Problem Set 2: Univariate Distributions

EAS 6134: Inverse Methods and Data Analysis in EAS

Assigned: 2/07/22
Due: 2/14/22

Name: _____

Others consulted: _____

Revision: 2022-02-06 23:35:52Z

- Please be neat and organized! Once you have found a way to the answer, please rewrite it in an orderly fashion so that others can follow your steps, and put a box around your final solution when appropriate.
- Include this page as the cover, listing all who helped with this set including me in the "Others consulted" line.
- Show all of your work.
- An answer with incorrect or absent units will be considered wrong.
- For electronic submission to Canvas, please create a single PDF of your written work. This can be done with most smartphones using tools such as *Google Drive*, *Adobe Scan*, or *Scanner App* (GoogleDrive tested, but any method acceptable).

Given the global earthquake catalog for events equal or greater than magnitude 6 between 1976 and the end of 2020 (from the Global CMT project), you will evaluate whether these data are well described by varying distributions.

Data: CMTS_geM6_1976-2020.txt

1. Read in the above data and convert the timing information into something more usable. It is generally beneficial to leave the original file in its current format and use your computer code to smartly parse the information. This can be a bit time-consuming, but can be extremely helpful if datasets get updated with some frequency (and similarly).

   When working with time-based data, it is generally most beneficial to convert that data into a usable format that minimizes/removes issues with wrapping around days/months/years etc. Fortunately, there are some very useful tools in modern programming languages to convert most time formats into epoch-time, a time format counting seconds from January 1, 1970. Submit the first few lines of your codes solutions for epoch times.

   In python with pandas you can make quick work of this (along with parsing and converting data params into integers):

```
# read in CMT data
import pandas as pd
import datetime
import numpy as np
CMTS=pd.read_csv('CMTS_geM6_1976-2020.txt', sep='\s+', comment="#")
```

```
for index, EQ in CMTS.iterrows():
    #print(index)
    # split date and convert all to integers
    year,mo,day= [int(x) for x in EQ.DATE.split('/')]
    # a little trickier for time of day since seconds are floats
    hr,mn,sec= EQ.TIME.split(':')
    hr=int(hr)
    mn=int(mn)
    sec=int(float(sec))  # float then int
    # there exists 1 occurrence of sec= 60sec (breaks the below)
    if sec >= 60:
        print(sec)
        sec=sec-60
        mn=mn+1
    eepoch=datetime.datetime(year,mo,day,hr,mn,sec,
                tzinfo=datetime.timezone.utc
                ).strftime('%s')
    CMTS.loc[index,'EPOCH'] = eepoch
```

2. Plot the temporal occurrence of this global dataset in a meaningful way. This can be tricky, as data frequently can be written in a wide range of formats, and even epoch time is not always the most meaningful, and something more like decimal year, may be better. Be sure to appropriately label all axes and give either a useful header, or a caption that describes what we're evaluating.

3. Plot the data to evaluate the functionality of a Pareto Power-law Distribution for these data ($\Pr[X \geq x]$). For this, you will need to establish a counting mechanism for events greater than or equal to a given magnitude. It's likely most useful to evaluate every 0.1 unit of magnitude ≥ 6.0, then ≥ 6.1, etc... To understand the frequency of occurrence, it's best to normalize the data the time window over which the data were collected. Plot should be log-scale in Y. Comment on any deviations you see from an expected pure power-law distribution.

4. Similarly, evaluate the functionality of these data for being described by a purely random, time-independent Poisson behavior. Try this for all data, and then for events only greater than magnitude 8.

5. Select another probability distribution of your choice that you may think would be useful for evaluating the entire dataset.

   (a) Apply it and present your results.

   (b) At what magnitude would we expect a 50% annual occurrence?