SIO 223A, Geophysical Data Analysis
Problem Set 2
due 01/30/2020

This problem set asks, as others have and will, for programs and plots. For problems with this element, you will be graded on presentation as well as correctness. For programs, presentation means code that is documented well enough to enable someone to understand what you have done, even someone who may not be familiar with the programming language you are using. For plots, that means adequate labeling of the axes and the contents of the plot. You should submit all results on paper, not via email. You don't have to be fancy, indeed you shouldn't be, but you should be clear. Remember that a major part of science is making clear to others what you have done.

2.1. Derive the procedure for computing Weibull-distributed random deviates.

2.2 Write a program to produce Gaussian, Cauchy, and Pareto-distributed variates from uniformly-distributed ones. For the Pareto, show a derivation of your procedure as well as the code; in the code, use a shape factor of 1.5. You should lay out the code that only calls a generator of uniform random numbers; don't just call an internal function that does this. For the Gaussian, use the polar method. Show a histogram with the pdf to check against errors. Now plot the mean and median of the first n terms of each set against n; that is, for the mean plot

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

and for the median plot the point that is halfway through a sorted list of the first n values, for n from 1 to 5000. Include the value for the mean and median of the theoretical pdf as appropriate. You should see the mean settle down rapidly for the Gaussian, slowly for the Pareto, and not at all for the Cauchy variables.

2.3 On the website, you will find a file of times for California earthquakes of magnitude 5 and greater since 1935 (roughly the start of the local networks). Compute the inter-event times in years (yes, this means dealing with the calendar). By trial and error (that is, by making normalized histograms and varying the parameters of the pdf, and plotting the pdf on top), see how well you can fit these with a gamma function and a log-normal distribution. From your best fits, find the time interval for there being a 50% chance of another earthquake of this size after one has happened.