LINKÖPING UNIVERSITY
Department of Computer and Information Science
Division of Statistics
Mattias Villani

2014-04-25
Advanced Bayesian Learning
PhD course

Approximate Methods - Computer Lab

Deadline:	May 6 at midnight
Teacher:	Mattias Villani
Grades:	Pass/Fail
Submission:	By email to Mattias Villani, use message header: 'ABL - ApproxBayes'

The lab is to be reported in a concise report.
You have to write your own code to solve the problems, in whatever language you like.
You are not allowed to use existing toolboxes
Existing basic functions for matrix algebra, random number generators, density functions etc are allowed.
Attach your code to the email as separate executable files.

1. **Variational Bayes**. Consider first the usual the iid Poisson model for count data with an expontial prior:

$$X_1, ..., X_n|\lambda \overset{iid}{\sim} Poisson(\lambda)$$
$$\lambda|\beta \sim Exponential(\beta)$$

The posterior for λ is easily shown to be a Gamma distribution (see my slides from the Bayesian Learning course at http://www.ida.liu.se/~732A46/2013/BayesLearnL2.pdf and note that $Exponential(\beta) = Gamma(1, \beta)$, so the prior used here is a special case of the Gamma prior used in the previous course). Let's now complicate things a bit by assuming that you are uncertain about β, and you want to assign an exponential prior also to β. This gives us the following hierarchial model

$$X_1, ..., X_n|\lambda \overset{iid}{\sim} Poisson(\lambda)$$
$$\lambda|\beta \sim Exponential(\beta)$$
$$\beta \sim Exponential(\theta)$$

We will assume that $\theta = 5$.

(a) Let the observed training data be $\mathbf{x} = (x_1, ..., x_n)$. Let $Gamma(a, b)$ denote a gamma distribution parametrized so that the mean is a/b. [Note that there are two parametrizations of the gamma (see wikipedia) and you need to find out which one is used in your programming language.] Set up a Gibbs sampler to sample from the joint posterior $\pi(\lambda, \beta|\mathbf{x})$ by iterating the following two updating steps:

i. Simulate a draw from $\lambda|\beta, \mathbf{x}$
ii. Simulate a draw from $\beta|\lambda, \mathbf{x}$

(b) Use your Gibbs sampler to approximate joint posterior $p(\lambda, \beta|\mathbf{x})$ in the dataset $\mathbf{x} =$ (1,0,3,4,2,6,2,3,1,1).

(c) Derive and implement a mean field VB approximation the joint posterior:

$$q(\lambda, \beta) = q(\lambda) \cdot q(\beta)$$

This takes some work (you will not solve this in 5 minutes ...), but it is only straightforward algebra. The derivation of the lower bound $\ln \underline{p}(\mathbf{x}; q)$ is quite messy (but also straightforward). If you are unable to get $\ln \underline{p}(\mathbf{x}; q)$, you can proceed with the implementation and use a fixed number of iterations rather than monitoring changes in $\ln \underline{p}(\mathbf{x}; q)$. [Hints: both $q(\lambda)$ and $q(\beta)$ are Gamma densities].

(d) Use the VB implementation to approximate joint posterior $p(\lambda, \beta|\mathbf{x})$ in the dataset $\mathbf{x} =$ (1,0,3,4,2,6,2,3,1,1).

(e) Compare the quality of the results from the Gibbs sampler and the VB approx. Compare computing times.

2. **ABC**. Use ABC to approximate the posterior from Problem 1 above. Make your own creative choice of summary statistic and distance function. Explore the sensitivity with respect to ε, and the choice of summary statistics. Compare with the Gibbs sampling results in Problem 1. Implement:

(a) Likelihood-free rejection sampler 2

(b) Likelihood-free MCMC sampler. Let $q(\cdot|\theta^{(t-1)})$ be simple random walk Metropolis kernel $\theta'|\theta^{(t-1)} \sim N(\theta^{(t-1)}, c^2 I)$, where $\theta = (\lambda, \beta)$ and $c > 0$ is a tuning constant.

(c) Correct the draws from a) and b) using post-processing of the ABC output with the regression approach.

Good luck! Remember that sometimes an approximate answer is the correct answer!