Cvičení P01, od 1.11.2021:

Při měření rychlosti $n = 50$ aut na Rohanském nábřeží naproti budově MFF jsme získali průměr $\overline{y} = 59.6$ km/h a výběrovou směrodatnou odchylku $s = 11.1$ km/h (měření probíhala v době před instalací dvou nových světelných křižovatek na rozích kancelářské budovy *River Garden*).

Vyšetřete aposteriorní rozdělení střední hodnoty rychlosti aut na Rohanském nábřeží za předpokladu normálního rozdělení rychlostí $\mathcal{N}(\mu, \sigma^2)$. Jako apriorní rozdělení pro μ a $\tau = \sigma^{-2}$ uvažujte (semi-konjugované) rozdělení:

$$p(\mu,\, \tau) = p(\mu)\, p(\tau),$$

kde $p(\mu) \propto 1$ a $\tau \sim \mathsf{Ga}(g,\, h)$, $p(\tau) \propto \tau^{g-1} \exp(-\, h\, \tau)$. Za g, h uvažujte postupně následující volby:

(i) $g = 0.001,\ h = 0.001$;

(ii) $g = 1,\ h = 0.005$;

(iii) $g = 0,\ h = 0$ (nevlastní rozdělení s $p(\tau) \propto \tau^{-1}$).

1. Pro kontrolu spočítejte obyčejný (tj. nebayesovský, frekventistický) 95% konfidenční interval pro neznámou střední hodnotu μ a taktéž pro směrodatnou odchylku σ a inverzní rozptyl τ.

2. Odvoďte marginální aposteriorní rozdělení pro μ a nakreslete jeho hustotu (do jednoho obrázku) pro výše uvedené volby g a h.

   S jakou aposteriorní pravděpodobností je střední hodnota rychlosti vyšší než 55 km/h (opět pro tři různé volby g a h)?

3. Odvoďte marginální aposteriorní rozdělení pro τ a nakreslete jeho hustotu (do jednoho obrázku) pro výše uvedené volby g a h. Do druhého obrázku nakreslete marginální aposteriorní hustoty pro $\sigma = \sqrt{1/\tau}$.

   Uvědomte si, že pro účely kreslení obrázku není potřeba explicitně odvozovat vzorec pro aposteriorní hustotu σ, máte-li k dispozici počítačovou funkci pro výpočet funkčních hodnot aposteriorní hustoty τ.

4. Nakreslete (do tří různých obrázků) *image/contour* graf sdružené aposteriorní hustoty $(\mu,\, \tau)$ pro tři výše uvedené volby hyperparametrů g, h.

   Uvědomte si, že pro výpočet funkčních hodnot sdružené hustoty $p(\mu,\, \tau \,|\, \mathbf{y})$ a její kreslení lze využít vztahu $p(\mu,\, \tau \,|\, \mathbf{y}) = p(\mu \,|\, \tau,\, \mathbf{y})\, p(\tau \,|\, \mathbf{y})$.

5. Nakreslete (do tří různých obrázků) *image/contour* graf sdružené aposteriorní hustoty $(\mu,\, \sigma)$ pro tři výše uvedené volby hyperparametrů g, h.

   Uvědomte si, že pro výpočet funkčních hodnot sdružené hustoty $p(\mu,\, \sigma \,|\, \mathbf{y})$ a její kreslení lze využít vztahu $p(\mu,\, \sigma \,|\, \mathbf{y}) = p(\mu \,|\, \sigma,\, \mathbf{y})\, p(\sigma \,|\, \mathbf{y})$, kde navíc $p(\mu \,|\, \sigma,\, \mathbf{y}) = p\big(\mu \,\big|\, \sigma^{-2},\, \mathbf{y}\big)$.

6. Pro každou výše uvedenou volbu hyperparametrů g a h spočtěte 95% ET (*equal-tail*) věrohodnostní intervaly pro μ, τ i σ. Jak se liší hodnoty a interpretace těchto intervalů od frekventistických (nebayesovských) protějšků?

7. Pro každou výše uvedenou volbu hyperparametrů g a h spočtěte (numericky, pokud si myslíte, že nelze jinak) 95% HPD (*highest posterior density*) věrohodnostní intervaly pro μ, τ i σ. Liší se tyto intervaly od ET věrohodnostních intervalů?

8. Napište krátkou funkci, pomocí níž lze generovat pseudonáhodná čísla ze (sdruženého) aposteriorního rozdělení (μ, τ) (apriorní hyperparametry specifikujte jako argumenty této funkce).

   Opět si uvědomte význam vztahu $p(\mu, \tau \,|\, \mathbf{y}) = p(\mu \,|\, \tau, \mathbf{y})\, p(\tau \,|\, \mathbf{y})$.

   Na základě nasimulovaného výběru z aposteriorního rozdělení (délky alespoň 10 000) spočtěte znovu (nyní Monte Carlo odhady pro) 95% ET věrohodnostní intervaly parametrů μ, τ a σ (opět pro tři volby hyperparametrů g a h). Liší se tyto intervaly od těch spočtených v bodu 6?

9. Výše uvedenou funkci na generování z aposteriorního rozdělení $p(\mu, \tau \,|\, \mathbf{y})$ mírně rozšiřte tak, aby bylo možno generovat též z prediktivního rozdělení rychlosti Y_{n+1} dalšího projíždějícího auta:

$$\begin{aligned} p(y_{n+1} \,|\, \mathbf{y}) &= \int p(y_{n+1}, \mu, \tau \,|\, \mathbf{y})\, \mathrm{d}(\mu, \tau) \\ &= \int p(y_{n+1} \,|\, \mu, \tau, \mathbf{y})\, p(\mu, \tau \,|\, \mathbf{y})\, \mathrm{d}(\mu, \tau) = \int p(y_{n+1} \,|\, \mu, \tau)\, p(\mu, \tau \,|\, \mathbf{y})\, \mathrm{d}(\mu, \tau). \end{aligned} \quad (1)$$

   Následně spočtěte (na základě minimálně 10 000 nasimulovaných hodnot z prediktivního rozdělení $p(y_{n+1} \,|\, \mathbf{y})$) 95% ET věrohodnostní interval pro Y_{n+1} (opět při třech volbách hyperparametrů g a h). Jak lze interpretovat věrohodnostní intervaly pro Y_{n+1}?

   Uvědomte si, že ke generování z $p(y_{n+1} \,|\, \mathbf{y})$ můžete využít (pouze na první pohled složitější) generování ze sdruženého rozdělení

$$p(y_{n+1}, \mu, \tau \,|\, \mathbf{y}) = p(y_{n+1} \,|\, \mu, \tau, \mathbf{y})\, p(\mu, \tau \,|\, \mathbf{y}) = p(y_{n+1} \,|\, \mu, \tau)\, p(\mu, \tau \,|\, \mathbf{y}).$$

10. Pomocí simulace dále aproximujte hodnoty prediktivní hustoty $p(y_{n+1} \,|\, \mathbf{y})$ (opět pro tři různé volby hyperparametrů g a h) a nakreslete je do jednoho obrázku.

    Uvědomte si, že k MC odhadu hodnot prediktivní hustoty, lze využít vztahu (1).

11. Po dalším mírném rozšíření výše uvedené funkce na generování z aposteriorního rozdělení $p(\mu, \tau \,|\, \mathbf{y})$ spočtěte Monte Carlo odhad pravděpodobnosti, že další projíždějící auto překročí nejvyšší povolenou rychlost o více než 30 km/h (tj. řidič spáchá přestupek, po němž následuje odebrání řidičského průkazu)?

    Uvědomte si obdobu vztahu (1):

$$\begin{aligned} \mathsf{P}(Y_{n+1} > 80 \,|\, \mathbf{y}) &= \int \mathsf{P}(Y_{n+1} > 80 \,|\, \mu, \tau, \mathbf{y})\, p(\mu, \tau \,|\, \mathbf{y})\, \mathrm{d}(\mu, \tau) \\ &= \int \mathsf{P}(Y_{n+1} > 80 \,|\, \mu, \tau)\, p(\mu, \tau \,|\, \mathbf{y})\, \mathrm{d}(\mu, \tau). \end{aligned}$$

Deadline pro odevzdání vypracovaného úkolu (e-mailem na `komarek[AT]karlin.mff...`) je pondělí 15.11. v 15:11 CET.