Homework # 4

ORFE 524

1. Problem 2.2.33.

2. Let $X_1, \cdots, X_{n+4}$ be i.i.d. $N(\mu, 1/2)$. Suppose that we observe $S_1 = X_1, \cdots, S_n = X_n, S_{n+1} = X_{n+1} + 2X_{n+2}$, and $S_{n+2} = 2X_{n+3} - X_{n+4}$.

(a) Give a simple method without using bivariate density or bivariate regression formula to show that

$$E(X_{n+3}|S_{n+2}) = (2S_{n+2} + 3\mu)/5, \quad E(X_{n+4}|S_{n+2}) = (-S_{n+2} + 6\mu)/5.$$ 

Hint: $E(2X_{n+3} - X_{n+4}|S_{n+2}) = S_{n+2}$ and you need to obtain another equation.

(b) Using the result in (a), write down the E-step and M-step for computing the maximum likelihood estimator. Hint: In the E-step, there is no need to evaluate constant terms that will not be used in the M-step.

3. (Lumped Hardy-Weinberg data) Following the notation in Example 12, let $X_{i1}, X_{i2}$ and $X_{i3}$ be the indicators respectively for whether the genotype of the $i$th individual is AA, Aa and aa. Suppose we only observe the data $S_i = (X_{i1}, X_{i2} + X_{i3}), \ i = 1, \cdots, m$, (namely, we can not differentiate the genotype between Aa and aa) and $S_i = (X_{i1} + X_{i2}, X_{i3}), \ i = m + 1, \cdots, n$ (namely, we can not differentiate the genotype between AA and Aa). Write down the E-step and M-step of the EM algorithm.

4. Problem 1.3.1 (Do the problem only for $\delta_1, \delta_3, \delta_5, \delta_7$ to save time).

5. Problem 1.3.4 (a) and (b).