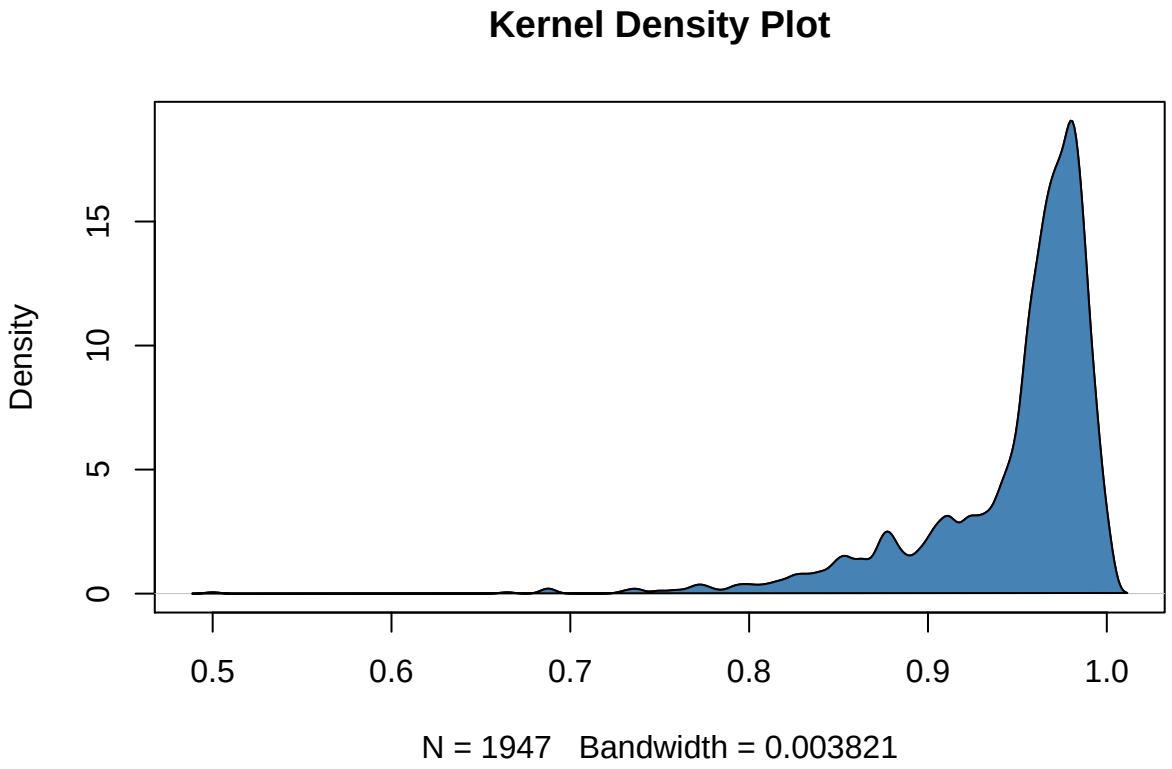
Data visualization

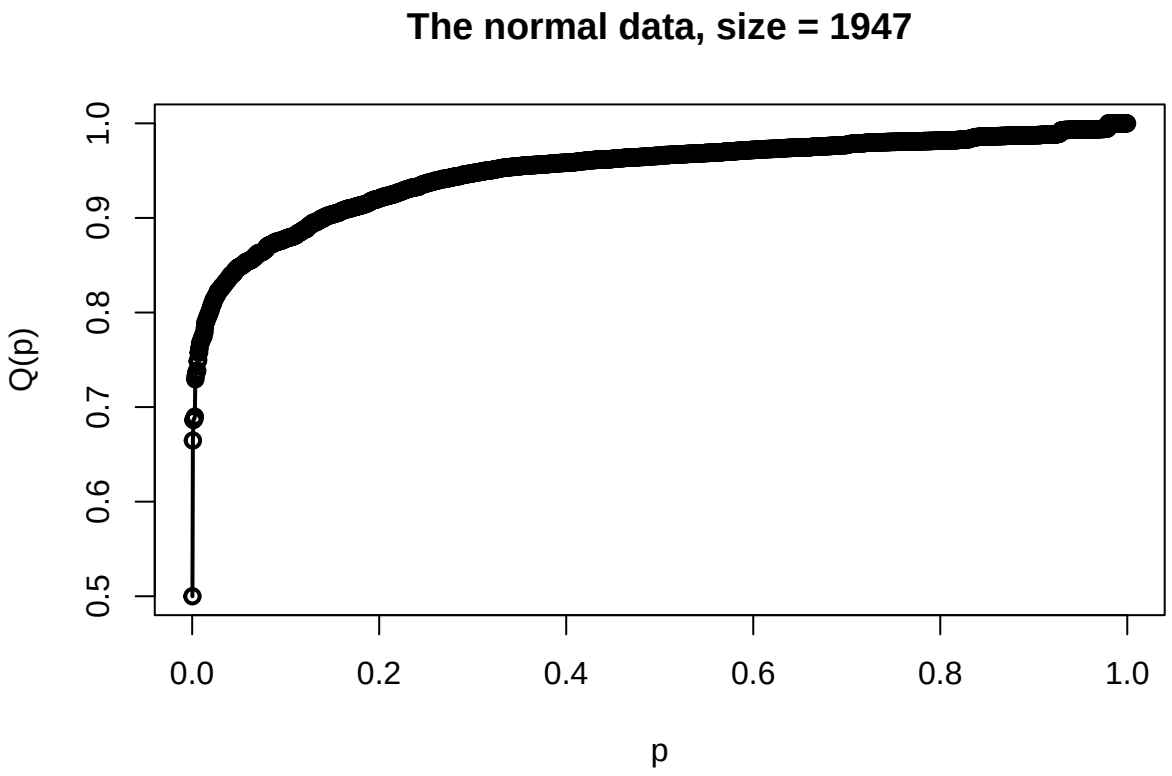
Bohan Zhang

4/11/2021

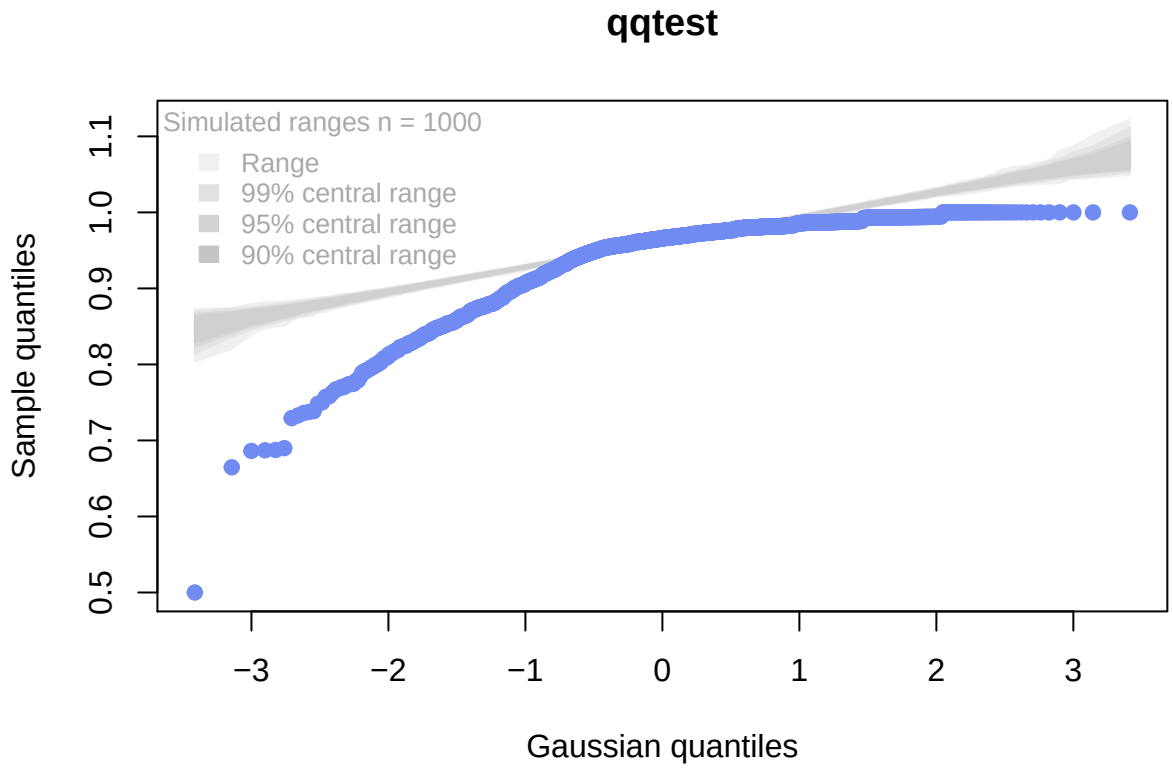
Analysis for average accuracy per day



Left skewed



Not fitting the shape for Gaussian distribution, should have a S shape with two long tails.

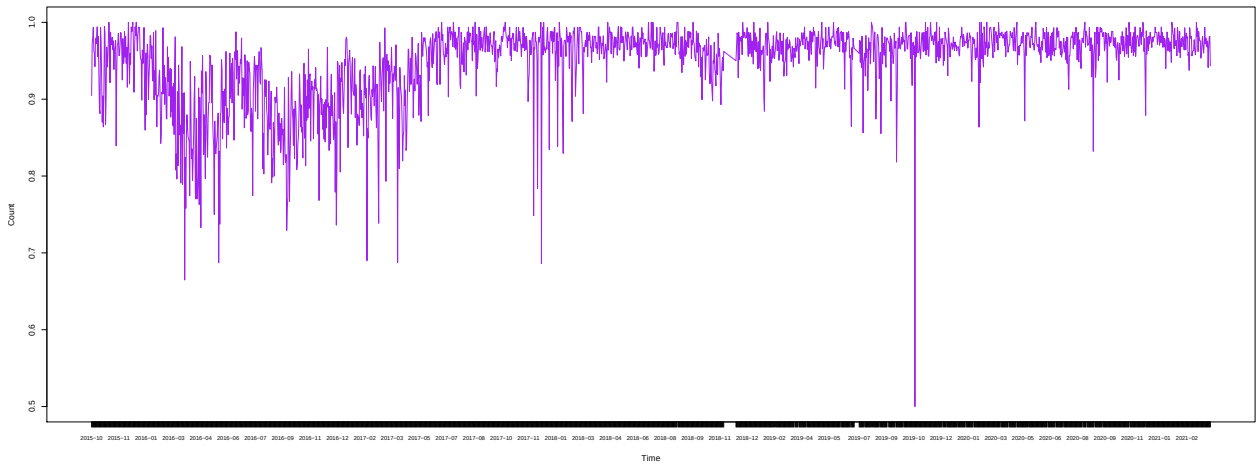


Most data is not in the envelop.

This concludes, the accuracy doesn't follow a normal distribution. Though the average is high, the model is not stable.

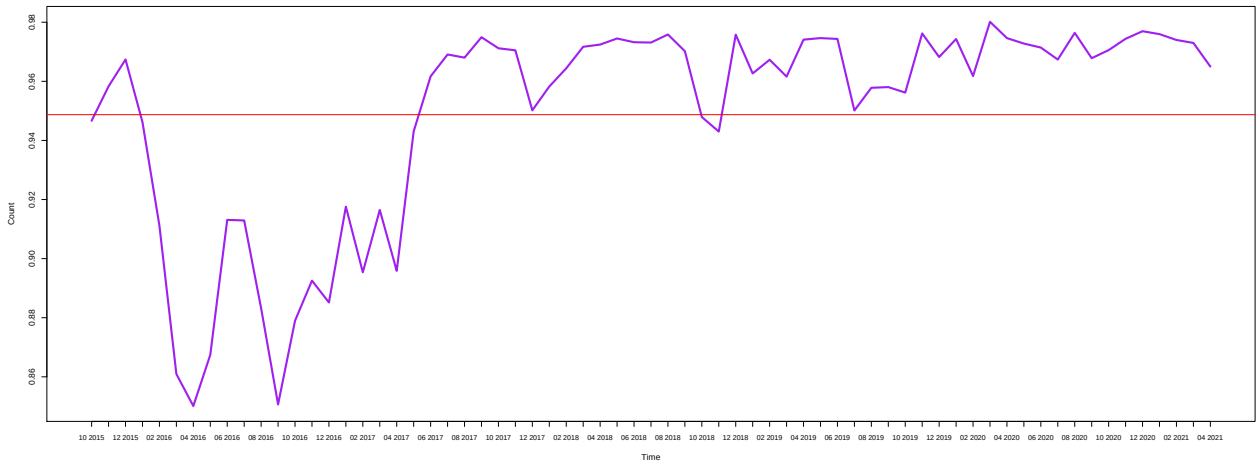
The trend of average accuacy per day

```
plot(dataProcess$Accuracy ~ dataProcess$Time, xlab="Time", ylab = "Count" , xaxt = "n", type = "l",
     col='purple',cex=19,lwd = 1)
axis(1, dataProcess$Time, format(dataProcess$Time, "%Y-%m"), cex.axis = .7)
```



Averge accuracy per month

```
plot(dataProcess2$Accuracy ~ dataProcess2$Time, xlab="Time", ylab = "Count" , xaxt = "n", type = "l",
     col='purple',cex=19, lwd = 4)
axis(1, dataProcess2$Time, format(dataProcess2$Time, "%m %Y"), cex.axis = 0.8)
abline(h = mean(dataProcess$Accuracy), col = "red")
```



Reason behind this?

From 900 dollars to 20,000 dollars: Bitcoin's Historic 2017 Price Run Revisited

We found out, Bitcoin price start change dramatically in May 2017. And that is the time, people start to get to know bitcoin and talking about bitcoin. And the performance of our model start to get better at that time.

So, the stability of our model could possibly get affected by the concentration of people. And how it could be quantified, and use it in our model?

We will explore the past data of Reddit.