P8160 - Hurricane Project Report

Wenbo Group#3

5/5/2022

1. Introduction

1.1. Background

Hurricanes are dangerous and can cause major damage from storm surge, wind damage, rip currents and flooding. They can happen along any U.S. coast or in any territory in the Atlantic or Pacific oceans. The amount of damage depends on the strength of a storm and what it hits. High winds are one of the primary causes of hurricane-inflicted loss of life and property damage. For better planning and prevention ahead to secure people from destructive hurricanes, it is extremely important and necessary to explore trajectories of hurricanes and predict each hurricane's wind speed.

1.2. Objectives

In this study, two data sets were explored.In the first part, we attempted to use the track data of 703 hurricanes in the North Atlantic area since 1950 to explore the seasonal differences and if there is any evidence showing that the hurricane wind speed has been increasing over years. First, we calculated the posterior distribution of four parameters $(B, \beta, \sigma^2, \Sigma^{-1})$ in proposed Bayesian model. Next, we designed an MCMC algorithm to generate the posterior distribution. Then, we used the MCMC chain we developed to estimate the parameters, and checked to see how well the model fits the data.

Furthermore, in order to forecast hurricane damage and deaths, another data set containing the damages and deaths caused by 46 hurricanes in the United States were used. We constructed a model and to determine which traits of hurricanes are more associated to damage and deaths.

2. Methods

2.1. Data Cleaning and Exploratory Analysis

In this study, there are two data sets. First one contains 703 hurricanes in the North Atlantic since 1950. It recorded the location (longitude and latitude) and maximum wind speed every 6 hours for every hurricanes. There are 8 variables and 22038 observations.

The second data set contains the damages and deaths caused by 46 hurricanes in the United States. There are 14 variables and 43 observations.

2.2. Posterior Distributions

The following Bayesian model was suggested.

$$Y_i(t+6) = \beta_{0,i} + \beta_{1,i}Y_i(t) + \beta_{2,i}\Delta_{i,1}(t) + \beta_{3,i}\Delta_{i,2}(t) + \beta_{4,i}\Delta_{i,3}(t) + \epsilon_i(t)$$

where $Y_i(t)$ the wind speed at time t (i.e. 6 hours earlier), $\Delta_{i,1}(t)$, $\Delta_{i,2}(t)$ and $\Delta_{i,3}(t)$ are the changes of latitude, longitude and wind speed between t and $t+6$, and $\epsilon_{i,t}$ follows a normal distributions with mean zero and variance σ^2, independent across t.

In the model, $\boldsymbol{\beta}_i = (\beta_{0,i}, \beta_{1,i}, ..., \beta_{7,i})$ are the random coefficients associated the ith hurricane, we assume that

$$\boldsymbol{\beta}_i \sim \mathcal{N}(\boldsymbol{\beta}, \boldsymbol{\Sigma}),$$

and we assume the following non-informative or weak prior distributions for σ^2, $\boldsymbol{\beta}$ and Σ.

$$P(\sigma^2) \propto \frac{1}{\sigma^2}; \quad P(\boldsymbol{\beta}) \propto 1; \quad P(\Sigma^{-1}) \propto |\Sigma|^{-(d+1)} \exp(-\frac{1}{2}\Sigma^{-1})$$

d is dimension of β.

Note from given Bayesian model:

$$\epsilon_i(t) = Y_i(t+6) - \Big(\beta_{0,i} + \beta_{1,i}Y_i(t) + \beta_{2,i}\Delta_{i,1}(t) + \beta_{3,i}\Delta_{i,2}(t) + \beta_{4,i}\Delta_{i,3}(t)\Big) \overset{i.i.d}{\sim} N(0, \sigma^2)$$

or

$$Y_i(t+6) \sim N(\boldsymbol{X}_i(t)\boldsymbol{\beta}_i^T, \sigma^2)$$

where $\boldsymbol{X}_i(t) = (1, Y_i(t), \Delta_{i,1}(t), \Delta_{i,2}(t), \Delta_{i,3}(t))$, and $\boldsymbol{\beta}_i = (\beta_{0,i}, \beta_{1,i}, \beta_{2,i}, \beta_{3,i}, \beta_{4,i})$. Therefore,

$$f_{Y_i(t+6)}(y_i(t+6) \mid \boldsymbol{X}_i(t), \boldsymbol{\beta}_i, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big\{ -\frac{1}{2\sigma^2}\Big(y_i(t+6) - \boldsymbol{X}_i(t)\boldsymbol{\beta}_i^T\Big)^2\Big\}$$

for hurricane i at time t. To show the likelihood function for hurricane i across all time points, t, we can write the multivariate normal distribution

$$(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \boldsymbol{\beta}_i, \sigma^2) \sim \mathcal{N}(\boldsymbol{X}_i\boldsymbol{\beta}_i^T, \sigma^2 I)$$

where Y_i is an m_i-dimensional vector and $\boldsymbol{X}_i$ is a $m_i \times d$ matrix. Finally, the joint likelihood function of all hurricanes can be expressed as

$$L_Y(\mathbf{B}, \sigma^2 I) = \prod_{i=1}^{n} \Big\{ \det(2\pi\sigma^2 I)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^T)^T(\sigma^2 I)^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^T)\Big)\Big\},$$

where I is an identity matrix with dimension consistent with Y_i. We can find the posterior distribution for Θ by

$$\pi(\mathbf{B}, \boldsymbol{\beta}, \sigma^2, \Sigma^{-1} \mid Y) \propto L_Y(\mathbf{B}, \sigma^2 I) \times \pi(\mathbf{B} \mid \boldsymbol{\beta}, \Sigma^{-1}) \times \pi(\boldsymbol{\beta}) \times \pi(\sigma^2) \times \pi(\Sigma^{-1}),$$

where $\pi(\mathbf{B} \mid \boldsymbol{\beta}, \Sigma)$ is the joint multivariate normal density of β,

$$\pi(\mathbf{B} \mid \boldsymbol{\beta}, \Sigma^{-1}) = \prod_{i=1}^{n} \Big\{ \det(2\pi\Sigma)^{-\frac{1}{2}} \exp(-\frac{1}{2}(\boldsymbol{\beta}_i - \boldsymbol{\beta})\Sigma^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\beta})^T) \Big\}.$$

So we have the following joint posterior distribution of Θ:

$$\pi(\mathbf{B}, \boldsymbol{\beta}, \sigma^2, \Sigma^{-1} \mid Y) \propto \prod_{i=1}^{n} \Big\{ (2\pi\sigma^2)^{-m_i/2} \exp\big\{ -\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^T)^T(\sigma^2 I)^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^T)\big\} \Big\}$$
$$\times \prod_{i=1}^{n} \Big\{ \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\big\{ -\frac{1}{2}(\boldsymbol{\beta}_i - \boldsymbol{\beta})\Sigma^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\beta})^T\big\} \Big\} \times \frac{1}{\sigma^2} \times |\Sigma|^{-(d+1)} \exp\big\{ -\frac{1}{2}\Sigma^{-1}\big\}.$$

We can use the joint posterior distribution to derive conditional posterior distributions for each of our parameters.

Let $\tau = \sigma^2$, then

$$(\tau|\beta, B, \Sigma^{-1}, Y) \propto \tau^{1+\frac{\sum_{i=1}^{n} m_i}{2}} exp(-\tau \times \frac{1}{2}\sum_{i=1}^{n}(Y_i - X_i\beta_i^T)^T I^{-1}(Y_i - X_i\beta_i^T)$$

Thus, σ^2 is from inverse-gamma distribution

$$\sigma^2 \sim \text{Inv-Gamma}(\frac{\sum_{i=1}^{n} m_i}{2}, \frac{1}{2}\sum_{i=1}^{n}(Y_i - X_i\beta_i^T)(Y_i - X_i\beta_i^T).$$

Parameter $\mathbf{B}$ has the following conditional posterior:

$$\pi(B|\beta, \sigma^2, \Sigma^{-1}, Y) \propto exp(-\frac{1}{2}\sum_{i=1}^{n}\left[(Y_i - X_i\beta_i^T)^T(\sigma^2 I)^{-1}(Y_i - X_i\beta_i^T) + (\beta_i - \beta)\Sigma^{-1}(\beta_1 - \beta)^T\right]) \tag{1}$$
$$\propto exp(-\frac{1}{2}\sum_{i=1}^{n}\left[\beta_i(X_i^T(\sigma^2 I)^{-1}X_i + \Sigma^{-1})\beta_i^T - 2\beta_i(X_i(\sigma^2 I)^{-1})Y_i + \Sigma^{-1}\beta^T\right]) \tag{2}$$

Let $V_i = X_i^T(\sigma^2 I)^{-1})X_i + \Sigma^{-1}$, and $U_i = X_i(\sigma^2 I)^{-1}Y_i + \Sigma^{-1}\beta^T$, then

$$\boldsymbol{\beta}_i \sim \mathcal{M}VN(V_i^{-1}U_i, V_i^{-1}).$$

Similarly, parameter $\boldsymbol{\beta}$ has a conditional posterior of:

$$\pi(\beta|B, \sigma^2, \Sigma^{-1}, Y) \propto exp(-\frac{1}{2}\sum_{i=1}^{n}(\beta_i - \beta)\Sigma^{-1}(\beta_1 - \beta)^T) \tag{3}$$
$$\propto exp(-\frac{1}{2}\sum_{i=1}^{n}\left[\beta\Sigma^{-1}\beta^T - 2\beta\Sigma^{-1}\beta_i^T\right]) \tag{4}$$

Let $V = n\Sigma^{-1}, U = \sum_{i=1}^{n}\Sigma^{-1}\beta_i^T$, then

$$(\boldsymbol{\beta}|B, \sigma^2, \Sigma^{-1}, Y) \sim \mathcal{M}VN(V^{-1}U, V^{-1}).$$

Finally, parameter Σ has the conditional posterior:

$$\pi(\Sigma^{-1}|\beta, B, \sigma^2, Y) \propto |\Sigma|^{-(d+1)}exp(-\frac{1}{2}tr(\Sigma^{-1})|\Sigma|^{-\frac{n}{2}}exp(-\frac{1}{2}\sum_{i=1}^{n}(\beta_i - \beta)\Sigma^{-1}(\beta_1 - \beta)^T) \tag{5}$$
$$\propto |\Sigma|^{-(d+1+\frac{n}{2})}exp(-\frac{1}{2}\left[tr(\Sigma^{-1}) + tr((\beta_i - \beta)\Sigma^{-1}(\beta_1 - \beta)^T)\right] \tag{6}$$
$$\propto |\Sigma|^{\frac{-(d+1+n+d+1)}{2}}exp(-\frac{1}{2}tr(\left[I + \sum_{i=1}^{n}(\beta_i - \beta)^T(\beta_i - \beta)\right]\Sigma^{-1})) \tag{7}$$

Thus,

$$\Sigma \sim \mathcal{W}_d^{-1}(\Psi, v),$$

where $v = d + 1 + n$, and $\Psi = I + \sum_{i=1}^{n}(\beta_i - \beta)^T(\beta_i - \beta)$.

2.3. Gibbs Sampling Algorithm

Now that we have conditional posterior distributions for each of our parameters, we can utilize the Gibbs Sampling MCMC algorithm to estimate model parameters. In Gibbs sampling, we use starting values $(\boldsymbol{\beta}_0, \Sigma_0, \sigma_0^2, \mathbf{B}_0)$ and for each $j = 1, 2, ..., n$:

1. Generate $\boldsymbol{\beta}_j$ from $\pi(\boldsymbol{\beta} \mid \Sigma = \Sigma_{j-1}, \sigma^2 = \sigma_{j-1}^2, \mathbf{B} = \mathbf{B}_{j-1})$;
2. Generate Σ_j from $\pi(\Sigma \mid \boldsymbol{\beta} = \boldsymbol{\beta}_j, \sigma^2 = \sigma_{j-1}^2, \mathbf{B} = \mathbf{B}_{j-1})$;
3. Generate σ^2 from $\pi(\sigma^2 \mid \boldsymbol{\beta} = \boldsymbol{\beta}_j, \Sigma = \Sigma_j, \mathbf{B} = \mathbf{B}_{j-1})$;
4. Generate $\mathbf{B}$ from $\pi(\mathbf{B} \mid \boldsymbol{\beta} = \boldsymbol{\beta}_j, \Sigma = \Sigma_j \sigma^2 = \sigma_j^2)$

to yield Θ_j. As j increases and the Markov Chain continues, estimates stabilize to yield our results.

3. Results

4. Discussion

4.1. Limitations

References