BAYESIÁNSKÁ ANALÝZA – CVIČENÍ 3

Toto cvičení je založeno na znalosti prvních tří kapitol z učebnice Koop (2003): *Bayesian econometrics*, případně na odpovídajících kapitolách podkladového učebního textu *Bayesiánská analýza*.

Co bude náplní cvičení?

✎ Odhad a posteriorní analýza normálního lineárního regresního modelu s přirozeně konjugovanou apriorní hustotou (více vysvětlujících proměnných).

✎ Citlivostní analýza volby apriorní husoty pravděpodobnosti.

✎ Odhad a posteriorní analýza na příkladech s využitím reálných dat.

Zadání příkladů

K řešení příkladů využijte již hotové funkce, případně si vytvořte své vlastní. Poslední příklady jsou z knížky Hill, Griffiths, Lim (2008): *Principles of Econometrics*. Pokud hovoříme o testování hypotéz, má se za to, že tento test provedeme za pomocí porovnávání modelů.

1. *Empirická ilustrace z třetí kapitoly Koop (2003)*, data o prodeji domů (soubor `HPRICE.TXT`). Projděte si řešení příkladu a podpůrné funkce a diskutujte nejasnosti. Rozšiřte analýzu o další vysvětlující proměnné a řešte příklad s případnými vlastními apriorními představami o hodnotách apriorních hyperparamterů.

2. *Odhad a Monte Carlo integrace v modelu vícenásobné regrese*.

   (a) Vytvořte umělý datový soubor pro velikosti $N = 100$ pro normální lineární regresní model s úrovňovou konstantou a jednou vysvětlující proměnnou. Úrovňovou konstantu položte rovnu 0 a koeficient sklonu regresní přímky položte roven jedné a $h = 1$. Vysvětlující proměnnou vezměte z uniformního rozdělení $U(0, 1)$.

   (b) Spočítejte posteriorní střední hodnotu a směrodatnou odchylku pro tato data 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 hustotu pro β_2, a to jak z definice její posteriorní marginální hustoty, tak pomocí Monte Carlo integrace (užijte histogram nebo funkci `ksdensity` pro vykreslení jádrové hustoty vašeho výběru). Pro různě velké velikosti výběru spočítejte numerickou standardní chybu aproximace střední hodnoty parametrů.

   (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 $x_2^* = 0.5$.

   (f) Proveďte citlivostní analýzu apriorní hustoty nastavením $\underline{V} = cI_2$ a opakujte kroky (2b), (2d) a (2e) pro hodnoty $c = 0.01, 1.0, 100.0, 1\times10^6$. Diskutujte citlivost posteriorní hustoty, Bayesova faktoru a predikční hustoty rozdělení.

   (g) Spočítejte posteriorní střední hodnotu a směrodatnou odchylku vektoru parametrů β za použití neinformativního prioru.

   (h) Spočítejte 99% HPDI pro β_2 užitím neinformativního prioru a užijte jej pro ověření hypotézy, že $\beta_2 = 0$. Porovnejte své výsledky s výsledky dosaženými v části (2d).

   Poznámka: Můžete samozřejmě rozšířit model o další proměnné (při generování umělých dat) a měnit jejich nastavení včetně volby priorů..

3. Soubor `cocaine.m` obsahuje 56 pozorování proměnných vztahujících se k prodeji kokainu v severovýchodní Kalifornii v období 1984-1991. Data jsou podmnožinou dat použitých ve studii Culkins, J.P. a Padman, R. (1993): „Quantity Discounts and Quality Premia for Illicit Drugs," *Journal of the American Statistical Association*, 88, 748-757. Proměnné jsou

   - $price$ = cena za gram kokainu v rámci dané transakce;
   - $quant$ = počet gramů kokainu prodaných v dané transakci;
   - $qual$ = kvalita kokainu vyjádřená jako procento čistoty;
   - $trend$ = časová proměnná s hodnotami od 1984=1 až po 1991=8.

   Předpokládejme regresní model

   $$price = \beta_0 + \beta_1 quant + \beta_2 qual + \beta_3 trend + \epsilon.$$

   (a) Jaká znaménka koeficientů byste očekávali u parametrů β_1, β_2 a β_3?

   (b) Odhadněte daný model (předpokládáme, že se jedná o NLRM s přírozeně konjugovanou apriorní hustotou). Zvolte si vhodné hyperparametry dle vašich zkušeností. Jsou znaménka parametrů v souladu s vašim očekáváním?

   (c) Říká se, že čím větší objem obchodů, tím větší riziko, že vás dostihne ruka zákona. Prodejci tak jsou ochotni akceptovat nižší cenu, pokud prodávají větší množství. Pokuste se testovat tuto hypotézu.

   (d) Ověřte hypotézu, že kvalita kokainu nemá vliv na jeho cenu.

   (e) Jaká je průměrná roční změna ceny kokainu? Zamyslete se nad tím, proč by se měla cena takto měnit.

4. Každé ráno mezi 6:30 a 8:00 opouští Bill Melbournské předměstí Carnegie, aby se dostal do práce na University of Melbourne. Čas, který Bill stráví cestou do práce, $time$, závisí na času odjezdu, $depart$, počtu červených světel na semaforech, $reds$ a počtu vlaků, kvůli kterým musí čekat na Murrumbeenském přejezdu, $trains$. Pozorování těchto proměných je celkem získáno za 231 pracovních dní v roce 2006 a jsou obsahem souboru `commute.m`. Proměnná $time$ je měřena v minutách, $depart$ je počet minut po 6:30, které uplynou než Bill vyrazí z domu.

   (a) Odhadněte rovnici
   $$time = \beta_0 + \beta_1 depart + \beta_2 reds + \beta_3 trains + \epsilon.$$

   (b) Jaká znaménka koeficientů byste očekávali u parametrů β_1, β_2 a β_3?

   (c) Otestujte hypotézu, že každé červené světlo zpozdí Billa nejméně o 2 minuty.

   (d) Testujte hypotézu, že čas odjezdu nemá vliv na čas strávený cestováním.

   (e) Otestujte hypotézu, čas cestování navíc díky čekání na jednom semaforu je stejný jako čas čekání průjezdu jednoho vlaku.