Theoretical Foundations of the Analysis of Large Data Sets

Laboratory 1, 02.03.2017, Due 16.03.2017

Bonferroni correction

1. For $t \in [0.2, 4]$ graphically compare $g_1(t) = 1 - \Phi(t)$ with $g_2(t) = \frac{\phi(t)}{t}$ and $g_3(t) = \phi(t)\frac{t}{t^2+1}$. Make two plots: represent $g_i(t)$ directly and plot the ratios $g_1(t)/g_2(t)$ and $g_1(t)/g_3(t)$.

2. For $p \in (10^2, 10^9)$ graphically compare $g_1(\alpha, p) = \Phi^{-1}\left(1 - \frac{\alpha}{2p}\right)$ for $\alpha = 0.01,\ 0.1$ and 0.5 with $c(p) = \sqrt{2 \log p}$ and with $g_2(\alpha, p) = \sqrt{B - \log B}$, where $B = 2 \log\left(\frac{2p}{\alpha}\right) - \log(2\pi)$. Make two plots: represent these functions directly and plot the ratios $g_1(\alpha, p)/c(p)$ and $g_1(\alpha, p)/g_2(\alpha, p)$. Use the log scale on the x axis.

3. Let $p = 10^8$ and $Y_1, \ldots, Y_p$ be iid from $N(0, 1)$. Simulate five trajectories of $M_k = max_{j \in \{1,\ldots,k\}}|Y_i|$, $k = 10^{ind}$, where $ind \in \{1, \ldots, 8\}$ and compare them to the graph of $g_k = \sqrt{2 \log k}$. Draw also the graph of trajectories of M_k/g_k. Use the log scale on the x axis.

4. For $p = 5000$ estimate the power of the Bonferroni, Fisher and chi-square tests against alternatives:

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

   b) $\mu_1 = \ldots = \mu_{1000} = 0.15\sqrt{2 \log p}$, $\mu_{1001} = \ldots, \mu_{5000} = 0$

Malgorzata Bogdan