Mattias Villani
Department of Statistics
Stockholm University



## Computer Lab 3

---

The labs should be done in R, since only R is available at the computer exam.
Your lab solutions will be available for you at the exam.
You work and submit your labs in pairs, but both of you should contribute equally and understand all parts of your solutions.
`It is not allowed to share exact solutions` with other student pairs.
Submit your solutions via Athena.

---

1. *Normal model, mixture of normal model with semi-conjugate prior.*
   The data `rainfall` 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, $y_1, ..., y_n|\mu, \sigma^2 \sim \mathcal{N}(\mu, \sigma^2)$ where both $\mu$ and $\sigma^2$ 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. Evaluate the convergence of the Gibbs sampler by suitable graphical methods, for example by plotting the trajectories of the sampled Markov chains.

   (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

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

   Use the Gibbs sampling data augmentation algorithm in `NormalMixtureGibbs.R` (available under Lecture 7 on the course page) to analyze the daily precipitation data. Set the prior hyperparameters suitably. Evaluate the convergence of the sampler.

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

2. *Metropolis Random Walk for Poisson regression.*
Consider the following Poisson regression model

$$y_i|\beta \sim \text{Poisson}\left[\exp\left(\mathbf{x}_i^T\beta\right)\right],\ i = 1, ..., n,$$

where $y_i$ is the count for the $i$th observation in the sample and $x_i$ is the $p$-dimensional vector with covariate observations for the $i$th observation. Use the data set `ebaycoins` from the library `SUdatasets`. This dataset contains observations from 1000 eBay auctions of coins. See `?ebaycoins` for variable definitions. Use `NBidders`, the number of bids in each auction, as the response variable. Use all remaining variables in the dataset as covariates, except for `FinalPrice`. Add a columns of ones to model the intercept. Make a log transform of the variable `BookVal` before using it as a covariate.

(a) Obtain the maximum likelihood estimator of $\beta$ in the Poisson regression model for the `ebaycoins` data [Hint: `glm.R`, don't forget that `glm()` adds its own intercept so don't input the covariate for the intercept]. Which covariates are significant?

(b) Let's now do a Bayesian analysis of the Poisson regression. Let the prior be $\beta \sim \mathcal{N}\left[\mathbf{0}, 100 \cdot (\mathbf{X}^T\mathbf{X})^{-1}\right]$ where $\mathbf{X}$ is the $n \times p$ covariate matrix. This is a commonly used prior which is called Zellner's g-prior. Assume first that the posterior density is approximately multivariate normal:

$$\beta|y \sim \mathcal{N}\left(\tilde{\beta}, J_{\mathbf{y}}^{-1}(\tilde{\beta})\right),$$

where $\tilde{\beta}$ is the posterior mode and $J_{\mathbf{y}}(\tilde{\beta})$ is the negative Hessian at the posterior mode. $\tilde{\beta}$ and $J_{\mathbf{y}}(\tilde{\beta})$ can be obtained by numerical optimization (`optim.R`) exactly like you already did for the logistic regression in Lab 2.

(c) Now, let's simulate from the actual posterior of $\beta$ using the Metropolis algorithm and compare with the approximate results in b). Code up a general function that uses the Metropolis algorithm to generate random draws from an *arbitrary* posterior density. In order to show that it is a general function for any model, I will denote the vector of model parameters by $\theta$. Let the proposal density be the multivariate normal density mentioned in Lecture 8 (random walk Metropolis):

$$\theta_p|\theta^{(i-1)} \sim N\left(\theta^{(i-1)}, c \cdot \Sigma\right),$$

where $\Sigma = J_{\mathbf{y}}^{-1}(\tilde{\beta})$ obtained in b). The value $c$ is a tuning parameter and should be an input to your Metropolis function. The user of your Metropolis function should be able to supply her own posterior density function, not necessarily for the Poisson regression, and still be able to use your Metropolis function. This is not so straightforward, unless you have come across *function objects* in `R` and the triple dot (...) wildcard argument. I have posted a note (HowToCodeRWM.pdf) on the course web page that describes how to do this in `R`. Now, use your new Metropolis function to sample from the posterior of $\beta$ in the Poisson regression for the eBay dataset. Assess MCMC convergence by graphical methods.

3. *Time series models in Stan*

(a) Write a function in `R` that simulates data from the AR(1)-process

$$x_t = \mu + \phi\left(x_{t-1} - \mu\right) + \varepsilon_t,\ \ \varepsilon_t \overset{iid}{\sim} N\left(0, \sigma^2\right),$$

for given values of $\mu$, $\phi$ and $\sigma^2$. Start the process at $x_1 = \mu$ and then simulate values for $x_t$ for $t = 2, 3 \ldots, T$ and return the vector $x_{1:T}$ containing all time points. Use $\mu = 10, \sigma^2 = 2$ and $T = 200$ and look at some different realizations (simulations) of $x_{1:T}$ for values of $\phi$ between $-1$ and 1 (this is the interval of $\phi$ where the AR(1)-process is stable). Include a plot of at least one realization in the report. What effect does the value of $\phi$ have on $x_{1:T}$?

(b) Use your function from a) to simulate two AR(1)-processes, $x_{1:T}$ with $\phi = 0.3$ and $y_{1:T}$ with $\phi = 0.95$. Now, treat the values of $\mu$, $\phi$ and $\sigma^2$ as unknown and estimate them using MCMC. Implement `Stan`-code that samples from the posterior of the three parameters, using suitable non-informative priors of your choice. [Hint: Look at the time-series models examples in the `Stan` reference manual, and note the different parameterization used here.]

  i. Report the posterior mean, 95% credible intervals and the number of effective posterior samples for the three inferred parameters for each of the simulated AR(1)-process. Are you able to estimate the true values?

  ii. For each of the two data sets, evaluate the convergence of the samplers and plot the joint posterior of $\mu$ and $\phi$. Comments?

(c) The data `campy` contain the number of cases of campylobacter infections in the north of the province Quebec (Canada) in four week intervals from January 1990 to the end of October 2000. It has 13 observations per year and 140 observations in total. Assume that the number of infections $c_t$ at each time point follows an independent Poisson distribution when conditioned on a latent AR(1)-process $x_t$, that is

$$c_t | x_t \sim \text{Poisson}\left(\exp\left(x_t\right)\right),$$

where $x_t$ is an AR(1)-process as in a). Implement and estimate the model in `Stan`, using suitable priors of your choice. Produce a plot that contains both the data and the posterior mean and 95% credible intervals for the latent intensity $\theta_t = \exp\left(x_t\right)$ over time. [Hint: Should $x_t$ be seen as data or parameters?]

(d) Now, assume that we have a prior belief that the true underlying intensity $\theta_t$ varies more smoothly than the data suggests. Change the prior for $\sigma^2$ so that it becomes informative about that the AR(1)-process increments $\varepsilon_t$ should be small. Re-estimate the model using `Stan` with the new prior and produce the same plot as in c). Has the posterior for $\theta_t$ changed?

HAVE FUN!