Problem Set 1:
The Gutenberg–Richter frequency–magnitude relation [gr]

GEOS 626/426: Applied Seismology, Carl Tape

Assigned: January 14, 2021 — Due: January 21, 2021

Last compiled: November 5, 2021

Overview and instructions

- The purposes of this problem set are:
  1. to practice using Python for scientific computing and plotting
  2. to think critically about histograms
  3. to review and apply some concepts of differentiation
  4. to understand the frequency–magnitude relation (Eq. 1)[1], including some of its subtler points

  If you are new to Python, let me know.

- Suggested reading:
  - *Stein and Wysession* (2003, Section 4.7.1)
  - *Gutenberg and Richter* (1944)

- Log into OpenSARlab and run the template notebook `hw_gr.ipynb`. You can work directly in this file.

- **Possible source of confusion:** Following standard mathematical notation, I will use '$\log_{10}$' to represent the base-10 logarithm and 'ln' to represent the natural logarithm. Note that in Python `log10` is $\log_{10}$ and `log` is ln. (Note that *Stein and Wysession* (2003) use 'log' for the base-10 logarithm.) Reminder: $\log_{10} x = \ln x / \ln 10$

- **Python tips:** `hw_gr.ipynb` shows an example of how to plot multiple items on a log-scaled plot. Make sure you follow the template script before proceeding.

- **Written responses:** Complete as many answers as possible from inside the Jupyter notebook. Other responses, such as equation derivations (if needed), can be done with pencil and paper, then scanned and uploaded to the google drive.

- **Latex option:** If you want to use Latex for your homeworks, see the templates in the folder `latex` (and see the `README`).

[1] This empirical formula is referred to as a + b, where a = "frequency–magnitude" or "Gutenberg–Richter" and b = "relation" or "relationship" or "distribution" or "law".

The Gutenberg–Richter frequency–magnitude relation

The Gutenberg–Richter frequency–magnitude relation (*Gutenberg and Richter*, 1944) is one of the cornerstone empirical relationships in seismology. It is given by

$$\log_{10} N = a - b\,M, \tag{1}$$

where

- N is the cumulative number of earthquakes having magnitudes larger than M that occur in region R within a particular time T
- M is the earthquake magnitude; we take this to be moment magnitude $M_{\rm w}$
- b controls the slope of the seismicity distribution in region R within a particular time T. The line will always increase to the left, so Equation (1) has a negative slope and therefore $b > 0$ by definition.
- $b \approx 1$ for most earthquake catalogs
- a indicates the seismic activity in region R within a particular time T

A subtle point is that in practice the magnitudes are *binned*, then a line is fit to the histogram. It is helpful to consider the discrete form of Equation (1):

$$\log_{10} N_i = a - b\,M_i, \tag{2}$$

where the index i refers to the magnitude bin, with M_i being the magnitude at the left boundary of the bin.

- The *cumulative distribution* is the cumulative number of events with $M \geq M_i$; the frequency–magnitude relation is expressed as a cumulative distribution.
- The *incremental distribution* is the number of events per magnitude bin, $M_i \leq M \leq M_{i+1}$.
- A *magnitude interval* is a range of magnitude, e.g., the magnitude interval $[8.7, 9.0]$. For the cumulative distribution, a magnitude interval should be expressed as an open interval, such as $M \geq 6.2$, rather than $6.2 \leq M \leq 6.3$.

Problem 1 (10.0). Histograms and earthquake statistics

Make sure that you are very comfortable with the meaning of Equation (1) before proceeding.

1. (0.2) We consider the Global Centroid Moment Tensor (GCMT) catalog (`www.globalcmt.org`), from 01-Jan-1976 to 30-June-2011. Thus T represents the duration of the catalog, and R represents planet Earth.

   Run the Python notebook `hw_gr.ipynb` to generate a global map of the catalog.

   (a) (0.0) What is the range of depths of events in the catalog?

   (b) (0.1) List three regions of the deepest seismicity (e.g., Tonga–Kermadec).

   (c) (0.1) What is the range of magnitudes of events in the catalog?

2. (0.8) Running `hw_gr.ipynb` will call the function `seis2GR.py` to obtain the cumulative and incremental distributions for the M_w values of the GCMT catalog, using a bin width of $\Delta M = 0.1$. Examine the output that appears.

   Note: The function `seis2GR.py` uses the variables names `Ncum` for the cumulative numbers N_i, `N` for the incremental numbers, and `Medges` for the bin edges M_i.

   (a) (0.2) What is the maximum value of the incremental distribution?
   What is its magnitude interval?

   (b) (0.2) What is the maximum value of the cumulative distribution?
   What is its magnitude interval?

   (c) (0.2) What is the minimum value of the incremental distribution over the interval $[4.2, 9.1]$?
   What is its magnitude interval?

   (d) (0.2) What is the minimum value of the cumulative distribution over the interval $[4.2, 9.1]$?
   What is its magnitude interval?

3. (3.0) Now it's time to write some lines of code.

   (a) (0.0) Using the output from `seis2GR.py` (as shown in `hw_gr.ipynb`), **plot the cumulative and incremental distributions on the same plot**, similar to the plot in *Stein and Wysession* (2003, Figure 4.7-2), but note that your x-axis is M_w, not $\log_{10} M_0$, and your y-axis is number of earthquakes, not number of earthquakes per year.

   Coding tip: There are two ways to deal with log scaling. One way is to transform N into $n = \log_{10} N$, then work with n and use the `plot` command. Another way is to work with N directly and use the `semilogy` command, as shown in the example at the bottom of `hw_gr.ipynb`.

   Which distribution—incremental or cumulative—has more scatter?

(b) (2.2)

i. (0.6) Find a best-fitting line, $\log_{10} N = a - bM$, for the 'most linear' section of the $\log_{10}$-scaled cumulative distribution. You can simply pick two points and compute the line, or use a command such as `polyfit` (and `polyval`) to apply a least-squares fit to a set of points. (Do not fit a line to the entire distribution!) What are your values for a and b? (Show your work.)

ii. (1.0) Plot your best-fitting line along with the full data set. Include both the incremental and cumulative distributions in your plot.

iii. (0.3) a is the y-intercept. What is its physical meaning?

iv. (0.3) b is the slope. What is its physical meaning?

(c) (0.6) Assume that the best-fitting distribution (not the GCMT catalog) is 'reality'. Based on the idealized cumulative distribution, what is largest earthquake expected over the duration of the GCMT catalog? (Show your work.)

Hint: Where does $N = 1$ intersect your best-fitting line?

Note: The expected value does not have to agree with what actually occurred within the GCMT catalog.

(d) (0.2) The "catalog completeness" (e.g., *Wiemer and Wyss*, 2000) M_c represents the smallest magnitude above which the frequency–magnitude distribution is representative for a particular seismicity catalog. What is the catalog completeness for GCMT? List your answer with 0.1 precision. (Provide a brief explanation, but no computation is necessary.)

4. (1.0) Instead of analyzing seismicity, let us now analyze seismicity rate by dividing all binned values by the duration of the catalog (T, in years).

(a) (0.1) Taking an average over the entire time interval of the GCMT catalog, how many earthquakes per year are there?

(b) (0.1) Why is seismicity rate more useful than seismicity?

(c) (0.5) What is the best-fitting line (namely, a and b) for the new distribution? You can determine this graphically or analytically, from your previous results. (Show your work.)

(d) (0.1) What magnitude interval averages >100 events per year?

(e) (0.1) How many $M \geq 0$ earthquakes are expected on Earth per year?

(f) (0.1) How many earthquakes (of any magnitude) are expected on Earth per year?

5. (1.0)

(a) (0.8) Plot the cumulative and incremental distributions for seismicity rate for bin widths of $\Delta M = 0.05, 0.10, 0.5,$ and 1.0.

(b) (0.2) What is the apparent relationship between bin width and the separation between the cumulative and incremental distributions?

6. (2.0) (*A computer is not needed for this problem.*)
   You will now try to show mathematically what you identified graphically in 5b.
   Define the bin width as

$$\Delta M = M_{i+1} - M_i \tag{3}$$

   where i increases to the right (the usual convention).

   The incremental distribution is given by

$$-\Delta N_i = -(N_{i+1} - N_i) = N_i - N_{i+1}. \tag{4}$$

   (a) (1.7) Using the discrete frequency–magnitude relation (Eq. 2), show that the incremental distribution can be written as

$$\log_{10}(-\Delta N_i) = a + \Delta a - bM_i \tag{5}$$

   where Δa is a function in terms of other variables (but not i).
   List the expression for Δa.

   (b) (0.1) What is the relationship between bin width and the shift in the y-intercept?

   (c) (0.2) If $b = 1$ and $\Delta M = 0.1$, what is Δa?

7. (1.5) **[GEOS 626]** (*A computer is not needed for this problem.*)

   (a) (0.3) Differentiate Equation (1) to obtain dN/dM.

   (b) (1.0) Using your expression from (a), as well as the mathematical definition of a derivative, derive an expression analagous to Equation (5) that is valid for small bin widths, $\Delta M \ll 1$. List the expression for Δa

   (c) (0.2) If $b = 1$ and $\Delta M = 0.1$, what is Δa?

8. (0.5) Earlier you determined the catalog completeness, M_c.

   (a) (0.1) Can the GCMT catalog, $M > M_c$ (see earlier part of this problem for M_c), be fit well with a single line?

   (b) (0.2) Compute an estimate b for $M > 7.5$.

   (c) (0.1) What does the different b value imply about large events in the catalog?

   (d) (0.1) What is a possible reason for this?

Problem

Approximately how much time *outside of class and lab time* did you spend on this problem set? Feel free to suggest improvements here.

References

Gutenberg, B., and C. F. Richter (1944), Frequency of earthquakes in California, *Bull. Seismol. Soc. Am.*, *34*(4), 185–188.

Stein, S., and M. Wysession (2003), *An Introduction to Seismology, Earthquakes, and Earth Structure*, Blackwell, Malden, Mass., USA.

Wiemer, S., and M. Wyss (2000), Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the western United States, and Japan, *Bull. Seismol. Soc. Am.*, *90*(4), 859–869.