8160 P3

2023-04-16

Contents

1. Objectives/background

As hurricanes affect people, the ability to forecast hurricanes is essential for minimizing the risks in suffered areas. Hereby, a hierarchical Bayesian strategy for modeling North Atlantic hurricane counts since 1950 is illustrated. Model integration would be expected to achieve through a Markov chain Monte Carlo algorithm. Contingent on the availability of values for the covariates, the models can be used to explore the seasonality among track data and make predictive inferences on hurricanes.

The data given has 703 observations and following features are recorded for each hurricanes in the North Atlantic:

- ID: ID of the hurricanes
- Season: In which the hurricane occurred
- Month: In which the hurricane occurred
- Nature: Nature of the hurricane (ET-Extra tropical;DS-Disturbance;NR-Not rated;SS-Sub tropical; TS-Tropical storm)
- time: dates and time of the record
- Latitude and Longitude: The location of a hurricane check point
- Wind.kt: Maximum wind speed (in Knot) at each check point

```
#EDA if needed
dt= read.csv("/Users/yujia/Downloads/hurrican703.csv")
```

2. Method

2.1 Derivation

Let t be time (in hours) since a hurricane began, and For each hurricane i, we denote $Y_i(t)$ be the wind speed of the i^{th} hurricane at time t. 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 $t-6$. In the model, $\epsilon_{i,t}$ is the random error, following a normal distributions with mean zero and variance σ^2, independent across t. $\boldsymbol{\beta}_i = (\beta_{0,i}, \beta_{1,i}, ..., \beta_{5,i})$ are the random coefficients associated the i^{th} hurricane.Therefore,

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

While our interest is to derive the posterior distribution of the joint parameter $\Theta = (\mathbf{B}^\top, \boldsymbol{\mu}^\top, \sigma^2, \Sigma)$, let $\mathbf{B} = (\boldsymbol{\beta}_1^\top, ..., \boldsymbol{\beta}_n^\top)^\top$, it could be written as:

$$P(\theta \mid Y) \propto f(Y|\Theta) \cdot P(\Theta) = \prod_{i=1}^{n} f(Y|B, \mu, \sigma^2, \Sigma) \cdot \prod_{i=1}^{n} f(B \mid \mu, \Sigma) \cdot P\left(\sigma^2\right) \cdot P(\mu) \cdot P\left(\Sigma^{-1}\right)$$

To begin the derivation of above equation, prior distributions for model parameters are assumed as following:

$$P(\sigma^2) \propto \frac{1}{\sigma^2}$$

$$P(\boldsymbol{\mu}) \propto 1$$

$$P(\Sigma^{-1}) \propto |\Sigma|^{-(d+1)} \exp(-\frac{1}{2}\Sigma^{-1}),$$

where d is the dimension of $\boldsymbol{\beta}$ and $\beta_i \sim MVN(\mu, \Sigma)$ follows a multivariate normal distributions with mean μ and positive-definite covariance matrix Σ. Next in line we consider the response variable Y. The response variable $Y_i(t)$ at time t for the i^{th} hurricane follows a normal distribution with mean $\mathbf{x}_i(t)^\top\mathbf{B} + \boldsymbol{\mu}$ and variance σ^2 for $t = 1, ..., T$ where T is the total number of time points. Therefore, the likelihood function for the response variable Y can be expressed as:

$$\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \boldsymbol{\beta}_i, \sigma^2, \Sigma \sim MVN(\boldsymbol{X}_i\boldsymbol{\beta}_i^\top, \sigma^2 I)$$

$$\prod_{i=1}^{n} f\left(Y \mid B, \boldsymbol{\mu}, \sigma^2, \Sigma\right) = \prod_{i=1}^{n}\left((2\pi)^{-\frac{ni}{2}} \cdot \left|\sigma^2 I_{ni}\right|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}\left(y_i - x_i\beta_i^T\right)^\top \left(\sigma^2 I_{n_i}\right)^{-1}\left(y_i - x_i\beta_i^T\right)\right\}\right)$$

where $B = \left(\beta_1^\top, \beta_2^\top, \ldots, \beta_n^\top\right)^\top$ and n is the number of hurricanes. Then recall β_i follows MVN, the last part to identify for the posterior of Θ can be derived from the pdf of β_i:

$$\begin{aligned}\prod_{i=1}^{n} f(B|\mu, \Sigma) &= \prod_{i=1}^{n} MVN(B|\mu, \Sigma) \\ &= \prod_{i=1}^{n}(2\pi)^{-\frac{5}{2}}|\det(\Sigma)|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}\left(\beta_i - \mu\right)^\top \Sigma^{-1}\left(\beta_i - \mu\right)\right\}\end{aligned}$$

Substituting the likelihood and the prior distributions of the parameters into the joint distribution, we obtain the posterior distribution of the joint parameter Θ without simplification as:

$$
\begin{aligned}
&P(\theta \mid Y) \propto f(Y|\Theta) \cdot P(\Theta) \\
&= \prod_{i=1}^{n} f(Y|B, \mu, \sigma^2, \Sigma) \cdot \prod_{i=1}^{n} f(B \mid \mu, \Sigma) \cdot P\left(\sigma^2\right) \cdot P(\mu) \cdot P\left(\Sigma^{-1}\right) \\
&= \prod_{i=1}^{n} (2\pi)^{-\frac{n_i}{2}} \cdot \left|\sigma^2 I_{ni}\right|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}\left(y_i - x_i\beta_i\right)^\top \left(\sigma^2 I_{n_i}\right)^{-1} \left(y_i - x_i\beta_i^T\right)\right\} |\Sigma|^{-\frac{1}{2}} \\
&\cdot \exp\left\{-\frac{1}{2}\left[\left(\beta_i - \mu\right)^\top \Sigma^{-1} \left(\beta_i - \mu\right)\right]\right\} \frac{1}{\sigma^2} |\Sigma|^{-(d+1)} \exp(-\frac{1}{2}\mathbf{\Sigma}^{-1})
\end{aligned}
$$

2.2 Markov Chain Monte Carlo Algorithm

3. Results

3.1 MCMC

3.2 Seasonal difference analysis

3.3 Prediction