Mattias Villani and Måns Magnusson
Division of Statistics and Machine Learning
Dept. of Computer and Information Science
Linköping University
mattias.villani@liu.se

2017-04-20

Bayesian Learning, 6 hp

Computer lab 3

- You are strongly recommended to use R for solving the labs since the computer exam will be in R.
- You are supposed to work and submit your labs in pairs, but do make sure that both of you are contributing.
- It is OK to discuss the lab with other student pairs in general terms, but `it is not allowed to share exact solutions`.
- Submit your *solutions* via LISAM no later than **May** 15 at **midnight**.

1. *Normal model, mixture of normal model with semi-conjugate prior.*
   The data `rainfall.dat` consist of daily records, from the beginning of 1948 to the end of 1983, of precipitation (rain or snow in units of $\frac{1}{100}$ inch, and records of zero precipitation are excluded) at Snoqualmie Falls, Washington. Analyze the data using the following two models.

   (a) *Normal model.*
   Assume the daily precipitation $\{y_1, ..., y_n\}$ are independent normally distributed, $y_1, ..., y_n|\mu, \sigma^2 \sim \mathcal{N}(\theta, \sigma^2)$ where both μ and σ^2 are unknown. Let $\mu \sim \mathcal{N}(\mu_0, \tau_0^2)$ independently of $\sigma^2 \sim Inv\text{-}\chi^2(\nu_0, \sigma_0^2)$.

      i. Implement (code!) a Gibbs sampler that simulates from the joint posterior $p(\mu, \sigma^2|y_1, ..., y_n)$. The full conditional posteriors are given on the slides from Lecture 7.

      ii. Analyze the daily precipitation using your Gibbs sampler in (a)-i. Investigate the convergence of the Gibbs sampler by suitable graphical methods.

   (b) *Mixture normal model.*
   Let us now instead assume that the daily precipitation $\{y_1, ..., y_n\}$ follow an iid two-component **mixture of normals** model:

   $$p(y_i|\mu, \sigma^2, \pi) = \pi\mathcal{N}(y_i|\mu_1, \sigma_1^2) + (1-\pi)\mathcal{N}(y_i|\mu_2, \sigma_2^2),$$

   where

   $$\mu = (\mu_1, \mu_2) \quad \text{and} \quad \sigma^2 = (\sigma_1^2, \sigma_2^2).$$

      i. Use the Gibbs sampling data augmentation algorithm in my code (or code your own!) in `NormalMixtureGibbs.R` (available under Lecture 7 on the course page) to analyze the daily precipitation data. Set the prior hyperparameters suitably.

ii. Investigate the convergence of the Gibbs sampler by suitable graphical methods.

(c) *Graphical comparison.*
Plot the following densities in one figure: 1) a histogram or kernel density estimate of the data. 2) Normal density $\mathcal{N}(\mu, \sigma^2)$ in (a); 3) Mixture of normals density $p(y_i|\mu, \sigma^2, \pi)$ in (b). Use the posterior mean value for all the parameters.

2. *Probit regression*

(a) Implement (code!) a data augmentation Gibbs sampler for the probit regression model

$$\Pr(y = 1|\mathbf{x}) = \Phi(\mathbf{x}^T\beta).$$

[Hint: `rtnorm` function in `the msm`-package + Lecture 8.]

(b) Compute the posterior of β in the probit regression for the `WomenWork` dataset from Lab 2 using the prior $\beta \sim \mathcal{N}(0, \tau^2 I)$, with $\tau = 10$.

(c) Do a normal approximation $\beta|\mathbf{y}, \mathbf{X} \sim N\left(\tilde{\beta}, J_{\mathbf{y}}^{-1}(\tilde{\beta})\right)$ of the posterior for β in the probit regression. Compare with the results from 2(b). Is the normal approximation accurate? [Hint: you can use your code from Lab 2 for the logistic regression. Just change the likelihood to a probit likelihood.]

HAVE FUN!