Theoretical Foundations of the Analysis of Large Data Sets

Multiple Testing, 27.04.2017, due - 18.05.2017

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 following procedures:

   a) Bonferroni

   b) Holm

   c) Hochberg

   d) Benjamini-Hochberg.

2. Large dimensional set-up: Using the settings from Problem 2 in Lab 3 compare FWER, FDR and Power (expected proportion of identified signals) of the following procedures:

   a) Bonferroni

   b) Holm

   c) Hochberg

   d) Benjamini-Hochberg.

   Additionally consider the setting $\mu_1 = \ldots = \mu_{100} = \sqrt{2 \log p}$, $\mu_{101} = \ldots = \mu_p = 0$.

Malgorzata Bogdan