Theoretical Foundations of the Analysis of Large Data Sets

Estimation risk

Due 15.06.2017

1. Simulate 500 realizations of the random vector $X = (X_1, \ldots, X_p) \sim N(\mu, \Sigma)$ where $p = 500$, $\Sigma_{i,i} = 1$, for $i \neq j$ $\Sigma_{i,j} = 0.4$ and the vector μ is as in Problem 1 on list 5.

   Compare the mean square error of the maximum likelihood estimate X with the extension of James-Stein estimate by Mary Ellen Bock (1975)

   $\mu_{MEB} = \left(1 - \frac{\tilde{p}-2}{X^T\Sigma^{-1}X}\right) X$, where $\tilde{p} = \frac{Tr(\Sigma)}{\lambda_{max}(\Sigma)}$.

2. Simulate 500 realizations of the random vector $X = (X_1, \ldots, X_p) \sim N(\mu, I)$ where $p = 500$ and the vector μ is equal to

   a) $\mu_1 = \ldots = \mu_5 = 3.5$, $\mu_6 = \ldots = \mu_{500} = 0$
   b) $\mu_1 = \ldots = \mu_{30} = 2.5$, $\mu_{31} = \ldots = \mu_{500} = 0$
   c) $\mu_1 = \ldots = \mu_{100} = 1.8$, $\mu_{101} = \ldots = \mu_{500} = 0$
   d) $\mu_1 = \ldots = \mu_{500} = 0.4$
   e) $\mu_i = 3.5 * i^{-1/2}$
   f) $\mu_i = 3.5 * i^{-1}$

   For each of these examples compare the mean square error of the

   a) maximum likelihood estimator
   b) James-Stein estimator
   c) hard-thresholding rule based on the Bonferroni correction with the nominal FWER equal to 0.1 (i.e. MLE when Bonferroni rejects H_{0i}, 0 otherwise)
   d) hard-thresholding rule based on the BH procedure with the nominal FDR equal to 0.1.

Malgorzata Bogdan