ST740 – Assignment 2 – Due 10/4

This homework is due in class on Wednesday, 10/4. Include code used to produce the results. You may work in groups, but write up separate solutions.

1. Let $Y_{ij}|\theta_i \stackrel{\text{ind}}{\sim} \text{Normal}(\theta_i, \sigma^2)$ and $\theta_i \stackrel{\text{iid}}{\sim} \text{Normal}(0, \sigma^2/m)$ for $i \in \{1, ..., p\}$ and $j \in \{1, ..., n\}$. The variance $\sigma^2 > 0$ and prior scaling factor m are fixed and known but the random effects $\theta = (\theta_1, ..., \theta_p)$ are unknown. Let $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_p)$ be the posterior mean of θ, $\theta^* = (\theta_1^*, ..., \theta_p^*)$ be the true value and the Euclidean discrepancy be $D(x, \theta^*) = \sum_{i=1}^p (x_i - \theta_i^*)^2$ for $x = (x_1, ..., x_p)$.
   (a) Compute $\text{E}[D(\hat{\theta}, \theta^*)]$ with respect to the distribution of the Y_{ij} given $\theta = \theta^*$.
   (b) Show that if p is fixed and $n \to \infty$, then $E[D(\hat{\theta}, \theta^*)] \to 0$.
   (c) Give conditions[1] on m, k and θ^* that ensure $E[D(\hat{\theta}, \theta^*)] \to 0$ as $p \to \infty$ with $n = p^k$.

2. Let Y_i be the observation in spatial region $i \in \{1, ..., n\}$. The observations are decomposed as $Y_i = \theta_i + \varepsilon_i$ for spatial process θ_i and error $\varepsilon_i \stackrel{\text{iid}}{\sim} \text{Normal}(0, \tau^2)$. The spatial terms are modelled using a conditionally autoregressive (CAR) model. The CAR model has PDF

$$\pi(\theta_1, ..., \theta_n|\sigma^2) \propto (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \sum_{j=1}^n A_{ij}(\theta_i - \theta_j)^2\right],$$

   where $A_{ij} = 1$ if regions i and j are adjacent and $A_{ij} = 0$ otherwise. To complete the Bayesian model, we select priors $\sigma^2, \tau^2 \sim \text{InvGamma}(a, b)$.
   (a) Using only the prior distribution, derive the full conditional distribution of θ_i given the other $n - 1$ parameters, θ_j for $j \neq i$. Describe why this is a reasonable prior.
   (b) Derive the univariate full conditional posterior distributions of θ_i, σ^2 and τ^2.
   (c) Download the `boats` data in `R`. The original image is 256×384 so $n = 98,304$. You may work with the reduced image obtained by retaining every fourth column and every fourth row, leaving $n = 6,144$.

```
library(imager)
plot(boats)
Y  <- grayscale(boats)[,384:1,1,1]
image(Y,col=gray.colors(100))
y  <- Y[1:nrow(y)%%4==2,] # Reduced image for speed
y  <- y[,1:ncol(y)%%4==2]
image(y,col=gray.colors(100))
```

   Apply Gibbs sampling to fit the CAR model to the boats data with $A_{ij} = 1$ for pixels that share an edge (so most observations will have four neighbors) and $a = b = 0.1$. Make trace plots of the variance components σ^2 and τ^2 and comment on convergence. Also, map the posterior mean and variance of $(\theta_1, ..., \theta_n)$ and comment on the fit to the data.

[1] the weaker the conditions, the better your score will be!