Some topics in the analysis of Large Data Sets

Multiple Testing

1. Consider a low dimensional setup: $p = 20$, for $i = 1, \ldots, 10$, $\mu_i = \sqrt{2 * \ln(20/i)}$ and $\mu_{11} = \ldots = \mu_{20} = 0$. Compare FWER, FDR and Power (proportion of identified alternative hypotesis among all alternative hypotheses) of the Bonferroni and the Benjamini-Hochberg procedure.

2. Large dimensional set-up.

   Let $p = 5000$ and

   a) $\mu_1 = 1.2\sqrt{2 \log p}$, $\mu_2 = \ldots = \mu_p = 0$

   b) $\mu_1 = \ldots = \mu_{100} = 1.2\sqrt{2 \log \left(\frac{p}{100}\right)}$, $\mu_{101} = \ldots = \mu_{5000} = 0$

   c) $\mu_1 = \ldots = \mu_{1000} = 1.2\sqrt{2 \log \left(\frac{p}{1000}\right)}$, $\mu_{1001} = \ldots = \mu_{5000} = 0$.

   In each of the above settings compare FWER, FDR and Power of the Bonferroni and the Benjamini and Hochberg multiple testing procedures.

Malgorzata Bogdan