Theoretical Foundations of the Analysis of Large Data Sets
Multiple Testing (2)

1. For n=100 and $q = 0.1$ generate 1000 replicates of the vector of p-values, according to the recipe of E.J.Candes and R. Foygel-Barber given to illustrate Theorem 1 in Lecture 9.

   a) Estimate the correlation matrix of the vector of p-values.

   b) Estimate the FWER of BH_q with $q = 0.1$ and compare to the value provided in class.

2. Generate 1000 replicates of the vector vector of random errors $\epsilon = (\epsilon_1, \ldots, \epsilon_{100})$ from the multivariate normal distribution $N(0, \Sigma)$ where all diagonal elements of Σ are equal to 1 and all off-diagonal elements are equal to

   a) 0.5

   b) 0.9

   c) are the same as the correlations calculated in Problem 1

   Now, for each of the above examples calculate 1000 replicates of the vector $X = \mu + \epsilon$, where

   i) $\mu = 0$

   ii) $\mu_1 = \ldots = \mu_5 = 3$ and $\mu_6 = \ldots = \mu_{100} = 0$

   iii) $\mu_1 = \ldots = \mu_{10} = 3$ and $\mu_{11} = \ldots = \mu_{100} = 0$

   – Use the values generated in i) to estimate the weak sense FWER for BH_q with $q = 0.1$ for all three correlations setups (a-c).

   – Use the values generated in ii) and iii) to estimate FDR for BH_q with $q = 0.1$ for all three correlations setups (a-c).

   – Use the values generated in ii) and iii) and plot the histograms of FDP and TPP (True Positive Proportion), and estimate the power and the standard deviation of FDP and TPP.

   Summarize the results referencing to the theoretical results provided in class.

3. Consider the data generated in Problem 2. For each of the above examples use the transformation $Y = \Sigma^{-1/2}X$ and apply SLOPE with $q = 0.1$, $\lambda = BH_q$ and $\sigma = 1$ to identify nonzero elements of the vector μ in the regression model $Y = \Sigma^{-1/2}\mu + \epsilon$.

   – Use the values generated in i) to estimate the weak sense FWER for $SLOPE_q$ with $q = 0.1$ for all three correlations setups (a-c).

   – Use the values generated in ii) and iii) to estimate FDR for $SLOPE_q$ with $q = 0.1$ for all three correlations setups (a-c).

   – Use the values generated in ii) and iii) and plot the histograms of FDP and TPP (True Positive Proportion), and estimate the power and the standard deviation of FDP and TPP. Compare to the results obtained in Problem 2.

Malgorzata Bogdan