STAT 460: Assignment 7

David Stenning

For this homework assignment, please complete the exercise below. It will require you to write a Gibbs sampler. You may use the sample code from lecture slides, previous homework solutions, or BDA3 as a guide, but you should not simply take code from the internet or rely on R packages (or similar for other languages) for implementing the algorithms.

Please also note that this homework assignment is worth 12 points total, not 10 points as for previous assignments, and will thus be weighted slightly more when computing your final grade.

Exercise 1: Consider the "Canadian house prices" example from the class lecture slides (Hierarchical Models: Part 1). Data for this example are contained in the file `housing_data.csv`. The first column is the log-price of a house and the second column indicates the province in which the house is located. For simplicity we consider only $J = 5$ provinces and obtain $n_j = 100$ prices per province.

Using the notation from class, y_{ij} is the log-price of house i in province j, and we consider the sample $y_{\cdot j} = y_{1j,\ldots,y_{n_j j}}$ as independent and identically distributed (iid) with density $\mathrm{N}(y_{ij}|\mu_j, \sigma^2)$, where μ_j and σ^2 are unknown model parameters. Note that the μ_j vary across provinces but σ^2 is common to all provinces.

We specify prior distributions $\mu_j \overset{\text{iid}}{\sim} \mathrm{N}(\mu_0, 25^2)$, for $j = 1, \ldots, 5$, and $\sigma^2 \sim \text{Inv-}\chi^2(1, 0.5^2)$. We also specify the hyperprior $\mu_0 \sim \mathrm{N}(0, 50^2)$.

(a) Write down the joint posterior distribution $p(\mu, \sigma^2, \mu_0|y)$, where $\mu = (\mu_1, \ldots, \mu_5)$ and $y = (y_{\cdot 1}, y_{\cdot 2}, \ldots, y_{\cdot 5})$. For this, you may use simplifying notation to represent known densities, such as $\mathrm{N}(y_{ij}|\mu_j, \sigma^2)$ to represent a Normal density with mean μ_j and variance σ^2 evaluated at y_{ij}. Be sure that your final formula includes the known constants for fixed values (i.e., it should not contain terms such as J and n_j as these are known, as are several of the parameters for the prior/hyperprior distributions). You do not need to simplify the equation; it should be the product of many densities, so you are being asked to be explicit regarding those densities. (**2 points**)

(b) Derive the following conditional distribution: $\mu_j|y, \mu_0, \sigma^2$. To do this, you may use previous results from single-parameter and multiparameter Normal models. (The final distributions are given in the class notes, but you should derive/justify your solutions on your own.) (**2 points**)

(c) Derive the following conditional distribution: $\sigma^2|y, \mu, \mu_0$. (**2 points**)

(d) Derive the following conditional distribution: $\mu_0|\mu, y, \sigma^2$. (**2 points**)

(e) Using the data in `housing_data.csv`, construct a Gibbs sampler using the results from parts (b) - (d) to obtain 10,000 after-burn-in draws from the joint posterior distribution, $p(\mu, \sigma^2, \mu_0|y)$. For this part, submit the trace plots for $\mu = (\mu_1, \ldots, \mu_5)$, σ^2, and μ_0. (**4 points**)