CPSC 535D: Assignement 1

February 16, 2007

1. Question 1: Bayesian Linear Model

Consider the following Bayesian linear model

$$Y = X\beta + \sigma\varepsilon$$

where $Y = (y_1, ..., y_n)^{\mathrm{T}}$, X is a $n \times p$ matrix, $\beta = \left(\beta_1, ..., \beta_p\right)^{\mathrm{T}}$ and $\varepsilon = (\varepsilon_1, ..., \varepsilon_p)^{\mathrm{T}}$ is $p \times 1$. We observe X and Y whereas $\varepsilon \sim \mathcal{N}(0, I_p)$.

• Compute the maximum likelihood estimate of $\left(\beta, \sigma^2\right)$.

We set the following prior distribution on $\left(\beta, \sigma^2\right)$

$$\begin{aligned} p\left(\beta, \sigma^2\right) &= p\left(\beta \middle| \sigma^2\right) p\left(\sigma^2\right) \\ &= \mathcal{N}\left(\beta; m, \sigma^2 V\right) \mathcal{IG}\left(\sigma^2; \frac{\mu}{2}, \frac{\nu}{2}\right) \end{aligned}$$

• Establish analytically the expression of the posterior distribution $p\left(\beta, \sigma^2 \middle| Y\right)$.

• Establish analytically the expression of the marginal posterior distribution $p\left(\beta \middle| Y\right)$ and $p\left(\sigma^2 \middle| Y\right)$.

• Establish analytically the expression of the marginal likelihood (also called evidence) $p(Y)$.

• Establish analytically the expression of the predictive distribution $p\left(y \middle| Y, x\right)$.

Consider a collection of potential models for the data

$$\mathcal{M}_i : Y = X_i \beta_i + \sigma_i \varepsilon$$

with prior $p\left(\beta_i, \sigma_i^2\right) = \mathcal{N}\left(\beta; m_i, \sigma_i^2 V_i\right) \mathcal{IG}\left(\sigma_i^2; \frac{\mu_i}{2}, \frac{\nu_i}{2}\right)$.

• Compute the Bayes factor $\frac{p(Y|\mathcal{M}_i)}{p(Y|\mathcal{M}_j)}$ and simplify the expression when we use a data-dependent prior known as the g-prior where

$$V_i = \delta^2 \left(X_i X_i^{\mathrm{T}}\right)^{-1}.$$

2. Question 2: Genetic Linkage Model

Assume that 197 animals are distributed into four categories as

$$\begin{aligned} Y &= (y_1, y_2, y_3, y_4) \\ &= (125, 18, 20, 34) \end{aligned}$$

with cell probabilities

$$\left(\frac{1}{2} + \frac{\theta}{4}, \frac{1}{4}(1-\theta), \frac{1}{4}(1-\theta), \frac{\theta}{4}\right).$$

• Derive and implement an Expectation-Maximization algorithm to estimate the Maximum Likelihood estimate of θ.

We assume that $\theta \sim \mathcal{U}[0,1]$, i.e. we set a uniform prior on $[0,1]$ for θ.

• Design and implement an accept/reject procedure to sample from the posterior $p(\theta|y)$.

• Design and impement an importance sampling method to sample from the posterior $p(\theta|y)$.

• Compare the efficiency of both algorithms.