Fall 2010 Math 541b Exam

1. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent pairs of independent random variables X_i and Y_i, where X and Y have continuous distribution functions $F(x)$ and $G(x)$, respectively. We want to test the hypothesis $H_0 : F = G$ versus the alternative $H_1 : F \neq G$. Let $Z_i = X_i - Y_i$ and rank the numbers $|Z_1|, |Z_2|, \cdots, |Z_n|$ in increasing order, and let r_i be the rank of $|Z_i|$, so that the smallest absolute value receives the rank of 1.

   Define the Wilcoxon signed-rank statistic W as

   $$W = \sum_{i=1}^{n} \text{sign}(Z_i) r_i,$$

   where $\text{sign}(z) = 1$ if $z > 0$, $\text{sign}(z) = -1$ if $z < 0$, and $\text{sign}(z) = 0$ if $z = 0$.

   (a) Calculate $P(\text{sign}(Z_i)) = 1)$ and $P(\text{sign}(Z_i)) = 0)$ under the null hypotheses H_0.

   (b) Calculate the mean and the variance of W under the null hypothesis H_0.

   (c) Propose a test for H_0 versus H_1 that has approximate Type I error level α when the sample size is large.

2. Let $X_1, X_2, \cdots, X_n$ be i.i.d observations drawn from a mixture of two normal densities $\mathcal{N}(\mu_1, 1)$ and $\mathcal{N}(\mu_2, 1)$, where α and $1 - \alpha$ are the probabilities that a given observation is taken from the first and second normal distribution, respectively. We suppose that (μ_1, μ_2, α) are unknown.

   (a) Write down the likelihood function of the observed data $(X_1, X_2, \cdots, X_n)$ as a function of $\theta = (\mu_1, \mu_2, \alpha)$.

   (b) In order to design an EM algorithm to estimate θ, define missing data $(Z_1, Z_2, \cdots, Z_n)$ where $Z_i = 1$ if X_i is drawn from the first population $\mathcal{N}(\mu_1, 1)$, and $Z_i = 0$ if it comes from the second population. Write the likelihood function of the complete data $((X_1, Z_1), (X_2, Z_2), \cdots, (X_n, Z_n))$.

   (c) Design an EM algorithm to estimate the parameter vector θ.