Project Submission Instructions

Please complete and save your project in pdf file. Before submission, please use the following naming convention:

`firstname-lastname`-P2.pdf

Your pdf file should be submitted to the e-mail address below:

`BIOS276.31r8w8ndztx2a19n@u.box.com`

Please do not use my personal e-mail address. After submission you should receive confirmation of successful uploading. I will let you know if I am missing your project by the day following the due date.

Please describe all your work in clear terms, before implementing `R` code. Each question should include a description of your approach with clear indication of where I can find the associated source code. Your code should be attached to your assignment or uploaded online to a repository I can freely access.

1. MCMC Sampling from the Banana Distribution

The two dimensional Banana distribution (Haario, 1999) is defined as:

$$\pi(x_1, x_2) \propto \exp\left\{-\frac{x_1^2}{2}\right\} \exp\left[-\frac{\left\{x_2 - 2\left(x_1^2 - 5\right)\right\}^2}{2}\right]; \qquad x_1, x_2 \in \mathbb{R}.$$

(a) Construct a Markov Chain to sample from $\pi(x_1, x_2)$ using the Metropolis Hastings algorithm. Describe your proposal strategy and compare convergence and mixing for different proposal jump sizes.

(b) Construct a Markov Chain to sample from $\pi(x_1, x_2)$ updating iteratively $\pi(x_1|x_2)$ and $\pi(x_2|x_1)$ in a component-wise fashion, using the Metropolis Hastings algorithm. Compare convergence and mixing for different proposal jump sizes.

(c) Repeat (b) using component-wise Gibbs sampling.

(d) Use the algorithms in (a, b, c) to estimate the following quantities: $E(x_1^2), E(x_2)$ and $Pr(x_1 + x_2 > 0)$. Discuss and quantify how different sampling strategies affect Monte Carlo variance.

2. Bayesian Adaptive Lasso

We consider the implementation of a posterior simulation algorithm for Bayesian adaptive LASSO. More precisely consider the following model:

$$Y \mid \beta, \sigma^2 \sim N_n(X\boldsymbol{\beta}, \sigma^2 I_n),$$

with $Y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n\times p}$, and $\boldsymbol{\beta} \in \mathbb{R}^p$. The model is completed with priors:

$$\boldsymbol{\beta} \mid \tau_1^2, \ldots, \tau_p^2 \sim N_p\left\{0, \text{diag}(\tau_1^2, \ldots, \tau_p^2)\right\},$$

$$p(\tau_1^2, ..., \tau_p^2 \mid \lambda^2) = \prod_{j=1}^{p} \frac{\lambda^2}{2} \exp\left\{-\frac{\lambda^2}{2}\tau_j^2\right\};$$

$$p(\sigma^2) \propto 1/\sigma^2$$

a. Consider $p = 1$. Simulate 5,000 Monte Carlo samples from the conditional prior $\boldsymbol{\beta} \mid \tau^2 = 1$ and obtain a plot of the density using the `R` function `density`.

b. Consider $p = 1$. Simulate 5,000 Monte Carlo samples from the marginal prior $\boldsymbol{\beta}$, considering $\lambda^2 = 2$, so that $E(\tau^2 \mid \lambda) = 1$. Obtain a plot of the density as in **a.**

c. Consider $p = 1$. Add a hyper prior on $\lambda^2 \sim Gamma(a, rate = b)$. Assess how the marginal prior of $\boldsymbol{\beta}$ changes for

$a = 1$ and values of $b \geq 1$.

d. Considering the hyper prior in **c**., describe a Markov Chain Monte Carlo algorithm to sample from the posterior distribution of $\boldsymbol{\beta}$ and σ^2.

e. Implement such algorithm in `R` and compare your results with estimates obtained using `glmnet()`. In particular, you should test your results on the `diabetes` data available from `lars`, (use the matrix of predictors x).

f. For the `diabetes` data, fix λ and produce a regularization path for adaptive Bayesian Lasso obtained on a grid of values for the tuning parameter λ. Describe your approach and compare your result with the path obtained using `glmnet()`.

g. Free λ and carry out a sensitivity analysis assessing the behavior of the posterior distribution of $\boldsymbol{\beta}$ and σ^2, as hyper parameters a and b are changed. Explain clearly the rationale you use to assess sensitivity and provide recommendations for the analysis of the diabetes data.