Project Team Members:
Haodong Lu hl821
Zhengjuan Fan zf67
Chengyuan Deng cd751
Xiao Liu xl500

Hurricanes Time Distribution

I. Introduction

Known as one of the natural disasters happen in the costal areas, though rare, hurricanes are always destructive and result in enormous losses, which are likely to be reduced at utmost if knowledge on the time distribution law of hurricanes can be applied to the prediction of hurricanes happening in a certain area. In this project, 3 reasonable distributions were tested based on the open-source data about Atlantic hurricanes in the past 170 years, namely, Poisson distribution, exponential distribution and normal distribution.

II. Data

1. Data resources:

   https://en.wikipedia.org/wiki/List_of_Category_4_Atlantic_hurricanes
   https://en.wikipedia.org/wiki/List_of_Category_5_Atlantic_hurricanes

   The first link includes information about category 4 hurricanes happened from 1851 to 2018, while the second only collects category 5 hurricanes from 1921 to 2018. Since the two data sets should not be integrated for analysis as a result that the time span doesn't match, the first data set is utilized for analysis and the second for test.

2. Data processing:

   We copy hurricanes data from data resources websites into Excel. There are 120 records of

ID	YearBegin	YearEnd		Number	number Per Year
1	1851	1860	120	2	0.2
2	1861	1870	118	1	0.1
3	1871	1880	117	3	0.3
4	1881	1890	114	2	0.2
5	1891	1900	112	5	0.5
6	1901	1910	107	2	0.2
7	1911	1920	105	5	0.5
8	1921	1930	100	7	0.7
9	1931	1940	93	9	0.9
10	1941	1950	84	12	1.2
11	1951	1960	72	12	1.2
12	1961	1970	60	9	0.9
13	1971	1980	51	5	0.5
14	1981	1990	46	7	0.7
15	1991	2000	39	14	1.4
16	2001	2010	25	17	1.7
17	2011	2018	8	8	1

category 4 hurricanes and 32 records of category 5 hurricanes. As the time spans of two category hurricanes are different and the records of category 5 hurricanes are not enough, we mainly focus on the data of category 4 hurricanes. We count the times that hurricanes happened in every 10 years through excel formulas. The raw data is shown in the above form.

III. Distribution

Given the fact that hurricanes are rare, analysis has to be based on a regular time span, which is assumed as 10 years in this case. From the data set, we can have an array of times that hurricanes happened in every 10 years from 1851 to 2018 of category 4 and 5 independently. Set the times of hurricanes appearance in every 10 years as a variable, we apply these samples to the three distributions.

For the very beginning, we drew the histogram of numbers of hurricanes in category 4 per ten years as fig 3.1

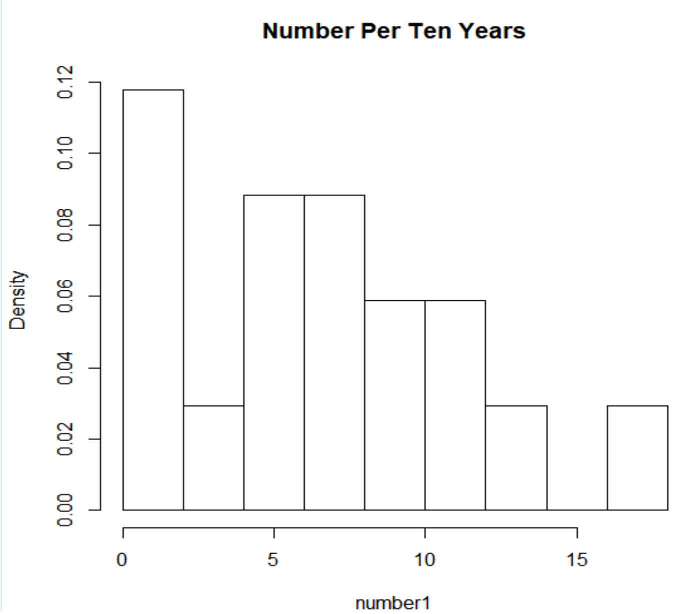

Fig 3.1 histogram of numbers of hurricanes in category 4

Consequently, the density curve was obtained as shown in fig 3.2

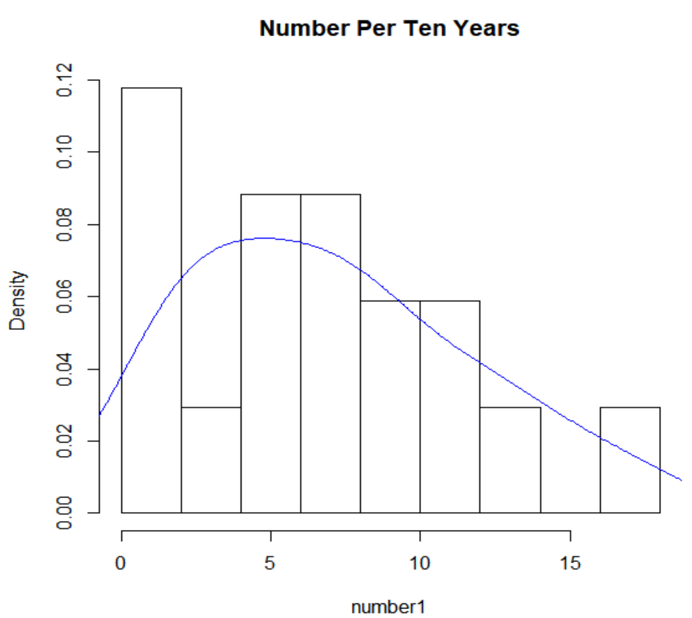

Fig 3.2 density curve

Based on the shape of density curve, we presented several conjectures of the distribution as follows: normal distribution, Poisson distribution and exponential distribution, the best approach to verify should be implement each possible distribution and compare with the original data.

We introduce function Quantile-Quantile(QQ) plot in R, which compares the fitted and empirical distributions in terms of the dimensional values of the variable by scatter plot. X-axis indicates theoretical quantiles while y-axis indicates empirical quantiles. Apparently, if all scatter points fit the reference line: y=x, the fitting distribution should be the best result, and the greater the departure is from the reference line, the more likely that the original data set should come from a population with a different distribution.

With the knowledge above, we applied QQ plot and added the reference line: y=x as shown in fig 3.3-fig 3.8, each two graph for one possible distribution.

3.1 Normal distribution

We call *rnorm()* function to generate random samples for the normal distribution with mean equal to hurricanes' mean and standard deviation equal to hurricanes' mean.

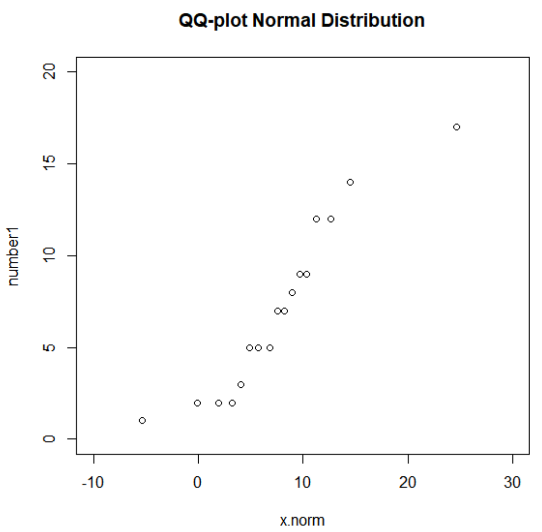

Fig 3.3

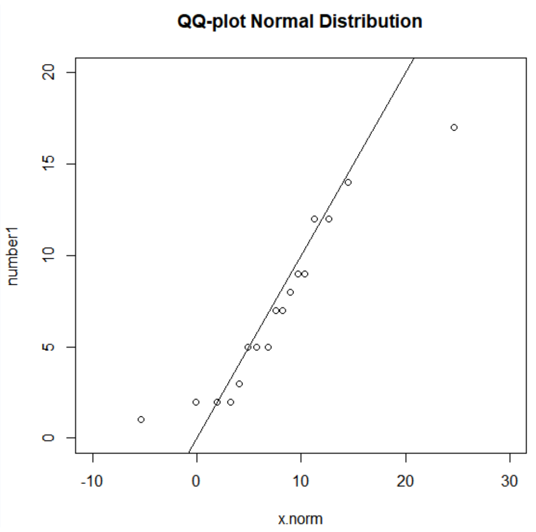

Fig 3.4

As fig 3.4 shows, most of the empirical quantiles have smaller values than theoretical quantiles.

3.2 Poisson distribution

In R, function *rpois()* is for random generation for the Poisson distribution with parameter lambda. We set lambda as the mean of hurricanes data.

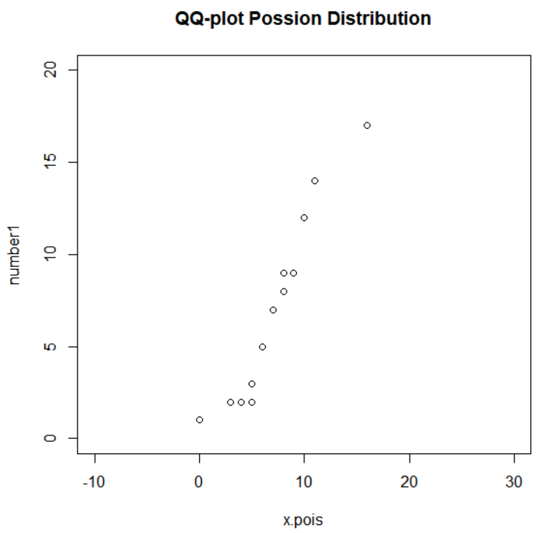

Fig 3.5

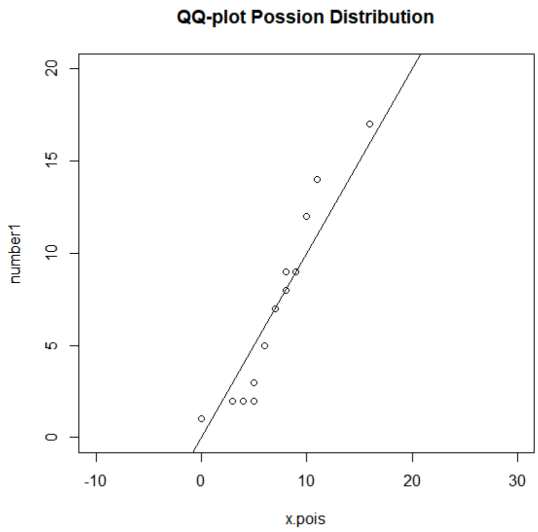

Fig 3.6

As fig 3.6 shows, the amount of points whose empirical quantiles are smaller than theoretical quantiles is almost the same as the amount of points whose empirical quantiles are greater than theoretical quantiles.

3.3 Exponential distribution

Function *rexp()* is called to generate random samples for the exponential distribution with rate 1/mean(hurricanes data).

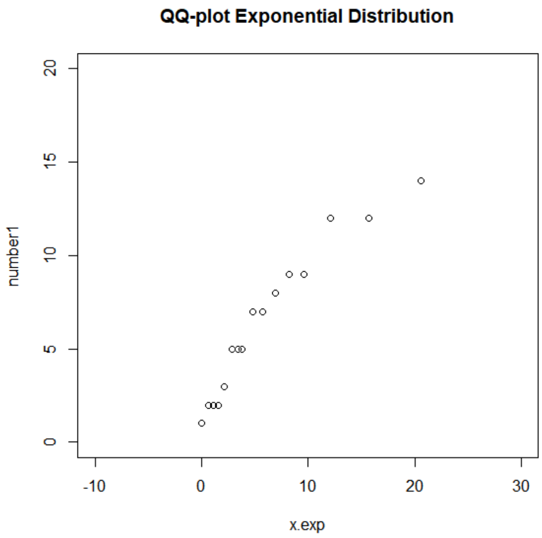

Fig 3.7

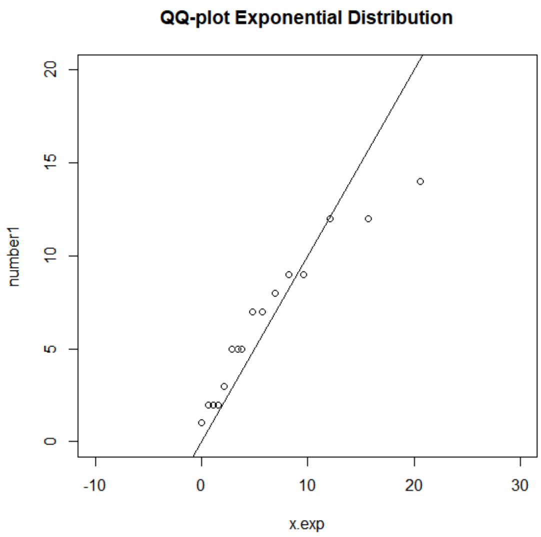

Fig 3.8

Fig 3.8 indicates that most of empirical quantiles are greater than theoretical quantiles.
From fig 3.3-fig 3.8, it seems the samples fit Poisson Distribution better than Normal Distribution and Exponential Distribution. Then we apply category 5 hurricanes samples to Poisson Distribution, Fig3.9. The plot shows that category 5 hurricanes fit Poisson distribution as well.

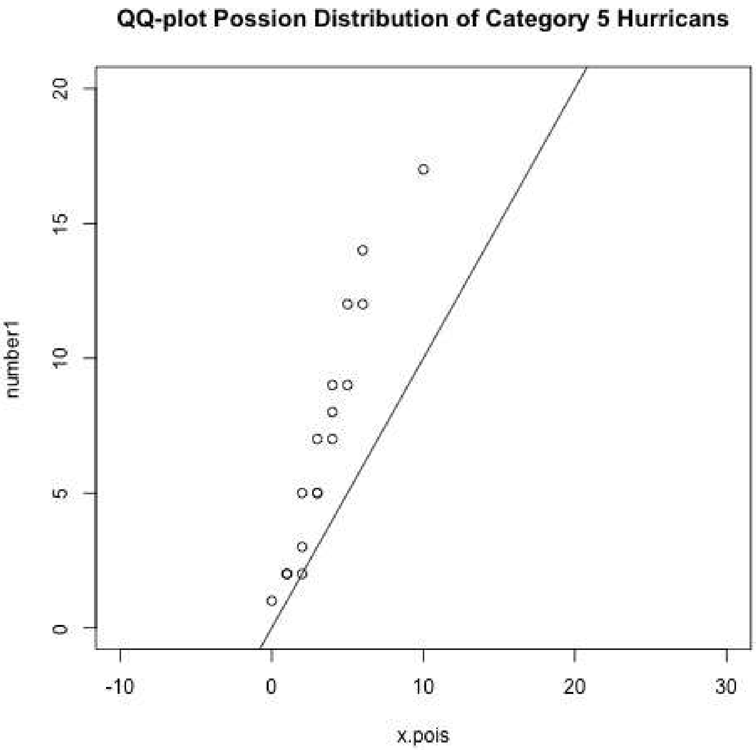

Fig 3.9

IV. Verification

To ensure the data indeed fits Poisson Distribution as it seems to be, means of test should be implemented.

1. Poisson process

To make the result more convincing, we would like to introduce Poisson process and expound that the occurrence of hurricanes is a Poisson process.

By definition, if a random process is Poisson process, it satisfies condition as follows:

(1)The process in ascending and independent, which means that process after time T is not related to the process happened.

(2)In a short span of time, the probability that the incident happens is proportional to the length of time span.

(3)In a very short time span, the probability that the incident happens more than once approaches to 0

To start with, the times that hurricanes happened will always be increasing, though there may be occasions that hurricanes happen mostly in rainy seasons, the history does not contribute to the formation of hurricanes, so condition 1 is satisfied. Given the fact that the occurrence of hurricanes is random and really rare, we can basically confirm that condition 2 and 3 are satisfied. Since the occurrence of hurricanes follows a Poisson process, the distribution along time should definitely be Poisson distribution.

2. Verifying the condition (1) of Poisson

With chronological order, we divide the data into two groups, which means there are 8 samples in each group. Applying QQ-Plot and Poisson distribution to two groups of data, we get Fig 4.1

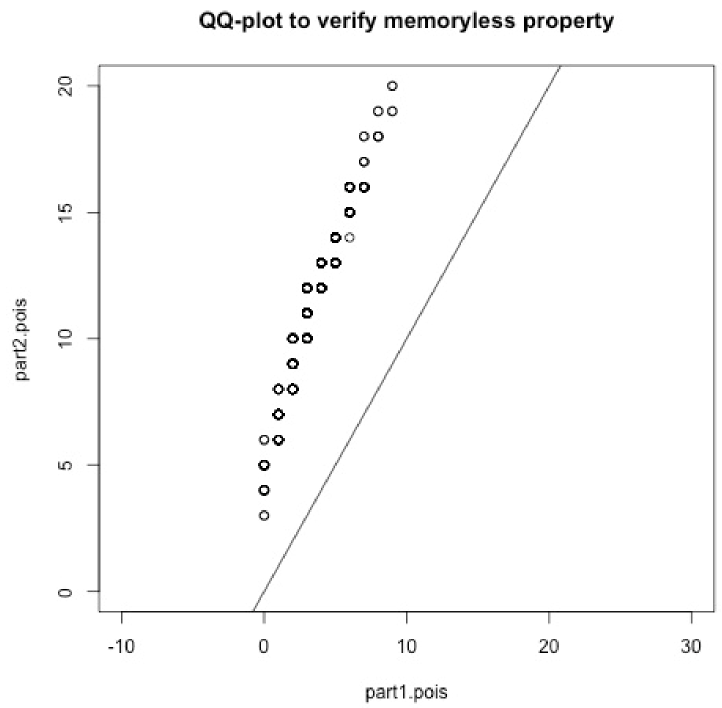

Fig 4.1

Obviously, Fig 4.1 fails to verify the condition (1) of Poisson. To find the deeper reason of the failure, we plot the numbers of category 4 and 5 hurricanes over time as Fig 4.2 and Fig 4.3. It shows that hurricanes are likely to get more frequent over time, especially after 1900.

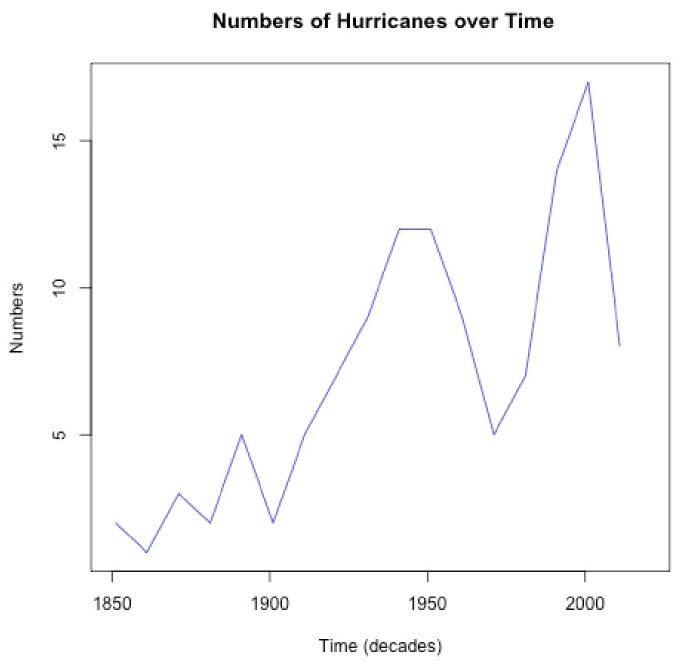

Fig 4.2

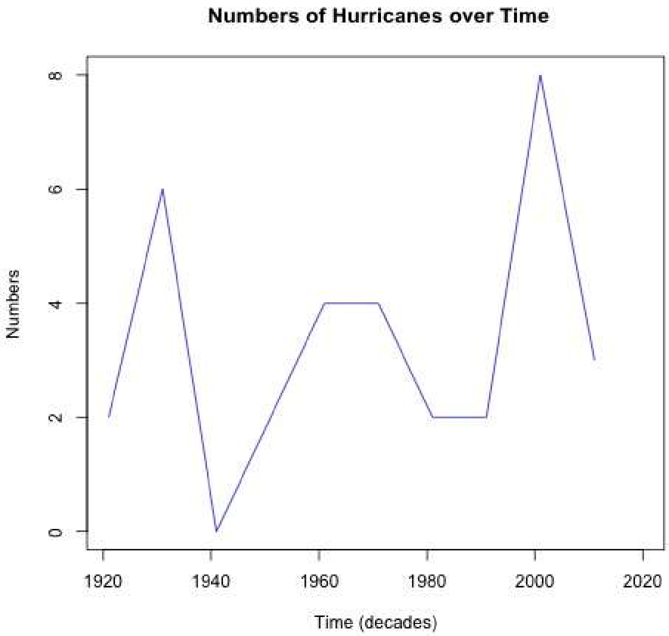

Fig 4.3

V. Conclusion

The Whole samples fit the Poisson distribution, but it fails to verify the properties of Poisson distribution. After 1900, hurricanes occurred much more frequently. Now we consider two factors, hurricane formation and data accuracy.

Since hurricane is a natural phenomenon, it requires four key ingredients in order to form: existing thunderstorms, warm water, moist air, and light winds. In the past one hundred years, the global warming exerted an increasing impact on world climate. The increase in named storms can be attributed to human-induced climate change. As a result, the world's oceans continue to warm at a fast rate, which means hurricanes are more likely. Hurricanes draw their energy from deep below the ocean's surface – up to depths of 2,000m. The temperature at these depths is measured by Ocean Heat Content, a metric that has soared since 1970, driven largely by four of the world's major oceans.

One thing that sets tropical cyclones apart from all the other storms on Earth is that they feed off of warm ocean waters. One theory states that a hurricane's strong winds help feed the storm heat from the ocean in a positive feedback loop. The heat helps make the storm stronger and, in turn, the stronger winds feed the hurricane even more heat from the warm ocean surface.

A similar process plays out at the ocean surface as a hurricane spins up. The WISHE theory states that strong winds generated by thunderstorms agitate the warm ocean surface and enhance evaporation. The evaporating water absorbs the sultry heat of the tropical ocean surface, and the water vapor travels up into the thunderstorms at the heart of the tropical cyclone.

This latent heat gets released into the atmosphere when the water vapor condenses back into water droplets, and it's this process that feeds the thunderstorms the warmth of the ocean. The newly-warmed air quickly rises through the storm, helping the thunderstorms grow stronger. Stronger thunderstorms suck more air up into the atmosphere, causing even lower air pressure at the surface. The lower pressure helps make the winds even stronger and the stronger winds cause even more evaporation. The process continues until you have a powerful, beefy hurricane; in a perfect environment, a hurricane's potential strength is only limited by the warmth of the water beneath the storm.

Both air and ocean warming may influence the formation of hurricanes, which is to say, if hurricanes occurrences fit Poisson distribution, the parameters of this assumed distribution are always changing due to some uncertain reasons.

Another factor is data accuracy. Before 1900, the technologies to detect hurricanes were deficient. The gathered data cannot precisely represent the actual situations. The Atlantic hurricane database extends back to 1851. However, because tropical storms and hurricanes spend much of their lifetime over the open ocean - some never hitting land - many systems were "missed" during the late 19th and early 20th Centuries. Starting in 1944, systematic aircraft reconnaissance was commenced for monitoring both tropical cyclones and disturbances that had the potential to develop into tropical storms and hurricanes.

With the improvement of technology, we can see hurricanes did not get more frequent after 1920 in both Fig 4.2 and Fig 4.3. For hurricanes striking the USA Atlantic and Gulf coasts, one can go back further in time with relatively reliable counts of systems because enough people have lived along coastlines since 1900. Thus, data accuracy may be the main reason that we fail to verify the properties of Poisson distribution.