BAYESIÁNSKÁ ANALÝZA – CVIČENÍ 4

Toto cvičení je založeno na znalosti čtvrté kapitoly z učebnice Koop (2003): *Bayesian econometrice*, případně na odpovídající kapitole podkladového učebního textu *Bayesiánská analýza*.

Co bude náplní cvičení?

- ✎ Odhad a posteriorní analýza normálního lineárního regresního modelu s nezávislou normální-gama apriorní hustotou a s omezeními ve tvaru nerovnosti.
- ✎ Osvojení si Gibbsova vzorkovače a importance sampling.
- ✎ Odhad a posteriorní analýza na příkladech s využitím reálných dat.

Zadání příkladů

1. *Gibbsův vzorkovač a jeho vlastnosti*: Předpokládejme elementární příklad modelu, kdy při odhadu jeho parametrů získáváme posteriorní hustotu odpovídající dvourozměrnému normálnímu rozdělení (to, že máme známou posteriorní hustotu bude sloužit k tomu, že výsledky simulátorů budeme schopni porovnat s výsledky analytickými):
$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$
kde $|\rho| < 1$ je známá posteriorní korelace mezi parametry θ_1 a θ_2.
   (a) Začněte si vytvářet skript, kdy si nejprve zadefinujete počet generovaných vzorků S (např. $S = 10000$), vytvořte vektor středních hodnot μ jakožto nulový vektor (později ho můžete změnit), korelační koeficient ρ a kovarianční matici Σ, kdy na diagonále budou jedničky a mimo ní koeficienty ρ (díky tomu odpovídá kovarianční matice přímo i korelační matici).
   (b) Vytvořte si vlastní funkci, která využívá Monte Carlo integraci k výpočtu posteriorní střední hodnoty, směrodatné odchylky parametrů θ_1 a θ_2 a numerické standardní chyby (NSE) pro střední hodnotu a rozptyl odhadu parametrů, popřípadě využijte dodanou funkci `MC_int.m`. V rámci této funkce můžete využít generátor náhodných čísel z vícerozměrného rozdělení buď přímo dostupný v rámci statistického toolboxu Matlabu (funkce `mvnrnd.m`) nebo funkci LeSageho ekonometrického toolboxu `norm_rnd.m`. *Poznámka:* Je třeba si uvědomit, že Monte Carlo integrace nevyužívá nic jiného než zákon velkých čísel, který říká, že výběrový průměr nějaké funkce parametrů bude konvergovat ke střední hodnotě této funkce parametrů pro rostoucí velikost vygenerovaných vzorků rozdělení, z něhož tyto parametry pocházejí. NSE je definováno jako $\sqrt{\frac{\sigma_g^2}{S}}$, kde σ_g^2 je rozptyl funkce parametrů (nahrazován odhadem), který nás zajímá.
   (c) Vytvořte funkci, která budeý využívat Gibbsův vzorkovač k výpočtu střední hodnoty a směrodatné odchylky parametrů θ_1 a θ_2 (využijte vlastností vícerozměrného normálního rozdělení pro výpočet odpovídajících podmíněných hustot pravděpodobnosti).
      i. Vyjděte z funkce sdružené hustoty normálního rozdělení (dvojrozměrné), se střední hodnotou $\mu = (\mu_1, \mu_2)'$ a kovarianční maticí Σ definovanou v úvodu příkladu:
$$\frac{1}{(2\pi)^{\frac{2}{2}}} |\Sigma|^{\frac{-1}{2}} \exp\left[-\frac{1}{2} (\theta - \mu)' \Sigma^{-1} (\theta - \mu) \right]$$
      ii. Odvoďtě podmíněné hustoty pro $\theta_1|\theta_2$ a $\theta_2|\theta_1$
      iii. Pokud jste šikovnější, odvoďte podmíněné hustoty pro obecné k-rozměrné normální rozdělení s arbitrárním dělením vektoru středních hodnot a kovarianční matice (zde je dobré využít teorém o inverzi a determinantu dělené matice z přílohy učebního textu či Koopovy učebnice). Důkaz inverze dělené matice lze nalézt např. zde a odvození podmíněných hustot zde.

iv. Výsledné funkce podmíněných hustot pravděpodobnosti jsou $\theta_1|\theta_2\ N(\mu_{(1|2)}, \Sigma_{(1|2)})$ a $\theta_2|\theta_1\ N(\mu_{(2|1)}, \Sigma_{(2|1)})$:

$$\mu_{(1|2)} = \mu_1 + \rho * (\theta_2 - \mu_2)\Sigma_{(1|2)} = 1 - \rho^2$$

(d) Nastavte $\rho = 0$ pro porovnání výsledků z části (a) a (b). Kolik replikací je nutných k odhadu středních hodnot a směrodatných odchylek parametrů θ_1 a θ_2 s přesností na dvě desetinná místa?

(e) Zopakujte část (c) pro $\rho = 0.5, 0.9, 0.95, 0.99, 0.999$. Jak velikost korelace ovlivní výkonnost Gibbsova vzorkovače? Pro srovnání vykreslete průběh Gibbsova vzorkovače pro prvních 50 a 1000 iterací a rovněž i výslednou sekvenci vzorků po odstranění počátečních S_0 vzorků, a to pro odlehlé počáteční hodnoty parametrů.

(f) Obohaťte své programy o výpočet numerické standardní chyby a v případě Gibbsova vzorkovače i o Gewekovu CD diagnostiku, konvergenční diagnostiku Gelmana a Rubina případně některý z typů konvergenčních diagnostik Brookse a Gelmana (nejlepší je vytvoření vlastních funkcí). Pro Gewekovu konvergenční diagnostiku využijte funkce `momentg.m` z LeSageho ekonometrického toolboxu případně funkce odhadu spektrální hustoty Matlabu (Signal processing toolboxu - funkce `psd`). Zopakujte části (c) a (d) s příslušnými kovergenčními diagnostikami. Je NSE dobrou charakteristikou přesnosti aproximace? Bude CD přesně signalizovat dosažení konvergence Gibbsova vzorkovače ve všech případech?

i. Funkce `momentg.m` vychází z odvození, které lze nalézt v rámci příkladu 11.14 na straně 144 z knihy Koop, Poirier, Tobias (2007): *Bayesian Econometric Methods*. Jedná se o způsob výpočtu NSE v rámci Gibbsova vzorkovače za předpokladu existence korelace mezi výběry z Gibbsova vzorkovače. Touto metodou (s využitím informace o korelacích resp. kovariancích v rámci vzorku) se tak aproximuje výpočet spektrální hustoty v bodě 0, která je přímo vztažena k rozptylu dané časové řady (zbaveného vlivu možných autokorelací).

ii. Konstrukce konvergenčních diagnostik není v zásadě nijak obtížná, neboť jde jen o vytvoření funkcí s odpovídajícími algoritmy. Zajímavým aspektem je však případ, kdy je potřeba generovat více řetězců s rozptýlenými počátečními hodnotami. Obvyklý způsob je ten, že se posteriorní rozdělení aproximuje směsí (kompozice) vícerozměrných hustot normálního rozdělení, se středními hodnotami odpovídajícími nalezeným modům posteriorní hustoty a rozptylem vycházející z druhých parciálních derivací jader posteriorních hustot vyhodnocených v těchto modech.

$$\widehat{P}(x) = \sum_{k=1}^{K} \omega_k (2\pi)^{-d/2} |\Sigma_k|^{-1/2} \exp\left(-\frac{1}{2}(x-\mu_k)'\Sigma_k^{-1}(x-\mu_k)\right)$$

Takto konstruované kompozice mají samozřejmě své váhy ω_k, které se získají tak, aby aproximativní hustota $\widehat{P}$ odpovídala původní hustotě P svou funkční hodnotou právě v oněch modech, tedy $\widehat{P}(\mu_k) = P(\mu_k)$ pro $k = 1, \ldots, K$. Pokud jsou mody dobře od sebe odděleny, je ω_k přibližně proporcionální $|\Sigma_k|^{1/2} P(\mu_k)$. V našem případě máme unimodální rozdělení a není tak další mody třeba hledat. Vzorky z přerozptýleného rozdělení (overdispersed distribution) de pa získají tak, že se generují náhodné výběr z aproximativní kompozice normálních rozdělení a následně se každý vektor násobí skalární náhodnou veličinou, kdy je obvyklá volba náhodné veličiny z χ^2_η dělená η. Výsledkem je pak v podstatě nové rozdělení, které odpovídá kompozici vícerozměrných t-rozdělení.

$$\overline{P}(x) \propto \sum_{k=1}^{K} \omega_k |\Sigma_k|^{-1/2} \left(\eta + (x-\mu_k)'\Sigma_k^{-1}(x-\mu_k)\right)^{-(d+\eta)/2}$$

Konzervativn volba je $\eta = 1$ (Cauchyho rozděleni), prakticky je obvyklá volba $\eta = 4$. Gelman a Rubin dále zavádějí k metodu *importance resampling* pro odstranění výběrů z příliš odlehlých oblastí posteriorní hustoty. Nejedná se o nic složitého, fakticky jde o metodu bootstrapu aplikovanou způsobem, kdy se vygeneruje N výběrů z vícerozměrného t-rozdělení (škálovaná náhodná veličina z vícerozměrného normálního rozdělení) a pro každý výběr se spočítá váha $P(x)/\overline{P}(x)$ (stačí počítat s jádry hustot). Následně se vybere jeden vzorek s pravděpodobnostmi danými oněmi spočítanými vahami. Tento postup výběru (bez nahrazování) se postupně opakuje (tzn. ve druhém kroku se vybírá z N-1 zbylých výběrů) až dokud nemáme potřebný počet m vzorků počátečních hodnot parametrů.

iii. Podrobnosti k metodě Brookse a Gelmana jsou obsahem prezentace „Konvergenční diagnostiky“ a samozřejmě i jejich (vcelku čtivého) článku.

2. *Importance sampling* Cílem tohoto úkolu je osvojení techniky importance sampling a pochopení vlastností tohoto postupu. Předpokládejme zcela jednoduchý model s jediným parametrem θ, jehož posteriorní hustota odpovídá $N(0, 1)$.

   (a) Vytvořte program počítající posteriorní střední hodnotu a směrodatnou odchylku parametru θ pomocí Monte Carlo integrace.

   (b) Vytvořte program počítající posteriorní střední hodnotu a směrodatnou odchylku parametru θ za použití techniky importance sampling a spočítejte rovněž střední hodnotu a směrodatnou odchylku použitých vah. Jako importance function použijte funkci hustoty pravděpodobnosti $t(0, 1, \nu)$.

      i. Metoda importance sampling využívá generování vzorků ze známé kandidátské hustoty (importance function) a zahrnuje počítání vah jakožto podílů jader posteriorní hustoty a kandidátské hustoty vyhodnocené ve vygenerovaném kandidátovi. Výpočet středních hodnot funkcí parametrů, které nás zajímají pak vyžaduje provedení Monte Carlo integrace, kde místo aritmetických průměrů využíváme vážené průměry s vypočtenými vahami z „importance sampleru“.

   (c) Proveďte Monte Carlo integraci a importance sampling pro různé hodnoty ν, např. $\nu = 2$, 5 a 100 pro daný počet replikací (např. $R = 10$). Porovnejte přesnost odhadů pro oba algoritmy a různou volbu ν. Všimněte se co se děje s váhami při zvyšujícím se ν.

   (d) Zopakujte část (c) při využití $t(3, 1, \nu)$ jakožto importance function. Diskutujte faktory ovlivňující přesnost importance sampling ve světle skutečnosti, že importance function jen slabě aproximuje posteriorní hustotu.

   (e) Vyzkoušejte i jiné importance function, např. $U(a, b)$, měňte i počet replikací R.

   (f) Vypočítat požadované momenty posteriorní hustoty s využitím importance samplingu je vcelku jednoduchá záležitost. Existuje ale způsob, jak získat pomocí importance samplingu a s využitím umění generovat vzorky z importance funkce i přímo vzorky odpovídající posteriornímu rozdělení?

      i. Myšlenka, jak na to není zas tak obtížná. Co máme k dispozici jsou jednak výběry z importance funkce, ale i velmi důležité váhy. Kdyby tedy existovala metoda, která by pro nějaký vzorek dat dokázala vybírat náhodné výběry, a to navíc i tak, že bychom mohli definovat s ajkou pravděpodobností se ten který prvek původního výběru může vybrat, tak máme vyhráno.

      ii. Touto metodou je metoda bootstrapu. V základním principu se nejedná o nic jiného, než o metodu, která vygeneruje z nějakého mepirického rozdělení náhodný výběr (o požadované délce) tak, že pravděpodobnost výběru z každého vzorku je stejná. Máme-li tedy výběr o velikosti 100, každý z těchto prvků se bude vybírat s pravděpodobností 1 %, příčemž si sami můžeme zvolit, jestli se bude jednat o výběr s nahrazováním (to znamená, že po výběru nějakého prvku z empirického rozdělení tento prvek zůstává a v dalším kole může být vybrán znovu) prvku nebo bez nahrazování (když nějaký prvek vybereme, už jej v druhém kole vybrat nemůžeme). V našem případě využijeme samozřejmě variantu s nahrazováním, což je právě princip bootstrapu (druhý případ by odpovídal situaci, kdybychom potřebovali vybrat náhodnou podmnožinu výběru). Takovýto boostrapový generátor si zvládneme vytvořit sami: stačí umět generovat čísla z uniformního rozdělení (pro 100 pozorování např. na intervalu 0 až 100, funkce `rand` transformovaná do požadovaného intervalu), která zaokrouhlíme na celá čísla (funkce `ceil`) a získáme tak index prvku, který z našeho vektoru dat máme vybrat. To opakuje tolikrát, kolik budeme chtít výběrů (počet výběrů nemusí odpovídat počtu prvků původního vektoru).

      iii. V našem případě ale nechceme vybírat prvky z kandidátské hustoty se stejnou pravděpodobností. Chtěli bychom předchozí proceduru modifikovat tak, aby pravděpodobnosti, s jakými budeme vybírat jednotlivé prvky vektoru byly proporcionální vahám získaným z importance samplingu. I zde bychom si příslušnou funkci mohli naprogramovat sami, např. tak, že si jednak nanormujeme váhy na jedničkový součet, vyhodíme prvky s nulovými vahami a vytvoříme nový umělý vzorek, ve kterém počet každého z prvků zvýšíme proporcionálně jeho váze (pravděpodobnosti). Následně pak na takto

rozšířený vzorek aplikujeme metodu bootstrapu popsanou výše (i když každý prvek má stejnou pravděpodobnost výběru, jeho početní zastoupení zajistí požadovanou pravděpodobnost danou původními vahami).

iv. Abychom si ale ulehčili práci, můžem využít funkci `bootstrp` Statistického toolboxu. Tato funkce má různé podoby volání, ale my budeme potřebovat podobu:

```
[bootstat,bootsam] = bootstrp(1,[],theta_IS,'Weights',weight_IS);
```

kde `bootstat` by za normálních okolností obsahovalo boostrapované statistiky či funkce původního vzorku, `bootsam` obsahuje boostrapované indexy ze vzorku proměnných (v našem případě vektoru) `theta_IS` (tyto indexy pak použijeme pro výběr z vektoru `theta_IS`), `[]` je prázdný argument, kde by za jiných okolností byla funkce nebo vektro funkcí aplikovaný na náš bootstrapovaný vzorek (např boostrapované střední hodnoty, směrodatné odchylky apod.), `theta_IS` je vektor parametrů z kandidátské hustoty a `weight_IS` je vektor odpovídajících vah z importance samplingu. `'Weights'` je dodatečná volba funkce `bootstrp`, umožňující zavést požadované váhy. Jednička ve funkci pak říká, že chceme jeden jediný bootstrapový vzorek (o stejné vleikosti jako vektor `theta_IS`). Kdybychom dali např. hodnotu 2, potom by se vytvořily dva vzorky a `bootsam` by byla matice indexů (o dvou sloupcích).

3. Lion Forest je velmi úspěšný profesionální hráč golfu. Ve věku 45 let však jeho hra přestala být tou, kterou bývala dříve. Svou profesionální kariéru začal, když mu bylo 20 a o 45. narozeninách se začal zajímat o analýzu historického průběhu svých vsledků s tím jak postupně stárnul. Soubor `golf.m` obsahuje údaje o jeho konečném skóre (skutečné skóre mínus par) ze 150 turnajů spolu s údaji o věku (v jednotkách odpovídajících deseti letům). V souboru jsou obsaženy výsledky z šesti hlavních turnajů (pro každý rok) v průběhu posledních 25 let. Označíme-li si výsledek turnaje jako $SCORE$ a jeho věk jako AGE, odhadněte následující model a pokuste se i o modelovou predikci jeho výsledků v rámci vzorku:

$$SCORE = \beta_1 + \beta_2 AGE + \beta_3 AGE^2 + \beta_4 AGE^3 + \epsilon.$$

   (a) Odhadněte model za předpokladu nezávislé normální-gama apriorní hustoty.

   (b) Ověřte (pomocí porovnání modelů), který z modelů je vhodnější, zda-li kvadratický nebo kubický.

   (c) Použijte modelové predikce (v rámci vzorku, tzn. pro věk od 20 do 45 let) pro zodpovězení následujících otázek:

      i. V jakém věku byl Lion na vrcholu své kariéry?

      ii. V jakém období jeho věku docházelo ke zlepšování jeho hry, a to rostoucím tempem?

      iii. V jakém období docházelo ke zlepšování Lionovy hry, a to klesjícím tempem?

      iv. Ve kterém věku začal hrát Lion hůře než na začátku své kariéry (ve věku 20 let)?

      v. Od kterého věku Lion už nebyl (či nebude) schopen hrát pod par? (v průměru)

   (d) Když bude Lionovi 70, bude (podle našho modelu) schopen zahrát turnaj na 100 úderů? Předpokládáme, že par je 72.

4. *Odhad hysterezní Phillipsovy křivky* Jednoduchý model Phillipsovy křivky je možno nalézt v článku Roberta Gordona z roku 1989. V tomto modelu je jednoduchá verze hypotézy přirozené míry nezaměstnanosti, která propojuje inflaci π_t a míru nezaměstnanosti U_t, zapsána následovně:

$$\pi_t = \alpha \pi_{t-1} + \beta(U_t - U_t^*). \tag{1}$$

Parametr α vyjadřuje setrvačnost v očekávání inflačního vývoje a jedná se tak o jistý druh adaptivních očekávání. Umožníme-li existenci jevu hystereze, můžeme definovat pravidlo, podle kterého se vyvíjí rovnovážná míra nezaměstnanosti U^* (reprezentována úrovní NAIRU):

$$U_t^* = \eta U_{t-1} + Z_t \tag{2}$$

Hystereze tedy nastává v případě, kdy U_t^* závisí na zpožděné hodnotě míry nezaměstnanosti U_{t-1} a na mikroekonomických determinantech reprezentovanými proměnnou Z_t. Tyto mikroekonomické determinanty můžeme ztotožnit s těmi, které uvádí Friedmann v rámci své hypotézy o přirozené míře nezaměstnanosti. Spojením obou vztahů získáme:

$$\pi_t = \alpha\pi_{t-1} + \beta(U_t - \eta U_{t-1} - Z_t). \tag{3}$$

Následná transformace vede k rovnici:

$$\pi_t = \alpha\pi_{t-1} + \beta(1-\eta)U_t + \beta\eta(U_t - U_{t-1}) - \beta Z_t. \tag{4}$$

Tuto rovnici využijte k empirickému testování hypotézy hystereze. Všimněte si teoretických aspektů a implikací, které nám předpoklad hysterezního charakteru nezaměstnanosti přináší. Je zřejmé, že pro $\eta = 1$ nastává případ "plné hystereze". V tomto případě již nebude existovat jedinečné U_t^* a rovnovážná úroveň nezaměstnanosti bude zcela variabilní veličinou nemající svou ustálenou (steady state) hodnotu.

„Plná hystereze“ má zásadní dopad na vztah inflace a nezaměstnanosti. Inflace v tomto případě nebude záviset na aktuální úrovni nezaměstnanosti, ale jen na změně v nezaměstnanosti. To je samozřejmě v protikladu s hypotézou o přirozené míře nezaměstnanosti, které by odpovídal případ $\eta = 0$. Rovnovážná úroveň nezaměstnanosti by v tomto případě plně reflektovala mikroekonomické determinanty reprezentované proměnnou Z_t. Jakýmsi kompromisem pak jsou hodnoty $\eta \in (0; 1)$, které připouštějí existenci inflačních tlaků jak ze strany aktuální úrovně nezaměstnanosti, tak i ze strany změn v míře nezaměstnanosti. Tento případ umožňuje existenci ustálené úrovně nezaměstnanosti, tedy úrovně, která nebude akcelerovat míru inflace a bude dlouhodobě udržitelná. Aktuální rovnovážná úroveň nezaměstnanoti bude mít tendenci k této ustálené úrovni konvergovat. Čím více se bude hodnota parametru η blížit jedné, tím pomalejší bude přizpusobování NAIRU svému ustálenému stavu a tím menší budou "inflační náklady"(v důsledku akceleračních tlaků na růst cenové hladiny) expanzivní, poptávkově orientované hospodářské politiky cílené na snížení míry nezaměstnanosti.

(a) *Úkol:* K dispozici máte čtvrtletní data o nezaměstnanosti (sezónně očištěné) a meziroční inflaci (pro ČR se jedná o čistou inflaci, pro Nový Zéland o inflaci spotřebitelskou, počítanou na základě CPI). Pro Českou republiku jsou data od 2. čtvrtletí 1995 do 3. čtvrtletí 2007, pro Nový zéland od 2. čtvrtletí 1991 do 3. čtvrtletí 2007. Níže uvedené úkoly řešte pro ekonomiku České republiky nebo ekonomiku Nového Zélandu (případně pro obě) a kriticky diskutujte dosažené výsledky. Alternativně se pokuste získat novější data pro zkoumané ekonomiky nebo data pro ekonomiku vlastní (viz např. databáze OECD, kdy pro plnohodnotný přístup je potřeba jít přes proxy-server naší knihovny, či databáze Eurostatu).

(b) Výchozí soubor je `gordon_bayes_zadani.m`, data jsou obsahem souborů `gordon_data_CZ.mat` a `gordon_data_NZ.mat`.

(c) Odhadovaný ekonometrický model má v souladu s rovnicí (4) následující podobu (model je chápán jako normální lineární regresní model s nezávislou normální-gama apriorní hustotou):

$$\pi_t = \lambda_1 + \lambda_2\pi_{t-1} + \lambda_3 U_t + \lambda_4(U_t - U_{t-1}) + \epsilon_t. \tag{5}$$

(d) Předpokládáme, že strukturální charakteristiky jsou v čase neměnné a náhodná složka splňuje obvyklé požadavky. Odhadněte parametry tohoto modelu (využijte Gibbsův vzorkovač) a na jejich základě pak zpětně získejte původní strukturální parametry. Apriorní hustotu obohaťte o informaci, týkající se přípustných hodnota parametrů η, tedy o informaci, že $\eta \in (0, 1)$. Nezapomeňte ověřit konvergenci.

   i. Využijte výsledky ze čtvrté kapitoly pro vytvoření odpovídajícího Gibbsova vzorkovače. Pracujeme totiž stále s lineárním regresním modelem s nezávislou normální-gama apriorní hustotou.

   ii. Zakomponování apriorního omezení na parametr η je velmi snadné. V rámci generování vzorků redukované formy modelu si vždy spočítáme původní strukturální parametry a vzorky v daném kroku replikací generujeme tak dlouho, dokud nebude splněna podmínka, že η je v intervalu nula až jedna. Dobré je rovněž uchovat si informaci o tom, kolik vzorků se vygenerovalo celkem, protože tím získáme důležitou hodnotu pro výpočet odpovídající integrační konstanty pro omezené vícerozměrné normální rozdělení (využitelné to je v rámci konstrukce Savage-Dickey ho poměru hustot, protože v něm potřebujeme znát plné hustoty a nikoliv jen jejich jádra).

(e) Vypočítejte jednotlivé pravděpodobnosti modelů, které odpovídají platnosti hypotézy o přirozené míře nezaměstnanosti, hypotézy hystereze a teorii NAIRU. Model odpovídající teorii NAIRU tak bude neomezený model, zbylé dva modely budou odpovídat vnořeným modelům $\eta = 0$ resp $\eta = 1$. Vypočítejte tedy příslušné Bayesovy faktory (na základě Savage-Dickey poměru hustot.

(f) Díky znalosti „law of motion“ pro vývoj NAIRU nasimulujte jeho trajektorii a sestrojte i příslušné 95% intervaly spolehlivosti.

i. NAIRU v kontextu oné rovnice není nic jiného než funkce pozorovaných (minulých) hodnot nezaměstnanosti a parametrů η a Z_t (chápané jeko část úrovňové konstanty).

ii. Střední hodnotu a rozptyl jakékoliv funkce paramtrů jsme pomocí Monte Carlo integrace schopni velmi snadno spočítat. V tomto případě máme totiž vygenerované platné výběry s psoteriorní hustoty) a tudíž jsme schopni snadno generovat i rozdělení NAIRU.

iii. Vcelku efektivní může být zachování vygenerovaných vzorků pomocí Matlabovské funkce `save`. (uložení do `.mat` souboru). Příslušný datový soubor si pak můžeme načíst v rámci nového skriptu věnovanému simulaci NAIRU.

5. 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 nezávislou normální gama 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.

6. 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 (v kontextu NLRM s nezávislou normální gama apriorní hustotou)

   $$time = \beta_0 + \beta_1 depart + \beta_2 reds + \beta_3 trains + \epsilon.$$

   (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.