Applied Stochastic Processes
Analysis of seismic activity in Japan

Arturo del Cerro

December 2020

Seismology is the scientific study of earthquakes and its properties. This field is more than 3000 years old and is known that about 90% of earthqueakes belong to the high-activity zone named Ring of Fire. Japan is in this region and have a high seismic activity each year, reaching a 9.1 magnitude earthquake in 2011. In this work we will study this stochastic variables referring to the frequency, magnitude, location, depth, correlacy and more of the earthquakes occurred in Japan in the last 20 years.

1 Data

The earthequakes catalog that have been used is given by the United States Geological Survey (USGS), available in [1]. This agency gives the option to download the seismic activity in a selected region during a selected period within a minimum and maximum magnitude values for each earthquake. In the case of this work, we have selected all the earthquakes from 01-01-2000 until 06-12-2020 (20 years of data) in the Japan and surroundings area with a magnitude grater or equal than 4.4. This agency gives the earthqueakes from a mininum magnitude of 2.7 but due to the vagueness and fluctuations of some seismographs and seismic activity meters, the number of earthquakes events with 2.7 to 4.3 magnitude is less than the number of earthquakes with magnitude 4.4, breaking the Gutemberg-Ritcher's Law about the frequency and magnitude earthquakes.
This USGS data offers information about date, time, latitude, longitude, depth, magnitude, number of seismic stations that report an earthqueake (*nst*), the largest azimuthal gap between azimuthally adjacent stations in degrees (*gap*), horizontal distance from the epicenter to the nearest station in degrees (*dmin*), horizontal location error, depth error, magnitude error and the number of seismic stations used to calculate the magnitude of an earthquake (*magnst*). A more detailed dictionary about the event terms in this dataset can be found in [2].

2 Does the seismic activity follows any distribution?

As it is known a Poisson process is an stochastic process that can be defined as a counting process, $\{N(t), t \geq 0\}$, that represents the number of events happened up to a time t. Some properties of a Poisson process are

- N(0) = 0
- Stationary and independent increments
- and $\mathbb{E}(N(t)) = \lambda$

being λ the mean of the Poisson distribution. So on, the probability of the random variable $N(t)$ being equal to k in a t interval is given by:

$$P(k\ events\ in\ interval\ t) = \frac{\lambda^k e^{\lambda}}{k!}$$

In this work we will study if this properties are satisfied by the seismic activity but as it is known the time event distribution in earthquakes does not verify the stationary increments conditions as we know that exists aftershocks when an earthquake occurs, so the memory-less of the process here is not being fullfilled. Nevertheless this seismic activity have other variables that may verify some other conditions.

Despite of this not shaping model for earthquakes, the poisson process have been used and is being used until today to measure the number of occurrences of earthquakes of magnitude $\geq M_{min}$. The reason of this is the simplicity and effectiveness of the model and the small number of parameters used. In the majority of the models that predict the seismic hazards appears this λ variable. This variable is full related with the Gutemberg-Ritcher's Law that cuantifies the relation between magnitude and frequency in a determined region. This law says that the number of earthquake events, $N(t)$ of a given magnitude M_{min} or greater in a time t is given by

$$\log_{10} N(t) = \alpha - \beta \cdot m$$

where α and β are constants related to the seismic nature of the region [3]. Then, for a given lower and upper bound magnitudes M_{min} and M_{max} respectively, the frequency of occurrences is

$$\lambda(t) = 10^{\alpha - \beta m} \text{ for } M_{min} \leq m \leq M_{max}$$

This lead us to the probability density function of magnitudes, $f_M(m)$, a truncated exponential given by

$$f_M(m) = \frac{\beta 10^{-\beta m}}{10^{-\beta M_{min}} - 10^{-\beta M_{max}}}$$

This lower bound appears because, as we said previously, the earthquakes from 4.4 to 0 in the dataset are undermeasured as they dont produce structural damages or they become unnoticed for seismographs. And this log in base 10 appears due to releasing energy.

3 Analysis

To begin with our work one could first see where are the earthquakes located. This is shown in Figure 1 where each dot represent an event. This spatial distribution doesn't seems to be uniform and seems

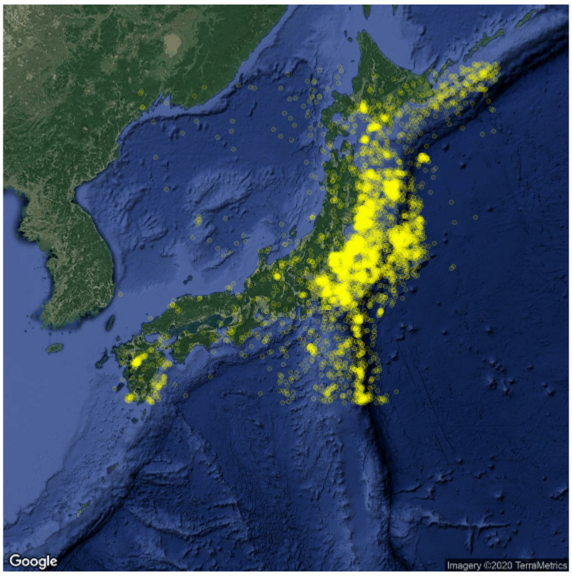


Figure 1: Earthquakes location

more to obey some correlation: in some locations occurs more events than in others. This is same as we talked previously about the Ring of Fire. In Japan adjoin three main plates: the Asian, the Pacific and the Philippine ones. That more shaded strips where the earthquakes distributes along are these plates. To continue, we see the first 80 events of our data showing how the magnitude of each earthquake fluctuates in Figure 2. From this small sample it seems that both magnitude values and time distribution have a dependence with the previous events. This makes sense: as is known, an earthquake is produced by a strong energy release due to the plates colliding. This collisions and movements generates more collisions and we end with global time windows with high and low seismic activity. This could be fixed

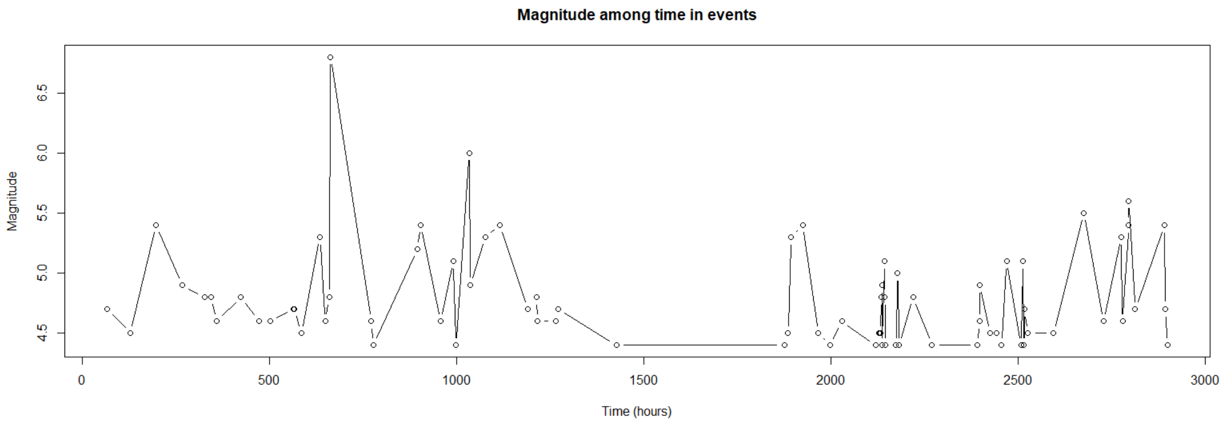


Figure 2: First 80 events

by removing the aftershocks of each event, a thing that we will study later on.

Here is where the theory about earthquake events we talked about before appears. We see in Figure 3 how this Gutemberg-Ritcher Law is satisfied.

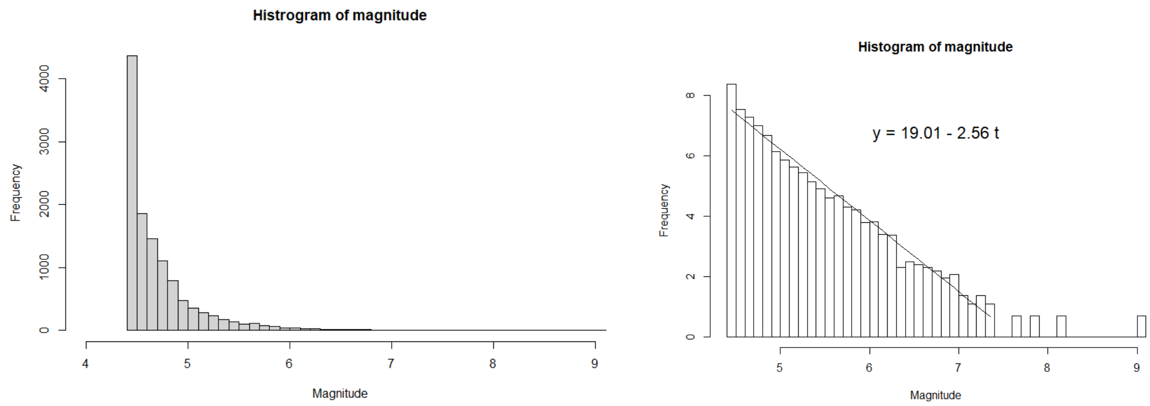


Figure 3: Magnitude histogram

In the right plot is considered the events with magnitude 7.3 and above as outliers in order to fit the slope. Also it need to be said that this enormous magnitudes doesn't fit with the probabily density function because we have to set a maximum magnitude. This is a problem for the Ritcher's magnitude scale and there are other models that takes in consider this.

3.1 Aftershock events

In all this way through the work we have been talking about the clear dependence between events and giving some idea of how we could make our data more randomly stochastic. In [4] can be found a similar analysis to the one that will be done here. In this paper is defined an event caused by an aftershock measured by distance and magnitud from the previous earthquake, which makes sense. By creating a window algorithm that drops the aftershocks, from an initial dataset with 11780 events we end with 1562 events not caused directly from other earthquake. This means that a 86% of our events was highly related to other events. Now we can understand more results given. To get some idea of how this data without aftershocks is we see a summary of characteristics of the time between events in both datasets:

```
Time between events with aftershocks
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
4.40    4.50    4.60    4.74    4.90    9.10
Time between events without aftershocks
Min. 1st Qu.  Median    Mean  3rd Qu.   Max.
4.40   4.60   4.80      4.99   5.20     9.10
```

In an attempt to show more of this difference we present in Figure 4 the autocorrelation function values for magnitude of the earthquakes with and without aftershocks.

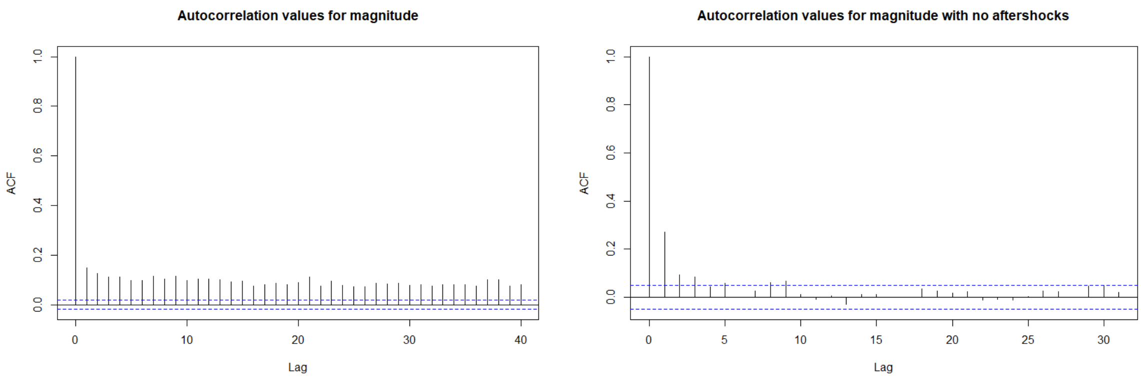


Figure 4: Autocorrelation function values

Clearly exists a difference between this magnitude variable when we force it to be enough independent: the full majority of values are below the threshold value in the dataset with no aftershocks, unlike the dataset with aftershocks.
Once we have studied this difference for the magnitude we could do the same for the time between events in order to see if happens the same or not. In Figure 5 is shown.

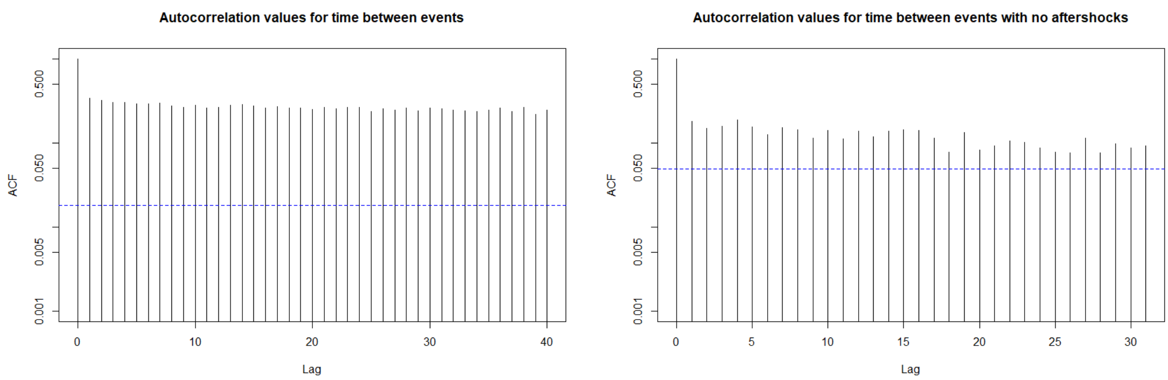


Figure 5: Autocorrelation function values

This time the result is not as pleasant as we would like it to be. As it can be shown both datasets with and without aftershocks present a significant autocorrelation values for the time between events. However, the second one, as it occurred before, seems to have a set of conditions more reliable to not to be autocorrelated as both the threshold and the autocorrelation function values are closer each other than in the first plot.
Now is time to see the periodogram for magnitude and time between events in order see an estimation of the spectral density. In Figure 6 we see that both for time and magnitude there is no frequency that rules the spectrum values and that it is flat. This is a great appointment in order to treat our data as stochastic.

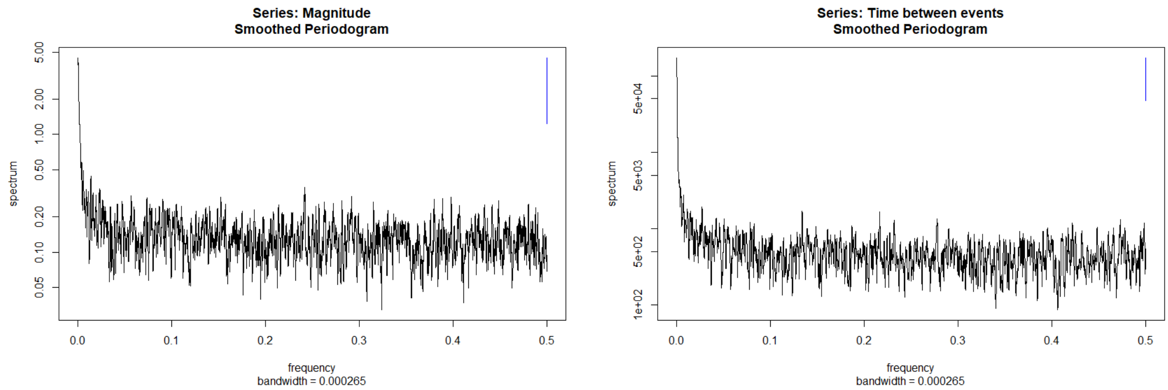


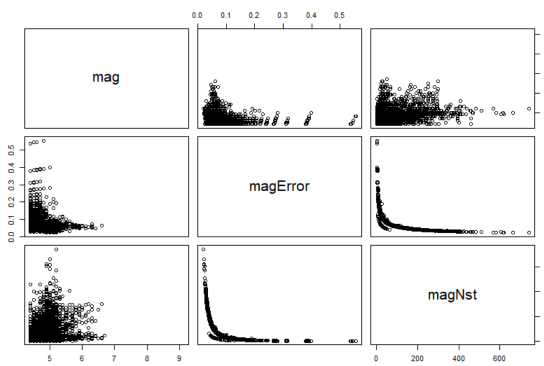
Figure 6: Periodogram

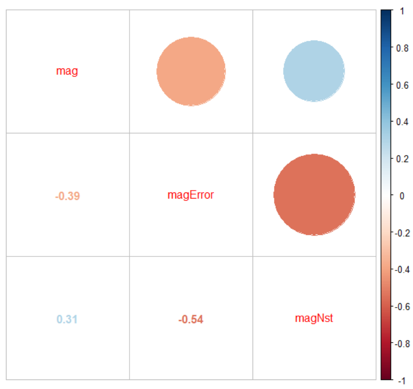
3.2 Other measures analysis

Up to this point we have only analized the variables magnitud and time between events in depth. However we have more other variables to study. We divide them into 2 clusters: depth and magnitude measures. In this section we will comment how this measurements relate among themselves. It is a must to say that the distance measures are not presented here as a result because they do not give significant conclussions about relations between themselves.

3.2.1 Magnitude measures

Our dataset provide us of the magnitude error values. We can check how it works depending on other values.

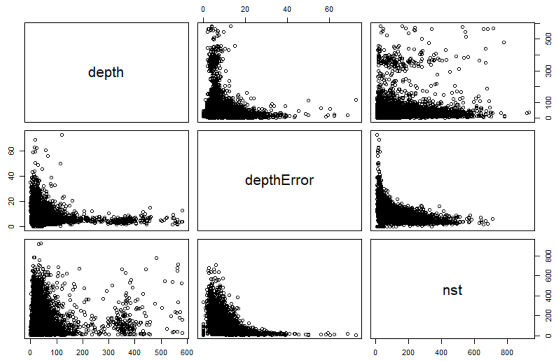




By this plots we can get some conclussions about how earthquakes occurs and how we measure them. First thing we can point out is the strong negative correlation between magError and the number of stations to measure the magnitude (*magNst*). It makes sense: the more stations to measure the less error obtained. Other guess we can do about how the measurement is made is that there are more errors when the magnitude of earthquakes is lower. This is a thing we commented before, as in the dataset there were fewer events with magnitude 3 than with magnitude 4, breaking the Gutemberg-Ritcher Law. So it means that higher magnitude earthquakes are more easy to detect, like it is logic.

3.2.2 Depth measures

As we know earthquakes can be produced in different depths of the terrestrial layer, giving out different wave forms and releasing more or less energy. As we have this depth data we can try to understand it and how it is measured. First thing to say is that altohugh it is not represented in the plots, depth and magnitude are not related variables. Again, as the previous paragraph, we have that the error in the depth measurement is negative correlated with the number of stations to measure it. But the key here is to realize that only near the surface the measurement errors are made. This can be explained in the way that in a more depth earthquake the different types of waves give more information as they travel



in the inner layers of the cortex.

References

[1] United States Geological Survey *(USGS)*. `https://earthquake.usgs.gov/earthquakes/search/`

[2] USGS terms dictionary. Event terms `https://earthquake.usgs.gov/data/comcat/data-eventterms.php#nst`

[3] T. Anagnos, A.S. Kiremidjian. *A review of earthquake occurrence models for seismic hazard analysis.* Probabilistic Engineering Mechanics, 1988, Vol. 3, No. 1

[4] J. K. GARDNER, L. KNOPOFF. *Is the sequence of earthquakes in souther california, with aftershocks removed, poissonian?* Bulletin of the Seismological Society of America, Vol. 64 October 1974 No. 5