Homework 4

Seismicity of Northern California

Purpose: Learn how we make predictions about seismicity from the relationship between the number of earthquakes that occur and their magnitudes.

We know that small earthquakes are common, while big ones are rare. But the large ones are the ones that affect our cities and our society! We don't want to wait for lots of big ones to happen to start to look for patterns. Can we look at patterns in smaller quakes to learn about the big ones…?

Background: The Number of Earthquakes

We know that there are more small earthquakes than large ones. If we can identify a pattern, and if we can express the pattern quantitatively, we can use the pattern to tell us what we might expect in terms of the sizes and numbers of earthquakes to come.

Let us consider one particular region of the world—Northern California, as is the focus of this class—and one particular time interval. We can use the magnitude **M** as a measure of the size of an earthquake and count them. For some particular magnitude **M**, we will find there are ***n*** earthquakes. For a different magnitude, we will find a different value of ***n***. The value ***n*** will be bigger for small earthquakes than for large ones. We say: "the number of earthquakes is a function of magnitude", or "***n*** is a function of **M."** We can write,

n(**M**) = *number of earthquakes with magnitude* **M**

As you might expect, the exact value of ***n*** for any particular magnitude depends on the length of time for which the count is made. To avoid this, we usually normalize ***n*** by the number of years for which we have counted (like a batting average). This gives us a new definition of ***n***,

n(**M**) = *number of earthquakes with magnitude* **M** *per year*

This measure of the number of earthquakes still has some problems. For example, this function requires two different magnitude values: the value of ***n*** for earthquakes between magnitude 3.0 and 3.5 is different from the value of ***n*** for earthquakes between magnitude 3.0 and 3.25. We can see, then, that this measure depends upon how accurately the magnitudes are measured and how precisely they are tabulated. A better way to express the number of earthquakes is to use the cumulative number of earthquakes above a give magnitude,

N(**M**) = *number of earthquakes per year with magnitude greater than or equal to* **M**

This measure of the number of earthquakes is much less sensitive to the way the data are saved. This function only requires a single magnitude to be Note that the number **N**(**M**) is also normalized by the length of time used and is given as the number of earthquakes per year.

For many areas in the world, scientists have graphed **N**(**M**) and have discovered that there is usually a simple relationship between **N**(**M**) and **M**. Specifically, when $\log_{10}$[**N**(**M**)] is plotted versus **M**, the data fall nearly on a straight line. Recall that if $10^x = y$, then $\log_{10}(y) = x$. So remember that the graph for a straight line is

$$y(x) = \boldsymbol{A} + \boldsymbol{b}x,$$

where x and y are variables (the axes), ***b*** is the slope of the line and ***A*** is the intercept of the line with the y-axis. So if we plot $\log_{10}$[**N**(**M**)] versus **M** and get a straight line, it suggests that we can write

$$\log_{10}[\mathbf{N}(\mathbf{M})] = \boldsymbol{A} - \boldsymbol{b}\ \mathbf{M}$$

where ***A*** and ***b*** are constants (– ***b*** means the line slopes downward to the right if ***b*** by itself is a positive number). Here ***A*** is the intercept of the earthquake occurrence curve at magnitude zero (remember, this does not mean "no earthquake", magnitude zero is just a small earthquake… earthquakes can even have negative magnitudes!). The value ***b*** is the negative slope of the line. It describes how many more small earthquakes there are for a given number of large earthquakes.
Using this formula it is very easy to characterize the seismicity of a region. The steps are:

(1) define a region,
(2) pick a time interval,
(3) count the earthquakes ***n*** for each magnitude **M** in the catalog,
(4) calculate **N**(**M**)
(5) plot $\log_{10}$[**N**(**M**)] versus **M,**

(6) fit a line to the plot and estimate the parameters ***A*** and ***b***.

By this process we can summarize the seismicity of a region using just two numbers. The number ***A*** gives an estimate of the general level of seismicity for the region: how many earthquakes (greater than magnitude 0) we can expect in the region during the course of the year. The number ***b*** tells us about what magnitudes we can expect. How many of those earthquakes will be big, or how often will big (or really big) earthquakes occur? For most regions, we have found that the number ***b*** is very close to 1.0. That means that if there are 100 earthquakes with **M** > 5, on average, there will be 10 with **M** > 6.

We must be cautious applying this method to characterize the number of earthquakes in a region.

1. At very large magnitudes you may only have one (or zero) events and the $\log_{10}$ of that is 0 (or not defined… one cannot choose a value for x such that $10^x = 0$). Including the counts (***n***(**M**)) of these magnitudes may cause the analysis to break down and estimation of a good line becomes difficult.
2. At very low magnitudes we may not be able to detect all the events and the counts (***n***(**M**)) of these magnitudes become unreliable. At this point the analysis breaks down and estimation of a good line becomes difficult.

(Note: This deficiency in the method can actually be used to extract additional information. If we assume that the linear relationship between $\log_{10}$[**N**(**M**)] and **M** is true, then the point at low magnitudes where the tabulated data begins to deviate from the straight line can be used to estimate the lowest magnitude for which we detect **all** of earthquakes.)

Procedure:

A. Earthquakes in Northern California

Northern California is a convenient region for checking the relationship between the number of earthquakes and their magnitudes described above. The earthquake catalogs for this region are available on the web page of the Northern California Earthquake Data Center (NCEDC, http://www.ncedc.org), which is hosted here at Berkeley. This catalog lists all earthquakes that have been recorded and located in the region with their magnitudes. In the NCEDC catalog, the magnitudes for local and regional earthquakes are given as M_L, the local magnitude. For

this lab, you will calculate the cumulative number **N**(**M**) using the magnitudes listed in the NCEDC catalog.
For this exercise you will estimate the statistics of earthquake occurrence in Northern California by using the data tabulated in NCEDC for one year. When you select a year to analyze, you should choose a year when there was no earthquake with many aftershocks. If the year is dominated by an aftershock sequence, then you will learn more about the statistics of this particular aftershock sequence than about the statistics of northern California, so avoid Oct 1989 to Oct 1990 – the Loma Prieta Earthquake. Also avoid April 1992 to April 1993 – a M7.2 near Eureka, followed that day by 3 greater than M6! Finally, it's probably a good idea to avoid anything more recent than Napa's August 2014 quake.

A1. Select one year available in the catalog (1984 – present, it does not have to be a calendar year – it could be from your birthday, 1995, to your birthday, 1996) and use the region from 124.5 W to 120.0 W (Note: W longitude is negative longitude) and 36.0 N to 42.0 N. Count the number of earthquakes in the magnitude ranges 2.0-2.49, 2.5-2.99, 3.0-3.49, 3.5-3.99, 4.0-4.49, etc, up to at least 7.0. For one of your magnitude ranges, print the first page (**only the first page!**) of output from the NCEDC for the request and submit it with the assignment.

NOTE: You should only print the first page. However, if you have multiple pages of earthquakes in your range, you still **must use all the data** on those pages! I just don't want 50 printed pages for each student sitting on my desk. I ask for this front page because I want to be able to see if you've made a mistake in the data request.

Use the NCSN catalog on the NCEDC web page: **http://www.ncedc.org/eps20/** You will need to tell the search engine the time period and the area you are interested in investigating.

Read the instructions, and select the following choices:

Input dataset: NCSN catalog (1967-present) Give the start and end dates for one year (follow the example format)
Min magnitude: 2.0
Max magnitude: 2.49
Min latitude: 36
Max latitude: 42
Min longitude: -124.5

Max longitude: -120

Use only earthquakes with reported magnitudes.
If you have problems figuring out what to enter, look at the instructions on the web site.

When you have done this for M2-2.49, copy and paste the output in a spreadsheet. That way you can quickly count how many earthquakes there were in that range! Record that number, and repeat this for the other magnitude ranges.

A2. Make a table with the magnitude range, the number of earthquakes in each range, ***n***(**M**). For each magnitude range, calculate the cumulative number of earthquakes, **N**(**M**) and include it in the table. So: a column for **M**, a column for ***n***(**M**), and a column for **N**(**M**). You may also find it useful to make a column for $\log_{10}$[**N**(**M**)]…

A3. Make a histogram of the values of ***n***(**M**) listed in the table.

A4. Plot $\log_{10}$[**N**(**M**)] versus **M**, either by hand or using a spreadsheet's graphing programs. Fit the points with a straight line and determine the parameters ***A*** and ***b***.

A5. How and why did you pick the section of the data on the graph that you used to decide where the straight line belonged? Did you use every magnitude range?

B. Recurrence Intervals

The parameters ***A*** and ***b*** you calculated in section **A** provide a useful characterization of the seismicity of a region. We can use them to estimate the rate at which future earthquakes will occur in the region we studied. One common way to express this rate is the **recurrence interval** for earthquakes of a given magnitude. We define this as

RI(**M**) = *time between earthquakes with magnitudes greater than or equal to* **M**

This is the same as the inverse of **N**(**M**):

RI(**M**) = 1 / **N**(**M**)

(remember, **N** is normalized to "the number of earthquakes per year" or earthquakes/year. So when we take the inverse, we get years/earthquake. Keep track of your units in this part… you could get lost if you don't!). If we know the parameters ***A*** and ***b*** for a region, we can calculate the value of **N**(**M**) for any magnitude **M**, and calculate 1 / **N**(**M**) to estimate the recurrence interval.

We can use this procedure to estimate recurrence intervals. For example:

1) We can extrapolate to small magnitudes where detection is incomplete and estimate how often earthquakes of that size may be occurring.

2) We can use it to extrapolate to large magnitudes, for events which seldom happen and estimate approximately how long we might have to wait for the next earthquake of that size or larger. Note that when using the linear relationship between $\log_{10}$[**N**(**M**)] and **M** to extrapolate in magnitude, we are making an important assumption: that the relationship is valid over the entire magnitude range.

B1. Using the values for ***A*** and ***b*** you calculated in part **A**, find the average recurrence intervals (**RI**(**M**)) for earthquakes of magnitude 2, 5, and 8 in northern California. (remember, the parameters ***A*** and ***b*** give $\log_{10}$[**N**(**M**)] and **RI** requires **N**(**M**)) If you like, you can check your **RI** against an earthquake you have lots of, such as M3.5. That way you can check if the numbers seem reasonable.

B2. Use the recurrence intervals you calculated in B1. How many earthquakes with magnitude 2 or greater do you expect in Northern California in one day? If the last magnitude 5 earthquake in Northern California occurred on May 23, 2013, when do you expect the next one? How about the next magnitude 8 earthquake, given the fact that the last one was in 1906?

Note: wolframalpha.com is a great tool to figure out hours or days or years after some event… and a number of other cool things.
(a digression… for example, you can just type in (area of Alaska)/(area of contiguous US) and it will give you the ratio of their areas! Alaska is a fifth the size of all other states combined…! Wow. Or type in "mars" and it will give you a map of where it is on the sky! If it's not below the horizon.)

B3. Think about how your analysis was performed. How can you get more reliable estimates of recurrence intervals? Name at least two ways!