

## M9750 Robustní a neparametrické statistické metody

### cvičení 1 - chování klasických odhadů a testů při porušení normality

1. Zkoumejte chování dvou odhadů střední hodnoty pro různá rozdělení: výběrového průměru a výběrového mediánu 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, jaké je rozdělení obou odhadů (histogram, jádrový odhad,...). Odhadněte střední hodnotu, rozptyl a střední čtvercovou chybu obou odhadů a výsledky porovnejte.
  - (c) Jaké by mělo být přesné (asymptotické rozdělení) těchto odhadů? Výsledek porovnejte se simulacemi z bodu (b).
2. Zkoumejte chování dvou testů o nulovosti střední hodnoty daného rozdělení: t-testu a jednovýběrového Wilcoxonova testu 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 daný test. Celý postup opakujte 10 000 krát. Zaznamenejte si  $p$ -hodnoty a počet zamítnutí nulové hypotézy. Hladinu významnosti  $\alpha$  volte 5 procent.
  - (b) Odhadněte skutečnou hladinu významnosti jako podíl zamítnutí nulové hypotézy mezi všemi 10 000 simulacemi. Dosahují všechny testy předepsané hladiny významnosti?
  - (c) Jaké rozdělení mají  $p$ -hodnoty daných testů?
  - (d) Uvažujte nyní parametr polohy všech rozdělení roven 0.5 a proveďte znovu body (a) – (c). Co teď znamená podíl zamítnutí nulové hypotézy? Který test je "lepší" pro dané rozdělení?

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

Funkce, které by se mohly hodit: `rnorm`, `dnorm`, `rlogis`, `dlogis`, `rt`, `dt`, `rcauchy`, `dcauchy`, `mean`, `median`, `t.test`, `wilcox.test`.