

P8160 Group Project 3: Bayesian Modeling of Hurricane Trajectories

Anyu Zhu | Haotian Wu | Wenhan Bao | Yiming Li | Qihang Wu

5/6/2022

Background and Objective

Hurricanes are large rotating tropical storms with winds in excess of 119 kilometers per hour (74 mph). They usually form between June 1 and November 30 in the Atlantic Ocean but can develop in other oceans as well. They are known as typhoons in the western Pacific and cyclones in the Indian Ocean[1].

When a hurricane approaches land, tremendous damage can occur to the nearby cities. Therefore, scientists continue to improve their ability to forecast hurricanes. The sooner they can access accurate information about a hurricane's location and intensity, the better the chances to minimize the hurricane's impacts.

In this project, we firstly built a Bayesian model based on the track data of 703 hurricanes in the North Atlantic area since 1950. A Markov Chain Monte Carlo (MCMC) algorithm was designed to generate the distribution of parameters. With the start time and type of each hurricane, the estimated coefficients from the Bayesian model were used to explore the seasonal differences and wind speed changes over years. Finally, we also explored the characteristics of hurricanes associated with the damage and deaths.

Data Descriptions and Preprocessing

The first dataset hurrican703.csv collected the track data of 703 hurricanes in the North Atlantic area since 1950. For all the hurricanes, their locations (longitude & latitude) and maximum wind speed were recorded every 6 hours. The variables include ID, Season, starting Month, Nature, time, Latitude, Longitude, Wind.kt. We created three new variables: changes of latitude and longitude, as well as the wind speed between the time t and $(t - 6)$. For splitting the train and test set, we removed the hurricanes with observations less than 5. We built a regression model predicting wind speed based on this dataset.

The second dataset hurricanoutcome2.csv recorded the damages and death caused by 46 hurricanes in the U.S, and some features extracted from the hurricane records. To explore the characteristics related with death and damage, we combined the coefficients results from the first model with the second dataset by hurricane ID.

Statistical Methods

Posterior Distribution and Conditional Distributions

Likelihood

For each hurricane i and k_i 's time points, we have the following Bayesian model:

$$Y_i(t + 6) = \beta_{0i} + \beta_{1i}Y_i(t) + \beta_{2i}\Delta_{i1}(t) + \beta_{3i}\Delta_{i2}(t) + \beta_{4i}\Delta_{i3}(t) + \varepsilon_i(t),$$

where $Y_i(t)$ is the wind speed at time t , Δ_{i1} , Δ_{i2} , and Δ_{i3} are the changes of latitude, longitude, and the wind speed between time point t and $t-6$, respectively. $\varepsilon_i(t)$ follows a normal distributions with mean zero and variance σ^2 . The above Bayesian model can be simplified as:

$$Y_i(t+6) = x_i(t) + \varepsilon_i(t),$$

where $\beta_i = (\beta_{0i}, \beta_{1i}, \dots, \beta_{4i}) \sim N(0, \sigma^2)$. Based on the property of the multivariate linear regression model, for each hurricane i , we have:

$$Y_i | X_i \sim N_{k_i}(x_i \beta_i, \sigma^2 I_{k_i}),$$

where I_{k_i} is an identity matrix with k_i dimensions.

Thus, we can consider the following distribution of each hurricane i :

$$f(y_i | \beta_i, \sigma^2) = [(2\pi)^{k_i} \cdot \det(\sigma^2 I_{k_i})]^{-\frac{1}{2}} \cdot \exp \left\{ -\frac{1}{2} (y_i - x_i \beta_i)^\top (\sigma^2 I_{k_i})^{-1} (y_i - x_i \beta_i) \right\}$$

.

From above, we derive the following likelihood function:

$$\begin{aligned} f(y | B, \sigma^2) &= \prod_{i=1}^n f(y_i | \beta_i, \sigma^2) \\ &= \prod_{i=1}^n \left([(2\pi)^{k_i} \cdot \det(\sigma^2 I_{k_i})]^{-\frac{1}{2}} \cdot \exp \left\{ -\frac{1}{2} (y_i - x_i \beta_i)^\top (\sigma^2 I_{k_i})^{-1} (y_i - x_i \beta_i) \right\} \right) \end{aligned}$$

Prior distributions

We assume the following non-informative prior distributions:

$$\beta_i = (\beta_{0i}, \beta_{1i}, \dots, \beta_{4i}) \sim N_5(\mu, \Sigma),$$

where $B = (\beta_1^\top, \beta_2^\top, \dots, \beta_n^\top)^\top$ and n is the number of hurricanes. So,

$$\pi(B | \mu, \Sigma^{-1}) = \prod_{i=1}^n f(\beta_i) \propto \det(\Sigma)^{-n/2} \cdot \exp \left\{ -\frac{1}{2} \sum_i [(\beta_i - \mu)^\top (\Sigma)^{-1} (\beta_i - \mu)] \right\}.$$

Also, $\pi(\sigma^2) \propto \frac{1}{\sigma^2}$; $\pi(\mu) \propto 1$; $\pi(\Sigma^{-1}) \propto |\Sigma|^{-(d+1)} \cdot \exp(-\frac{1}{2}\Sigma^{-1})$.

Conditional posteriors

The posterior distribution is the product of the likelihood and the prior:

$$g(B, \sigma^2, \mu, \Sigma^{-1} | y) \propto f(y | B, \sigma^2) \cdot \pi(B | \mu, \Sigma^{-1}) \cdot \pi(\sigma^2) \cdot \pi(\mu) \cdot \pi(\Sigma^{-1}),$$

so we have:

$$\begin{aligned} \pi(\sigma^2 | \cdot) &\propto \prod_{i=1}^n \det(\sigma^2 I_{k_i})^{-\frac{1}{2}} \cdot \exp \left\{ -\frac{1}{2} \sum_i [(y_i - x_i \beta_i)^\top (\sigma^2 I_{k_i})^{-1} (y_i - x_i \beta_i)] \right\} \cdot \sigma^{-2} \\ &= (\sigma^2)^{-\frac{1}{2} \sum_i k_i} \cdot \exp \left\{ -\frac{1}{2\sigma^2} \sum_i [(y_i - x_i \beta_i)^\top (y_i - x_i \beta_i)] \right\} \cdot \sigma^{-2} \\ &= (\sigma^2)^{-1 - \frac{1}{2} \sum_i k_i} \cdot \exp \left\{ -\frac{1}{2\sigma^2} \sum_i \sum_{t_i} (y_{i,t} - x_{i,t} \beta_i)^2 \right\} \end{aligned}$$

Therefore, $\sigma^2 \sim \text{Inverse Gamma} \left(\frac{1}{2} \Sigma_i k_i, \frac{1}{2} \Sigma_i \Sigma_{t_i} (y_{i,t} - x_{i,t} \beta_i)^2 \right)$.

$$\begin{aligned}
\pi(\Sigma^{-1} \mid \cdot) &\propto \det(\Sigma)^{-n/2} \cdot \exp \left\{ -\frac{1}{2} \Sigma_i (\beta_i - \mu)^\top \Sigma^{-1} (\beta_i - \mu) \right\} \cdot \det(\Sigma)^{-(d+1)} \cdot \exp \left\{ -\frac{1}{2} \Sigma^{-1} \right\} \\
&= \det(\Sigma)^{-(n/2+d+1)} \cdot \exp \left\{ -\frac{1}{2} \left[\Sigma^{-1} + \Sigma_i (\beta_i - \mu)^\top \Sigma^{-1} (\beta_i - \mu) \right] \right\} \\
&\propto \det(\Sigma^{-1})^{(n+2d+2)/2} \cdot \exp \left\{ -\frac{1}{2} \text{tr} \left[\Sigma^{-1} \cdot \left(I + \Sigma_i (\beta_i - \mu) (\beta_i - \mu)^\top \right) \right] \right\} \\
&\propto \det(\Sigma^{-1})^{(n+3d+3-d-1)/2} \cdot \exp \left\{ -\frac{1}{2} \text{tr} \left[\Sigma^{-1} \cdot \left(I + \Sigma_i (\beta_i - \mu) (\beta_i - \mu)^\top \right) \right] \right\}
\end{aligned}$$

Thus $\Sigma^{-1} \sim \text{Wishart} \left(n + 3d + 3, [I + \Sigma_i (\beta_i - \mu) (\beta_i - \mu)^\top]^{-1} \right)$, that is:

$$\Sigma \sim \text{Inverse Wishart} \left(n + 3d + 3, I + \Sigma_i (\beta_i - \mu) (\beta_i - \mu)^\top \right)$$

$$\begin{aligned}
\pi(\mu \mid \cdot) &\propto \exp \left\{ -\frac{1}{2} \Sigma_i \left[(\beta_i - \mu)^\top \Sigma^{-1} (\beta_i - \mu) \right] \right\} \\
&= \exp \left\{ -\frac{1}{2} \Sigma_i \left(\beta_i^\top \Sigma^{-1} \beta_i + \mu^\top \Sigma^{-1} \mu - 2 \beta_i^\top \Sigma^{-1} \mu \right) \right\} \\
&= \exp \left\{ -\frac{1}{2} \left(\Sigma_i \beta_i^\top \Sigma^{-1} \beta_i + \mu^\top n \Sigma^{-1} \mu - 2 \Sigma_i \beta_i^\top \Sigma^{-1} \mu \right) \right\} \\
&= \exp \left\{ -\frac{1}{2} \left(\mu^\top n \Sigma^{-1} \mu - 2 \Sigma_i \beta_i^\top \Sigma^{-1} \mu + \Sigma_i \beta_i^\top \Sigma^{-1} \beta_i \right) \right\} \\
&= \exp \left\{ -\frac{1}{2} \left(\mu^\top \underbrace{n \Sigma^{-1}}_M \mu - 2 \mu^\top \underbrace{\Sigma_i \Sigma^{-1} \beta_i}_N + \Sigma_i \beta_i^\top \Sigma^{-1} \beta_i \right) \right\} \\
&= \exp \left\{ -\frac{1}{2} \left[(\mu - M^{-1} N)^\top M (\mu - M^{-1} N) \right] \right\}.
\end{aligned}$$

Therefore, $\mu \sim \text{MVN} (M^{-1} N, M^{-1})$.

$$\begin{aligned}
\pi(B \mid \cdot) &\propto \exp \left\{ -\frac{1}{2} \Sigma_i \left[(y_i - x_i \beta_i)^\top (\sigma^2 I_{k_i})^{-1} (y_i - x_i \beta_i) \right] \right\} \cdot \exp \left\{ -\frac{1}{2} \Sigma_i \left[(\beta_i - \mu)^\top (\Sigma)^{-1} (\beta_i - \mu) \right] \right\} \\
&= \exp \left\{ -\frac{1}{2} \Sigma_i \left[(y_i - x_i \beta_i)^\top (\sigma^2 I_{k_i})^{-1} (y_i - x_i \beta_i) + (\beta_i - \mu)^\top \Sigma^{-1} (\beta_i - \mu) \right] \right\} \\
&= \exp \left\{ -\frac{1}{2} \Sigma_i \left[y_i^\top \sigma^{-2} I_{k_i} y_i + \beta_i^\top x_i^\top \sigma^{-2} I_{k_i} x_i \beta_i - 2 y_i^\top \sigma^{-2} I_{k_i} x_i \beta_i + \beta_i^\top \Sigma^{-1} \beta_i + \mu^\top \Sigma^{-1} \mu - 2 \mu^\top \Sigma^{-1} \beta_i \right] \right\} \\
&= \exp \left\{ -\frac{1}{2} \Sigma_i \left[y_i^\top \sigma^{-2} I_{k_i} y_i + \mu^\top \Sigma^{-1} \mu + \beta_i^\top (\Sigma^{-1} + x_i^\top \sigma^{-2} I_{k_i} x_i) \beta_i - 2 (y_i^\top \sigma^{-2} I_{k_i} x_i + \mu^\top \Sigma^{-1}) \beta_i \right] \right\}
\end{aligned}$$

We can define the following terms:

$$\begin{aligned}
R &= y_i^\top \sigma^{-2} I_{k_i} y_i + \mu^\top \Sigma^{-1} \mu \\
V &= \Sigma^{-1} + x_i^\top \sigma^{-2} I_{k_i} x_i \\
M &= \sigma^{-2} x_i^\top y_i + \Sigma^{-1} \mu
\end{aligned}$$

Thus, $\pi(B \mid \cdot) \propto (\beta_i - V^{-1} M)^\top V (\beta_i - V^{-1} M) \sim \text{MVN} (V^{-1} M, V^{-1})$

Gibbs Sampling

Since directly generating the above jointly density is rather complicated, we can implement the gibbs sampler to sample each variable in turn. Based on the previous conditional distributions, we updated the four parameters including B , σ^2 , \sum , and μ in sequence. Finally, we selected 10,000 iterations. Note that we ignored the first 8,000 iterations to ensure the stationary distribution of the Markov chain was reached.

Lasso and Poisson Regression Model

Results

Gibbs Sampling

Task 5

Task 6

Discussion and Limitations

Contributions

Reference

Appendix

For codes please click [here](#)