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%/20% of records were randomly assigned to the training/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 $\mathbf{B}_i \sim \pi(\mathbf{B}|\boldsymbol{\mu}_{i-1}, \sigma_{i-1}^2, \Sigma_{i-1}, \mathbf{Y})$

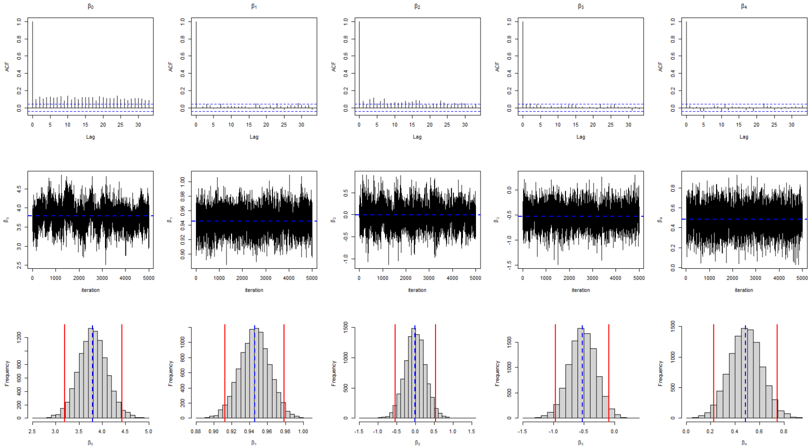
Sample $\boldsymbol{\mu}_i \sim \pi(\boldsymbol{\mu}|\mathbf{B}_i, \sigma_{i-1}^2, \Sigma_{i-1}, \mathbf{Y})$

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

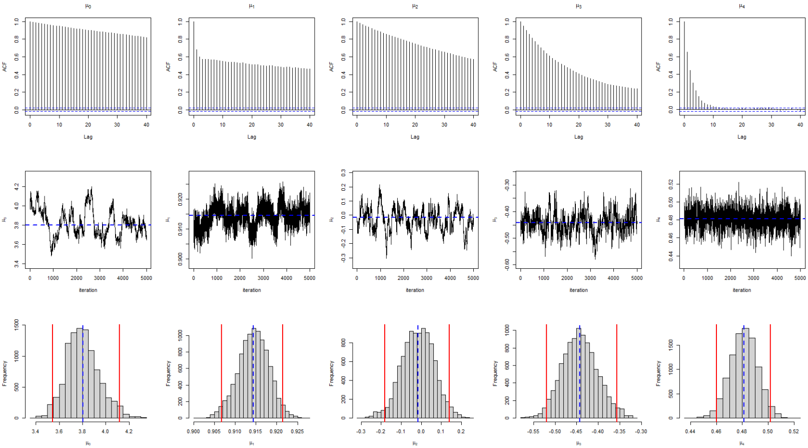
Sample $\Sigma_i \sim \pi(\Sigma|\mathbf{B}_i, \boldsymbol{\mu}_i, \sigma_i^2, \mathbf{Y})$

end for

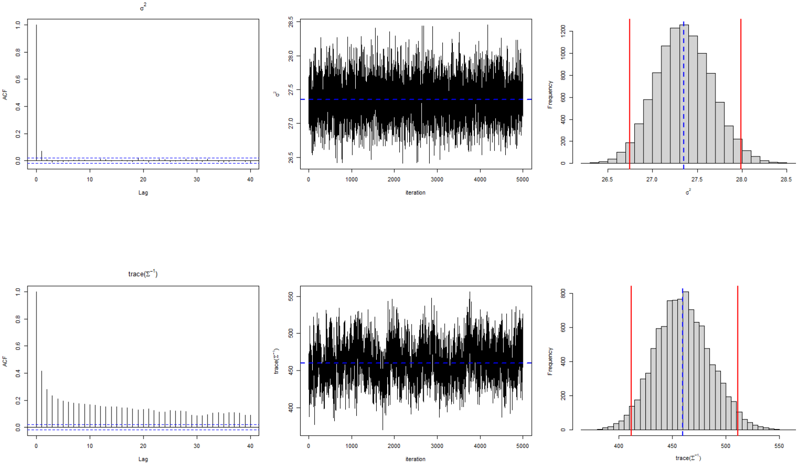
Results



Results

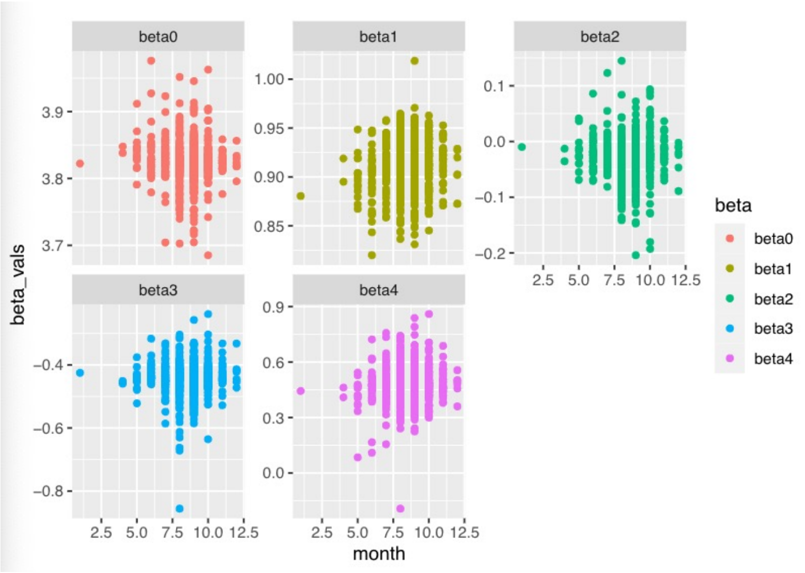


Results



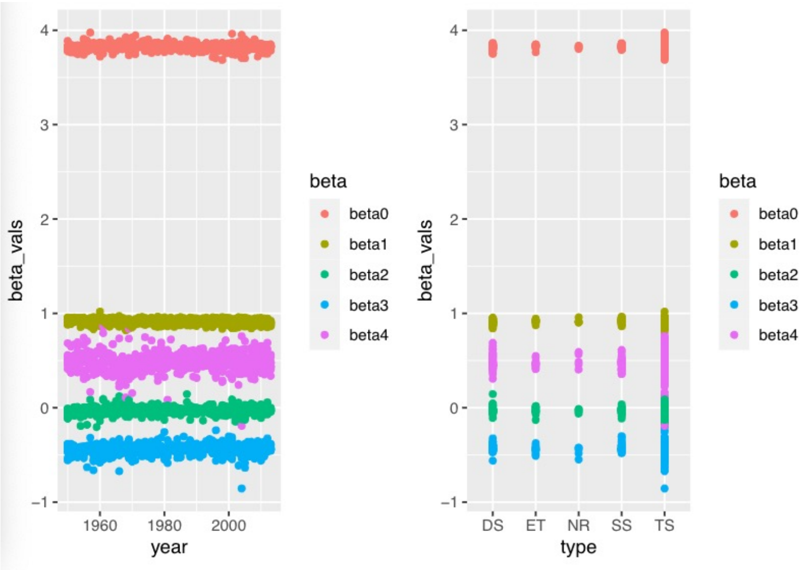
σ²
ACF
Lag
σ²
iteration
Frequency
σ²
trace(Σ⁻¹)
ACF
Lag
trace(Σ⁻¹)
iteration
Frequency
trace(Σ⁻¹)

EDA of β_i



beta0
beta1
beta2
beta3
beta4
beta_vals
month
beta
3.9
3.8
3.7
1.00
0.95
0.90
0.85
0.1
0.0
−0.1
−0.2
−0.4
−0.6
−0.8
0.9
0.6
0.3
0.0
2.5 5.0 7.5 10.0 12.5

EDA of β_i



beta_vals
year
1960
1980
2000
4
3
2
1
0
−1
beta
beta0
beta1
beta2
beta3
beta4
beta_vals
type
DS
ET
NR
SS
TS

Seasonal Analysis

Model 1:

$$Beta_i = \alpha_0 + \alpha_1 I(\mathrm{Month} = M) + \alpha_2 \times \mathrm{Year} + \alpha_3 I(\mathrm{Type} = N)$$

Coefficients of Model 1 for β_i

	Beta0		Beta1		Beta2		Beta3		Beta4	
	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)
(Intercept)	3.8722305	0.0000000	3.8722305	0.0000000	3.8722305	0.0000000	3.8722305	0.0000000	3.8722305	0.0000000
factor(month)4	0.0211436	0.5897652	0.0211436	0.5897652	0.0211436	0.5897652	0.0211436	0.5897652	0.0211436	0.5897652
factor(month)5	0.0202412	0.5405684	0.0202412	0.5405684	0.0202412	0.5405684	0.0202412	0.5405684	0.0202412	0.5405684
factor(month)6	0.0159306	0.6239449	0.0159306	0.6239449	0.0159306	0.6239449	0.0159306	0.6239449	0.0159306	0.6239449
factor(month)7	0.0078141	0.8090619	0.0078141	0.8090619	0.0078141	0.8090619	0.0078141	0.8090619	0.0078141	0.8090619
factor(month)8	0.0019068	0.9527290	0.0019068	0.9527290	0.0019068	0.9527290	0.0019068	0.9527290	0.0019068	0.9527290
factor(month)9	0.0009337	0.9768273	0.0009337	0.9768273	0.0009337	0.9768273	0.0009337	0.9768273	0.0009337	0.9768273
factor(month)10	0.0074737	0.8163761	0.0074737	0.8163761	0.0074737	0.8163761	0.0074737	0.8163761	0.0074737	0.8163761
factor(month)11	0.0057884	0.8588527	0.0057884	0.8588527	0.0057884	0.8588527	0.0057884	0.8588527	0.0057884	0.8588527
factor(month)12	0.0048248	0.8874129	0.0048248	0.8874129	0.0048248	0.8874129	0.0048248	0.8874129	0.0048248	0.8874129
year	-0.0000290	0.6794636	-0.0000290	0.6794636	-0.0000290	0.6794636	-0.0000290	0.6794636	-0.0000290	0.6794636
factor(type)ET	0.0075408	0.4379949	0.0075408	0.4379949	0.0075408	0.4379949	0.0075408	0.4379949	0.0075408	0.4379949
factor(type)NR	0.0005575	0.9705947	0.0005575	0.9705947	0.0005575	0.9705947	0.0005575	0.9705947	0.0005575	0.9705947
factor(type)SS	0.0074733	0.2505823	0.0074733	0.2505823	0.0074733	0.2505823	0.0074733	0.2505823	0.0074733	0.2505823
factor(type)TS	0.0057948	0.2474418	0.0057948	0.2474418	0.0057948	0.2474418	0.0057948	0.2474418	0.0057948	0.2474418

Seasonal Analysis

Model 2:

$$Beta_i = \alpha_0 + \alpha_1 I(\text{Season} = S) + \alpha_2 \times \text{Year} + \alpha_3 I(\text{Type} = N)$$

Coefficients of Model 2 for β_i

	Beta 0		Beta 1		Beta 2		Beta 3		Beta 4	
	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)
(Intercept)	3.8777185	0.0000000	1.4515958	0.0000000	-0.1108749	0.4777627	-0.7432739	0.0009368	0.8246599	0.0469629
factor(season)spring	0.0165743	0.0386414	-0.0161701	0.0195181	0.0080903	0.3613959	-0.0095303	0.4529259	-0.0700505	0.0029939
factor(season)summer	0.0017442	0.4921921	-0.0054774	0.0126824	-0.0021923	0.4356701	-0.0020772	0.6061277	-0.0145208	0.0520294
factor(season)winter	0.0004419	0.9695014	-0.0175939	0.0782636	0.0012626	0.9214169	0.0101974	0.5781775	-0.0099913	0.7687012
year	-0.0000297	0.6721280	-0.0002706	0.0000093	0.0000429	0.5814228	0.0001589	0.1539417	-0.0001639	0.4270083
factor(type)ET	0.0090086	0.3383681	0.0086688	0.2860439	-0.0058182	0.5765098	-0.0167604	0.2616162	-0.0206719	0.4547698
factor(type)NR	0.0017339	0.9077767	0.0079185	0.5401551	-0.0073224	0.6586863	-0.0384250	0.1059825	0.0085343	0.8462012
factor(type)SS	0.0077248	0.2318002	0.0080589	0.1486481	-0.0023967	0.7374924	0.0006525	0.9492157	-0.0222179	0.2419930
factor(type)TS	0.0047623	0.3404235	0.0024950	0.5629051	-0.0029696	0.5913088	-0.0157040	0.0477883	-0.0093683	0.5233984

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?

- $\beta_i \sim N(\beta, \Sigma)$ may be a too strong assumption

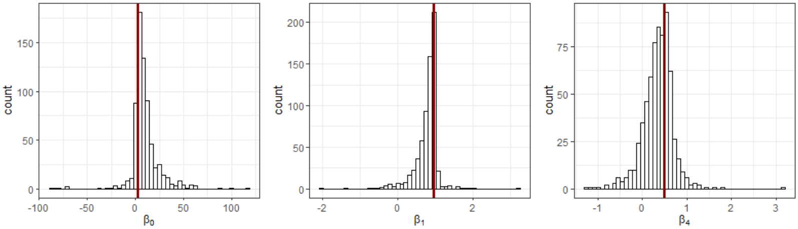


Figure 1: Distribution of β_is obtained by performing OLS for each hurricane (red line: β obtained by performing OLS on the whole training dataset)

- **Future work**: use a more adequate distribution assumption of β_i which can account for skewness