Lab2 : Probability, IT and Optimization

1. Given *n* independently and identically distributed (iid) samples, (x_i, y_i) from a bivariate Gaussian source: (1) derived the likelihood function of the observations; (2) draw a scatter plot of the samples on a 2-D *x*-*y* plane; (3) can you plot a 3-D histogram of the data to get a sense about the unknown parameters? (4) derive the maximum likelihood (ML) estimates for the five parameters in the mean vector and covariance matrix; (5) given the data set in hw2-binormal.txt which lists 5000 pairs of data, estimate the above set of five parameters using ML as above with 500 pairs of samples, then the next 500, until all 5000 pairs are exhausted, compare the ten sets of estimated parameters. Can you guess what were the parameters used to generate the data? What can you say to summarize the results?
2. A discrete random vector $R=(r_1, r_2, \ldots, r_m)$ has a multinomial distribution as shown on Slide 16 of Lecture 2: (1) what is the expected mean vector, E[*R*], note: the sum of all *m* variables is *n*; (2) can you derive the expected covariance, $E\{(R-E[R])^t (R-E[R])\}$? (3) given a sequence of *n* observations from the above multinomial distribution the total number of samples that belongs Category *i* is $\text{Count}(r_i)=q_i$, what is the maximum likelihood for r_i, note: sum of p_j is equal to 1, you need to use this constraint and Lagrange theory to derive the formula for the ML estimate; (4) you can consider the 27 symbols in Lab1 follow a multinomial distribution, the results you have to compute unigram is an estimate of $P(\text{observing } i\text{th symbol})=p_i$;(5) for the case of observing pair of letters in Lab1, can we modeling this event as a multinomial distribution? how many distinct cases we need to consider, i.e. what is the value of m? (6) Are we estimating the maximum likelihood estimate of p in Lab1 for letter pairs? (7) any relation between entropy and probability?
3. Use the WSJ data in Lab1 to do the following: (1) Pick two events, current and following letters, and evaluate the average mutual information for them; (2) Compute the sample point-wise mutual information for the top and bottom five letter pairs you obtained in HW1; (3) What can you conclude about the results?

Center of Signal and Image Processing
Georgia Institute of Technology