end{cases}$$

Now we want to recover $R_1, \dots, R_n$ from the perturbed S and the information given by w_{ij}. Proposed approach:

i. Denote by $\tilde{S}$ the matrix whose blocks are $w_{ij} B_{ij}$ (that is, perturbed blocks are zeroed). What is $\tilde{S} v_j$?

ii. Now, try to rescale the rows of $\tilde{S}$ so that v_j will be an eigenvector of the rescaled version.

iii. How can you now recover the R_i's from this matrix?