

# BAYESIÁNSKÁ ANALÝZA – DOMÁCÍ SKUPINOVÝ ÚKOL Č. 1

Pro účely zpracování tohoto úkolu si vytvořte pracovní skupinu o velikosti max. 4 osoby. Úkoly zpracujte a odevzdejte do stanoveného termínu. Veškeré podklady (např. skripty Matlabu) a komentáře, odevzdávejte do příslušné odevzdárny.

**Termín odevzdání:** 21. 10. 2016 (včetně)

## Zadání úkolu

1. (*Tvorba umělých datových souborů*) Vytvořit si vlastní datový soubor je užitečné pro pochopení vlastností modelu a pro analýzu kvality počítačových algoritmů. Pokud si totiž zvolíme hodnoty parametrů, víme jaké hodnoty by nám měly přibližně vycházet, použijeme-li tu či onu ekonometrickou metodu.
  - (a) Vygenerujte umělý datový soubor v rámci normálního lineárního regresního modelu:
    - Zvolte si hodnoty  $\beta_1$ ,  $\beta_2$ ,  $h$  a  $N$  (např.  $\beta_1 = 1$ ,  $\beta_2 = 2$ ,  $h = 1$  a  $N = 100$ ).
    - Vygenerujte  $N$  hodnot pro vysvětlující proměnnou z rozdělení dle vašeho výběru (např. proveďte  $N = 100$  náhodných výběrů z uniformního rozdělení  $U(0, 1)$ ).
    - Vygenerujte  $N$  chybových členů pomocí  $N$  i.i.d. výběrů z rozdělení  $N(0, h^{-1})$ .
    - Vytvořte vektor vysvětlované proměnné  $y$  v rámci NLRM (tj.  $y_i = \beta_1 + \beta_2 x_i + \epsilon_i$  pro  $i = 1, \dots, N$ ).
    - Vytvořte bodový graf (graf XY) takto vygenerovaných dat.
  - (b) Na základě vygenerovaných dat odhadněte posteriorní střední hodnotu a posteriorní směrodatnou odchylku pro parametry tohoto modelu při použití přirozeně kojugované normální-gama apriorní hustoty s  $\underline{\beta} = (0, 1)'$ ,  $\underline{V} = I_2$ ,  $\underline{s}^{-2} = 1$  a  $\underline{\nu} = 10$ .
  - (c) Vykreslete posteriorní hustoty pro všechny odhadované parametry (buď ze znalosti analytických posteriorních hustot nebo na základě Monte Carlo integrace, kdy použijte histogram nebo funkci `ksdensity` pro vykreslení jádrové hustoty vašeho výběru).
  - (d) Spočítejte Bayesův faktor porovnávající model  $M_1 : \beta_2 = 0$  s  $M_2 : \beta_2 \neq 0$ .
  - (e) Vykreslete predikční hustotu pro pozorování s hodnotou  $x^* = 0.5$ .
  - (f) Odhadněte model (posteriorní střední hodnoty a směrodatné odchylky parametrů  $\beta$  a  $h$ ) za použití neinformativní apriorní hustoty. Spočítejte Bayesův faktor porovnávající model  $M_1 : \beta_2 = 0$  s  $M_2 : \beta_2 \neq 0$ . Porovnejte a komentujte výsledky s výsledky dosaženými v části (1d).
  - (g) Spočítejte 99% HPDI pro odhadované parametry s využitím jak informativní, tak i neinformativní apriorní hustoty a použijte jej pro ověření hypotézy, že  $\beta_2 = 0$ . Porovnejte své výsledky s výsledky dosaženými v částech (1d) a (1f).

2. Archiv souborů `karate_LK2013.zip` obsahuje data z experimentálního měření síly úderu v karate. Sílu úderu (zjišťována jako průměr tří měření úderu pravou rukou na měřící vestu) mohou ovlivňovat různé faktory jednotlivce (např. věk, výška, váha, zkušenosti z karate). Zajímalo by nás tedy, jak vlastně jednotlivé faktory ovlivňují sílu úderu karatistů. V datech obsažené proměnné jsou:

- *VEK* = věk jedince, hodnota udávána v letech
- *VYSKA* = výška jedince, udávána v centimetrech
- *VAHA* = hmotnost jedince, udávána v kilogramech
- *POHLAVI* = pohlaví jedince (žena=0, muž=1)
- *ZKUSENOSTI* = počet let, který se jedinec věnuje karate
- *TECHNICKY\_STUPEN* = jedná se o hodnotu technického stupně, kterého daný jedinec dosáhl (čím vyšší hodnota, tím vyšší stupeň)
- *PRUMER* = vysvětlovaná proměnná, která je průměrem 3 úderů pravou rukou na měřící vestu (měřeno v joulech).

- (a) Odhadněte lineární regresní model se všemi vysvětlujícími proměnnými (Model A), věcně interpretujte dosažené výsledky (věcný význam parametrů). K odhadu využijte přirozeně konjugovanou apriorní hustotu (volba použití informativní nebo neinformativní hustoty je na vás). Testujte a komentujte významnost jednotlivých parametrů (s využitím Bayesova faktoru a 95% HPDI).
- (b) Testujte v Modelu A (v modelu vícenásobné regrese) hypotézu, že muži mají větší sílu úderu než ženy (s využitím Bayesova faktoru nebo 95% HPDI).
- (c) Modifikujte model vícenásobné regrese (Model A) tak, aby obsahoval jen statisticky významné proměnné (na základě vhodně zvolené hodnoty Bayesova faktoru). Věcně interpretujte dosažené výsledky (věcný význam parametrů) a komentujte, jestli se odhady parametrů proměnných změnily oproti těm z vícenásobné regrese se všemi proměnnými.
- (d) Na základě výsledného modifikovaného modelu (se všemi významnými proměnnými, označený jako Model B) otestujte hypotézu, že každý technický stupeň navíc zvyšuje sílu úderu nejméně o 6 joulů. Svůj přístup k testování a vyhodnocení hypotézy jasně popište a zdůvodněte.
- (e) Modifikujte dále Model B tak, abyste dokázali otestovat, že vliv zkušeností na sílu úderu se projevuje jen u lidí, kteří se věnují karate více než 5 let. Takto modifikovaný model odhadněte a hypotézu vhodně otestujte.