wenhao part for report

Wenhao Gou | wg2364

4/25/2021

```
library(tidyverse)
```

```
## -- Attaching packages -------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.0     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

THIS IS THE DISUSSION PART ON INTERCEPT - WENHAO GOU

Discussion on MC chain:

We can see that, expected for the intercept, all other variables converged in very few steps. To investigate this problem in detail, several steps are considered:

1. Create a very long chain:

One potential problem lied in this may due to the fact that we do not have a long-enough chain for this variable. So, we extended our chain to 20000, and got the following plot:
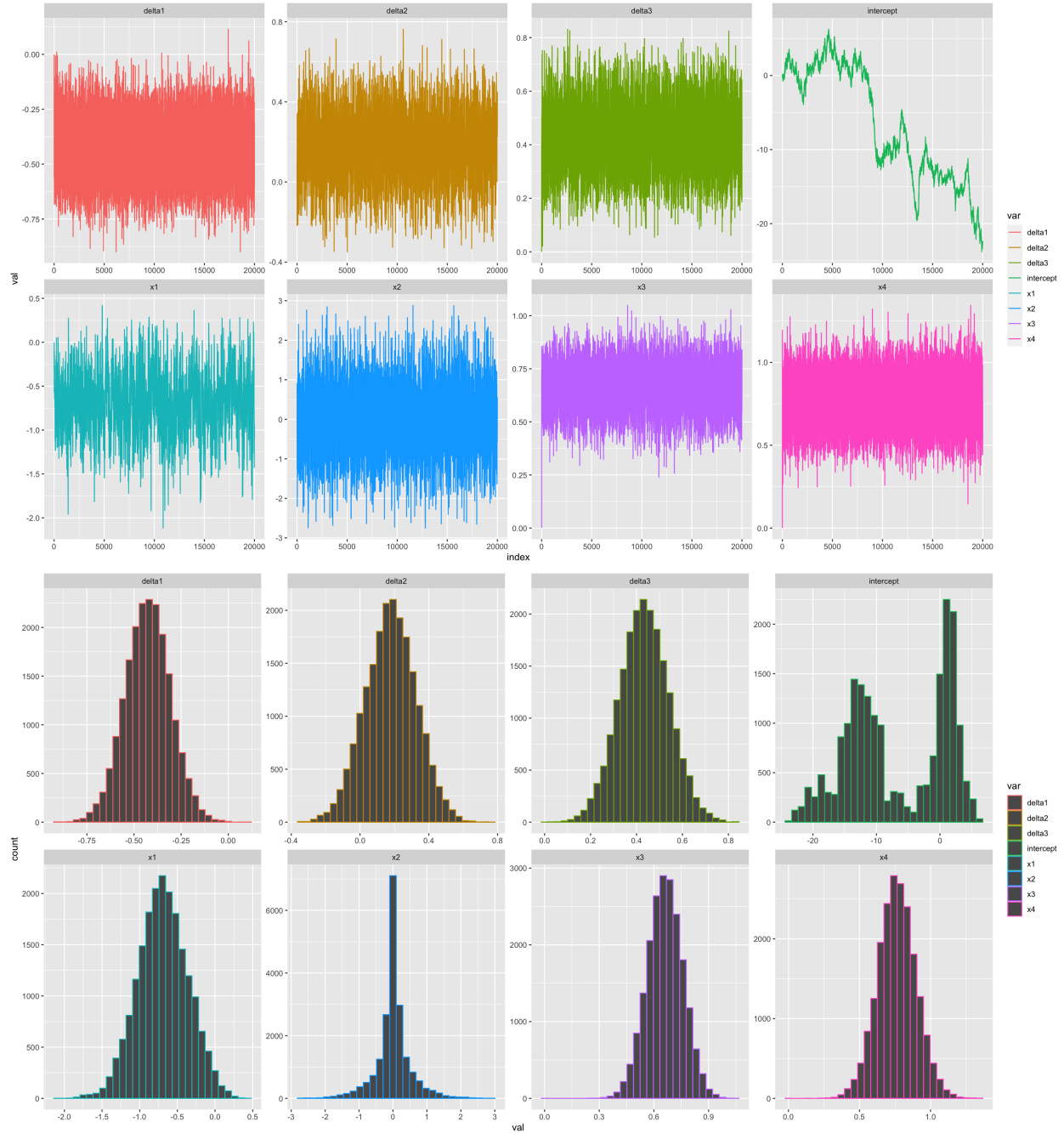
we can see that, the intercept still diverge, and appeared as a two-peak normal distribution. This suggested that some other parts in the model itself is questionable.

2. Investigate the models root mean square error:

Though the intercept do not converge, we can use other methods to see whether our *whole model* converged. To do so, SSR (sum square error) is considered. For each iteration, SSR for each hurricane can be calculated. If the SSR converged after some steps, it may be proper to say that our chain converged. Investigation on the *long chain generated in part 1(NOT DONE, I MAY RERUN IT TOMORROW)* have been extracted:

(MY R JUST DOWN, PLS MAKE A TRACE PLOT OF RSS AND DISTRIBUTINO PLOT OF RSS, LEAVE THE CODE! => use ggplot pls, use code in try.gibbs as it have buid a function to extract rss =>

2

can be used in further analysis without intercept too)

(THIS IS A DEMO:)

Histogram of .

.
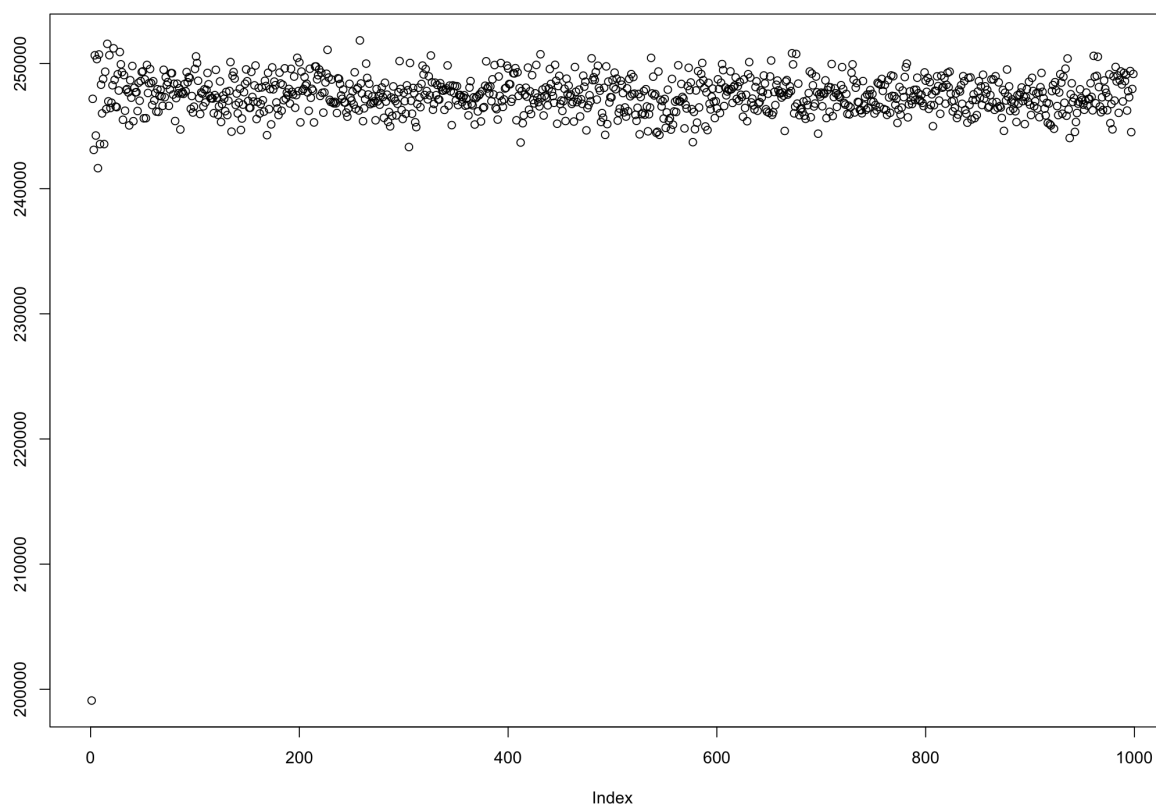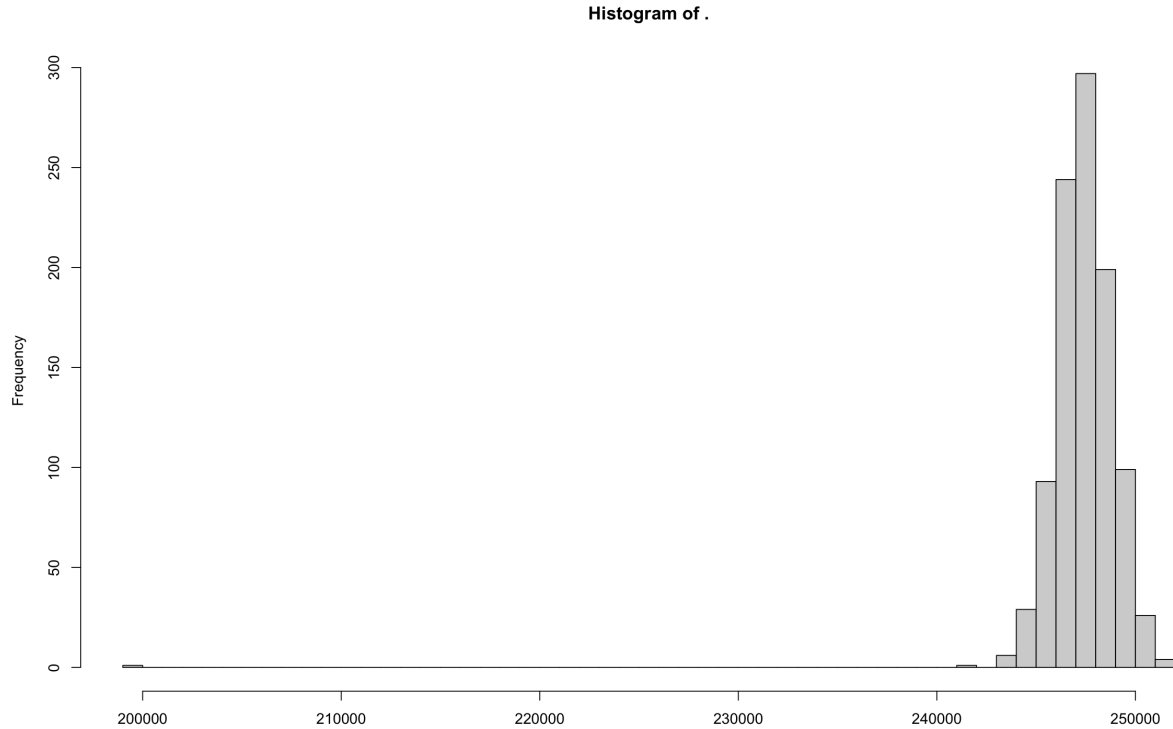
We can see that, SSR is normally distributed and converged. This suggested that our model is actually converged.

3. Issue in weak prior

Under the weak prior, by equation (), the posteria distribution will be a multivariate normal distribution. This suggested that the marginal distribution of the intercept will be a normal distribution. In our model setting, each intercepts for each hurricanes can be view as a baseline effect toward each wind speed. But, wind speed of each hurricanes are heterogeneous, and the distribution of wind speed is highly right skewed (see fig xxx), this suggested that use a prior that result in the posteria that individual level of intercepts follow a normal distribution may be invalid. This issue can be viewed as a

```
# hist(data$wind_kt)
```

To predict the wind speed at $t + 6$ in our model, we also include the lag 1 term x_4, which described the wind speed at time t. From previous results, the coefficient of this lag term is very close to 1, suggested that the variability in wind speed can be mainly explained by this term. Combined with other "significant" non-zero predictors, variability may be fully explained.

So, it may be proper to fit a model without the intercept term:

95% CI Intervals:

```
load("./no_int_betasummary.Rdata")
library(bayestestR)

ci_x1 <- ci(betasummary$x1, method = "ETI", ci = 0.95)
hat_x1 <- mean(betasummary$x1)
```

```r
ci_x2 <- ci(betasummary$x2, method = "ETI", ci = 0.95)
hat_x2 <- mean(betasummary$x2)

ci_x3 <- ci(betasummary$x3, method = "ETI", ci = 0.95)
hat_x3 <- mean(betasummary$x3)

ci_x4 <- ci(betasummary$x4, method = "ETI", ci = 0.95)
hat_x4 <- mean(betasummary$x4)

ci_d1 <- ci(betasummary$delta1, method = "ETI", ci = 0.95)
hat_d1 <- mean(betasummary$delta1)

ci_d2 <- ci(betasummary$delta2, method = "ETI", ci = 0.95)
hat_d2 <- mean(betasummary$delta2)

ci_d3 <- ci(betasummary$delta3, method = "ETI", ci = 0.95)
hat_d3 <- mean(betasummary$delta3)


ci_beta <- tibble(
  x1 = c(ci_x1$CI_low, hat_x1, ci_x1$CI_high),
  x2 = c(ci_x2$CI_low, hat_x2, ci_x2$CI_high),
  x3 = c(ci_x3$CI_low, hat_x3, ci_x3$CI_high),
  x4 = c(ci_x4$CI_low, hat_x4, ci_x4$CI_high),
  d1 = c(ci_d1$CI_low, hat_d1, ci_d1$CI_high),
  d2 = c(ci_d2$CI_low, hat_d2, ci_d2$CI_high),
  d3 = c(ci_d3$CI_low, hat_d3, ci_d3$CI_high)
) %>% t()

colnames(ci_beta) = c("2.5%","Beta_hat", "97.5%")

ci_beta %>% knitr::kable()
```

	2.5%	Beta_hat	97.5%
x1	-1.4676786	-0.7634907	-0.1076683
x2	-1.1794612	0.0166475	1.2195209
x3	0.4750182	0.6608306	0.8522768
x4	0.5142001	0.7687571	1.0351883
d1	0.2156073	0.4292543	0.6372581
d2	-0.1097564	0.1757811	0.4506797
d3	-0.6569219	-0.4243904	-0.1912513

```r
#fit model here
```

```r
#and the conclusions
```

From the trace plot and density plot, we can see that, all 7 left coefficients converged very quick, and the result do not have much discrepancy compare to the previous results (need check jareds results).To do further comparison on this two models, RSS are calculated for both model:

We can see that, the didn't depart much, and the distribution is all near normal distribution

(FURTHER ILLUSTRATIONS)