STP598sta: Spatiotemporal Analysis

Homework 2

Name: Your name; NetID: Your ID

Due 11:59pm Friday October 9, 2020

Question 1

For Monty Hall problem, assume that there are 4 doors. The contestant chooses a door, then Monty opens one of the three other doors and asks her whether she wants to switch or not. Find the probability of winning if she switches. Let's say she decides to stay with the door she chose first. Then Monty opens one of the other two doors that are still closed and asks her whether she wants to switch. What is the probability of winning if she switches this time?

Question 2

Consider a simple normal model for the height (y) of students where $y|\theta \sim N(\theta, 42)$. Assume a conjugate prior $\theta \sim N(0, 32)$. We measure the height of two students and observe $y_1 = 68$ and $y_2 = 71$.

(a) Find the posterior distribution of θ given y when the data come sequentially (i.e., we first observe y_1, we update our prior, then observe y_2 and update our prior again).

(b) Find the posterior distribution of θ given y when data come simultaneously (i.e., we update our prior after observing both y_1 and y_2).

(c) Plot the prior, the likelihood and all the different posterior distributions. Discuss your findings (just one paragraph).

Question 3

Consider `iris` data set in `R`. There are 3 species `setosa`, `versicolor` and `virginica` with 50 records for each. Let's consider a binary classification problem with the first 100 records (thus there are two levels `setosa` and `versicolor`).

```
iris2 = subset(iris,Species!=levels(iris$Species)[3])
head(iris2)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

```
iris2$Species = droplevels(iris2$Species)
levels(iris2$Species)
```

```
## [1] "setosa"     "versicolor"
```

(a) Fit a generalized linear model using `glm` function in R.

(b) Build a Bayesian logistic regression model (encode `setosa` as 1 and `versicolor` as 0):

$$P[y=1] = \frac{\exp(X\boldsymbol{\beta})}{1+\exp(X\beta)}$$
$$\boldsymbol{\beta} \sim N(\mathbf{0}, \mathbf{\Lambda}_0)$$

You can specify $\mathbf{\Lambda}_0$ as you like (diagonal matrix or scale matrix). But keep in mind that prior should be broad to include more possible models.

- Draw posterior samples of $\boldsymbol{\beta}|Y$ using either random walk Metropolis (RWM, as vector or component by component) or slice sampler (component by component). Samplers are provided on the course website.
- Diagnose the convergence of Markov chain using `geweke.diag` or `gelman.diag` (multiple chains) in `coda` package of `R`.
- Plot each of the five posterior densities of $\beta_i|Y$ $(i = 0, 1, \cdots, 5)$ using `density` function in one figure.
- Obtain posterior estimates (mean or median) of $\boldsymbol{\beta}$. Compare the results with GLM.

Question 4

Derive the forms of the full conditionals for $\boldsymbol{\beta}$, σ^2 and $\mathbf{W}$ in the exponential kriging model (τ^2 is assumed to be fixed)

$$\mathbf{Y}|\boldsymbol{\theta}, \mathbf{W} \sim N(X\boldsymbol{\beta} + \mathbf{W}, \tau^2 I)$$
$$\mathbf{W}|\sigma^2, \phi \sim N(\mathbf{0}, \sigma^2 H(\phi))$$
$$p(\boldsymbol{\beta}, \sigma^2, \phi) \propto \frac{p(\phi)}{(\sigma^2)^\alpha}$$

Question 5

The lithology data set (see https://www.counterpointstat.com/uploads/1/1/9/3/119383887/lithology.dat) consists of measurements taken at 118 sample sites in the Radioactive Waste Management Complex region of the Idaho National Engineering and Environmental Laboratory. At each site, bore holes were drilled and measurements taken to determine the elevation and thickness of the various underground layers of soil and basalt. Understanding the spatial distribution of variables like these is critical to predicting fate and transport of groundwater and the (possibly harmful) constituents carried therein; see Leecaster (2002) for full details. For this problem, consider only the variables `Northing`, `Easting`, `Surf Elevation`, `Thickness`, and `A-B Elevation`, and only those records for which full information is available (i.e., extract only those data rows without an "NA" for any variable).

(a) Produce image plots of the variables `Thickness`, `Surf Elevation`, and `A-B Elevation`. Add contour lines to each plot and comment on the descriptive topog- raphy of the region.

(b) Taking `log(Thickness)` as the response and `Surf Elevation` and `A-B Elevation` as covariates, fit a univariate Gaussian spatial model with a nugget effect, using the Mat'ern covariance functions. You may start with flat priors for the covariate slopes, Inverse Gamma(0.1, 0.1) priors for the spatial and nugget variances, and a Gamma(0.1,0.1) prior for the spatial range parameter. Modify the priors and check for their sensitivity to the analysis. (Hint: You will need `spBayes` for the Mat'ern model.)

(c) Perform Bayesian kriging on a suitable grid of values and create image plots of the posterior mean residual surfaces for the spatial effects.