Report

Margaret Gacheru, Melanie Mayer, Kee-Young Shin, Adina Zhang

May 10, 2020

Introduction

Hurricanes are type of storms with high speed winds that form over tropical or subtropical waters. Typically, hurricanes bring about strong winds, storm surges, and heavy rainfall that can lead to flooding, tornadoes, etc. The Saffir-Simpson Hurricane scale is used to rate hurricanes by their wind speed – the higher the category, the greater the possibility of landfall damage. It is worthwhile to understand aspects of hurricanes and make accurate predictions in order to provide appropriate information and resources to the public, make informed decisions, and potentially save lives.

For this project, we have information about 356 hurricanes in the North Atlantic area since 1989, including the storm's location and maximum windspeed for every 6 hours. Additionally, the dataset includes information about season and month which hurricane occured, nature of the hurricane, and timing of the storm. We are interested in building a bayesian model in order to predict hurricane trajectories, particularly the wind speed at a specific time point. We will utilize Markov Chain Monte-Carlo to estimate the model parameters and their 95% credible intervals through the posterior distributions. Furthermore, we are interested in evaluating our model's ability to accurately predict windspeed and track the hurricanes.

Methods

Bayesian and MCMC Basics

Unlike frequentist approaches, Bayesian analysis treats parameters as random variables and utilizes prior beliefs about the parameters. We start out with a prior distribution of the parameter of interest and then update it with observed data to obtain the posterior distribution. With the posterior distribution, we can obtain parameters estimates such as posterior mean. Bayes Theorem allows us to observe the relationship between prior distribution ($\pi(\theta)$), likelihood ($f(x|\theta)$), and posterior distribution ($\pi(\theta|x)$)

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)} = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta}$$

Obtaining the posterior distribution is often a difficult task – either there is no closed form or it is computationally intensive to directly sample from $\pi(\theta|x)$. Therefore, MCMC is used to create a Markov chain of the parameter of interest such that their distribution converge to the posterior.

Hurricane Bayesian Model

The Bayesian model used to predict hurricane trajectories can be expressed as follows: $Y_i(t+6) = \mu_i(t) + \rho_i Y_i(t) + \epsilon_i(t)$ where ρ_j is the autoregressive correlation and the error ϵ_i follows the normal distribution with mean 0 and variance σ^2 independent across t. $\mu_i(t)$ represents the function mean that can be written as follows:

$$\mu_i(t) = \beta_0 + \beta_1 x_{i,1}(t) + \beta_2 x_{i,2}(t) + \beta_3 x_{i,3}(t) + \sum_{k=1}^{3} \beta_{3+k} \triangle_{i,k}(t-6)$$

where $x_{i,1}(t)$, $x_{i,2}(t)$, and $x_{i,3}(t)$ are the day of the year, calendar year, and type of hurricane at time t, respectively. and

$$\Delta_{i,k}(t-6) = Y_{i,k}(t) - Y_{i,k}(t-6), k = 1, 2, 3$$

are the change of latitude, longitude, and wind speed from time $t-6$ to t. The following prior distributions were assumed: $\pi(\boldsymbol{\beta})$ is jointly normal with mean 0 and variance $diag(1, p)$, $\pi(\rho_j)$ follows a truncated normal $N_{[0,1]}(0.5, 1/5)$, and $\pi(\sigma^{-2})$ follows an inverse-gamma $(0.001, 0.001)$.

Metropolis-Hastings Algorithm

To estimate the parameters, a component-wise Metropolis-Hastings (MH) algorithm was suggested, with a uniform proposal for β and ρ, and a inverse-gamma proposal for σ^2. The MH algorithm constructs a Markov chain by accepting candidate points with probability

$$\alpha(y|x^{(t)}) = min(1, \frac{\pi(y)q(x^{(t)}|y)}{\pi(x^{(t)})q(y|x^{(t)})})$$

The random-walk Metropolis is a special class where the proposed transition q is symmetric. As a result, the acceptance probability is only related to $\pi(y)$, in our case the prior distributions and the likelihood. The response wind speed was assumed to follow a normal distribution and the log-likelihood could be expressed as

$$l(\beta, \rho) = -\frac{n}{2}log(2\pi) - nlog(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y - X^T\beta - \rho Y)^2$$

The partial posterior distribution can be formed by the product of the log-likelihood and the priors. For the MH algorithm, a burn in was incorporated to ensure analysis of a stationary distribution.

Results

Discussion