

M9750 Robustní a neparametrické statistické metody

cvičení 4 - robustní odhady parametru polohy

1. Nagenertejte si data - uvažujte náhodný výběr z normálního rozdělení $N(0, 1)$ o rozsahu $n = 30$. Uvažujte Huberův odhad daný funkcí $\psi_H(x)$, resp. $\rho_H(x)$ pro $k = 1, 2, 5$.
 - (a) Nejprve jej spočítejte jako řešení rovnice $\sum_{i=1}^n \psi_H(X_i - \hat{\theta}) = 0$.
 - (b) Poté jako řešení minimalizační úlohy definované funkcí $\rho_H(x)$.
 - (c) Nakreslete grafy obou předchozích funkcí.
 - (d) Co způsobuje volba k ?
 - (e) Uvažujte nyní data pocházející z Cauchyho rozdělení a najděte odpovídající Huberův odhad.
 - (f) Body (a)–(e) proveďte i pro další funkce $\psi(x)$ uvedené na přednášce.
2. Datový soubor `chem` v knihovně `MASS` obsahuje výsledky 24 měření obsahu mědi ve vzorku celozrnné mouky.
 - (a) Proveďte exploratorní analýzu dat.
 - (b) Odhadněte průměrný obsah mědi ve vzorku pomocí průměru, mediánu, Huberova odhadu, 5% a 10% useknutého a winsorizovaného průměru a Hodgesova-Lehmannova odhadu.
 - (c) Tyto výsledky porovnejte s případem, kdy odlehlé pozorování vyloučíte z analýzy.
 - (d) Tyto výsledky porovnejte s případem, kdy odlehlé pozorování opravíte na správnou hodnotu.
3. Zkoumejte chování předchozích odhadů pro různá rozdělení na základě náhodného výběru o rozsahu $n = 30$. Uvažujte následující rozdělení (vždy s nulovou střední hodnotou a jednotkovým rozptylem; POZOR na parametrizaci v R): normální, logistické, t-rozdělení s 3 stupni volnosti a Cauchyho (s parametry 0,1).
 - (a) Nejprve vygenerujte náhodný výběr o rozsahu n z daného rozdělení a proveďte dané odhady. Celý postup opakujte 10 000 krát.
 - (b) Odhadněte střední hodnotu, rozptyl a střední čtvercovou chybu všech odhadů a výsledky porovnejte.

Aby všem vycházely stejné výsledky, nastavte před každým během simulace generátor náhodných čísel pomocí příkazu `set.seed(1234)`.

Funkce, které by se mohly hodit: `optimize`, `uniroot`.