P8160 - Bayesian Modeling of Hurricane Trajectories

Paulina Han, Waveley Qiu, Yida Wang,
Lin Yang, Jibei Zheng, Haolin Zhong

2022-05-09

Introduction

Data Praparation

- Self-join was performed to generate data for Gibbs sampling
- For each hurricane, 80% record was randomly selected and assigned to the train dataset, and the rest are in the test dataset. Hurricanes with less than 5 records were removed.

Gibbs Sampling

Initialize $\Theta_0 = (\mathbf{B}_0, \boldsymbol{\mu}_0, \sigma_0^2, \Sigma_0)$

for iteration i = 1,2,... **do**

Sample $\mathbf{B}_i \sim \pi(\mathbf{B}|\boldsymbol{\mu}_{i-1}, \sigma_{i-1}^2, \Sigma_{i-1}, \mathbf{Y})$

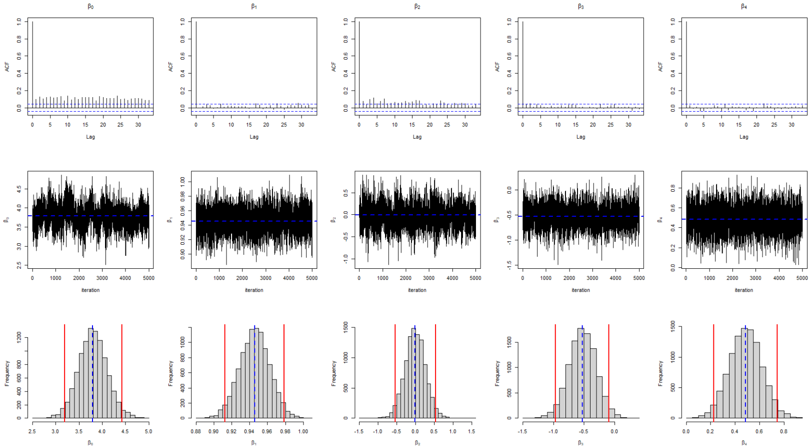
Sample $\boldsymbol{\mu}_i \sim \pi(\boldsymbol{\mu}|\mathbf{B}_i, \sigma_{i-1}^2, \Sigma_{i-1}, \mathbf{Y})$

Sample $\sigma_i^2 \sim \pi(\sigma^2|\mathbf{B}_i, \boldsymbol{\mu}_i, \Sigma_{i-1}, \mathbf{Y})$

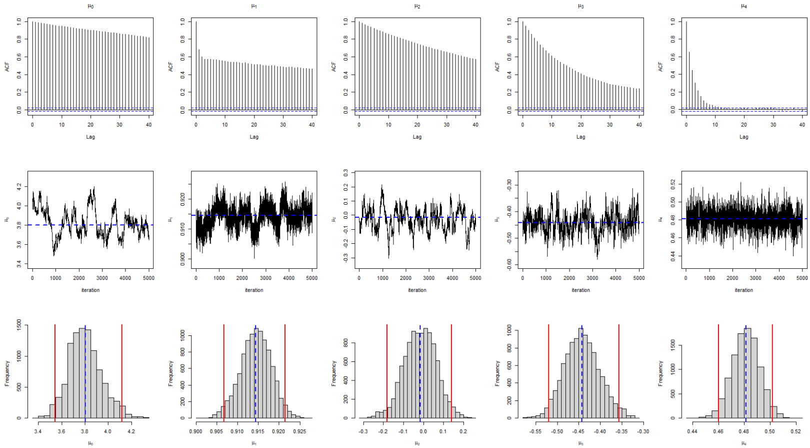
Sample $\Sigma_i \sim \pi(\Sigma|\mathbf{B}_i, \boldsymbol{\mu}_i, \sigma_i^2, \mathbf{Y})$

end for

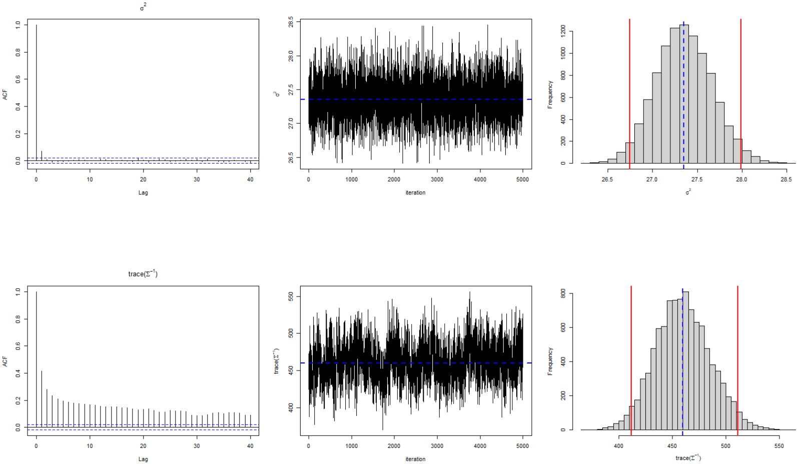
Results



Results



Results



Forcasting Hurricane Impact

ID: ID of the hurricanes
Season: In which year the hurricane occurred
Month: In which month the hurricane occurred
Nature: Nature of the hurricane
Damage: Financial loss (in Billion U.S. dollars) caused by hurricanes
Deaths: Number of death caused by hurricanes
Maxspeed: Maximum recorded wind speed of the hurricane
Meanspeed: Average wind speed of the hurricane
Maxpressure: Maximum recorded central pressure of the hurricane
Meanpressure: Average central pressure of the hurricane
Hours: Duration of the hurricane in hours
Total.Pop: Total affected population
Percent.Poor: % affected population that reside in low GDP counties
Percent.USA: % affected population that reside in the United States

LASSO Model for Damage

	Coefficients
(Intercept)	-533.5099174
season	3.3837824
deaths	0.0000000
monthJuly	0.0000000
monthJune	0.0000000
monthNovember	0.0000000
monthOctober	0.0000000
monthSeptember	0.0000000
natureNR	0.0000000
natureTS	0.0000000
maxspeed	1.2117851
meanspeed	0.0000000
maxpressure	0.0000000
meanpressure	0.0000000
hours	0.0000000
total_pop	0.3187361
percent_poor	0.0000000
percent_usa	0.7073409
beta0	0.0000000
beta1	0.0000000
beta2	0.0000000
beta3	0.0000000
beta4	0.0000000

Refitted Linear Regression Model

	Coefficients
(Intercept)	-1316.7386136
season	0.6485139
maxspeed	0.1968674
total_pop	0.0000033
percent_usa	0.1356486

Poisson Model for Dealths

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-199.5331639	11.8792784	-16.7967411	0.0000000
season	-0.0404185	0.0028048	-14.4104991	0.0000000
damage	0.0220163	0.0005679	38.7649762	0.0000000
monthJuly	-10.2286750	0.1604645	-63.7441688	0.0000000
monthJune	0.3928062	0.0989170	3.9710698	0.0000716
monthNovember	1.8733767	0.1664682	11.2536625	0.0000000
monthOctober	-1.6041896	0.0787720	-20.3649754	0.0000000
monthSeptember	1.2490015	0.0575033	21.7205350	0.0000000
natureNR	2.0903864	0.1371766	15.2386495	0.0000000
natureTS	-1.1903051	0.1118619	-10.6408484	0.0000000
maxspeed	0.0035207	0.0013988	2.5168778	0.0118400
meanspeed	-0.1978651	0.0039977	-49.4953412	0.0000000
maxpressure	0.0048106	0.0075485	0.6372945	0.5239331
meanpressure	0.0021204	0.0001759	12.0515409	0.0000000
total_pop	0.0000009	0.0000000	31.4237737	0.0000000
percent_poor	0.0873434	0.0010058	86.8433730	0.0000000
percent_usa	-0.0080185	0.0004884	-16.4173000	0.0000000
beta0	41.3531048	0.5634443	73.3934206	0.0000000
beta1	132.8784572	1.9305164	68.8305252	0.0000000
beta2	-10.7339527	0.5001340	-21.4621524	0.0000000
beta3	-0.4736994	0.5091748	-0.9303277	0.3522014
beta4	4.4919244	0.1971025	22.7897893	0.0000000

Discussion

Strength

- Unlike classical modeling methods, the MCMC approach bypass coefficient optimization process and directly sample coefficients from their distributions
- Optimization methods may vary from models to models, while we only need to derive posterior conditional distribution for each coefficients when using Gibbs Sampling.

Limitation

- MCMC approaches are often computationally expensive and can be inefficient since they involve thousands of rounds of sampling and updating.
- Convergence is not guaranteed.

Why Non-convergence?

- $\boldsymbol{\beta}_i \sim N(\boldsymbol{\beta}, \Sigma)$ may be a too strong assumption

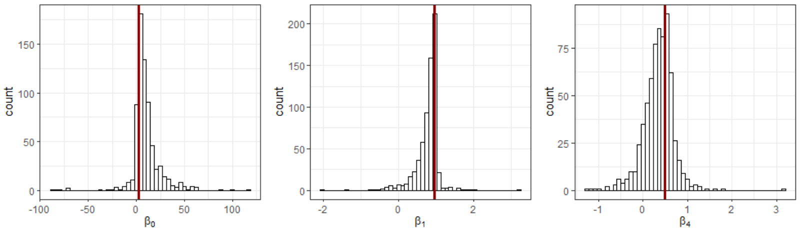


Figure 1: Distribution of $\boldsymbol{\beta}_i$s obtained by performing OLS for each hurricane (red line: $\boldsymbol{\beta}$ obtained by performing OLS on the whole training dataset)

- **Future work**: use a more adequate distribution assumption of $\boldsymbol{\beta}_i$ which can account for skewness