Assignment 2: Parametric Distributions

In this assignment you will be exploring the use of parametric distributions. You may complete the assignment using your programming language of choice. Feel free to use built in functions but make sure you have read the documentation about these functions and are confident they are indeed conducting the calculations you intend. Please submit your assignment and the code used to generate any results by uploading the files to Canvas by the assignment due date. Your code should be well commented so that others can easily understand what has been done and marks may be removed from your assignment if this is not the case.

1. In this question, you will use the same Penn State weather station data as in Assignment 1. As was done in Assignment 1, remove all missing data and use only days with precipitation greater than zero (no trace amounts). In this question we will be using the Tmax and precipitation data from this record to explore the Central Limit Theorem and the Extremal Types Theorem.

- Compute the annual average Tmax and precipitation amounts for each year in the record. According to the central limit theorem the resulting annual means should follow a Gaussian distribution. Fit a Gaussian distribution to these annual average Tmax and precipitation amounts, estimating the parameters μ and σ using a method of moments.
- For both Tmax and precipitation, create a histogram of the annual average values and superpose the fitted Gaussian distribution. Ensure you use a density histogram so that there is a correspondence between your histograms and your PDFs. Include in your legend the value of the parameters that have been fitted to the data. Discuss the resulting histograms and the Gaussian fit.
- Repeat the analysis above but this time for the annual maximum Tmax and annual maximum precipitation, which is the maximum daily Tmax or daily precipitation amount in each year. According to the Extremal Types Theorem the resulting annual maximums should follow a generalized extreme value distribution. For simplicity, we will assume this fits a Gumbel distribution ($\kappa = 0$). Use a method of moments to estimate the parameters β and ζ.
- For both Tmax and precipitation, create a histogram of the annual maximum values and superpose the fitted Gumbel distribution. Ensure you use a density histogram so that there is a correspondence between your histograms and your PDFs. Include in your legend the value of the parameters that have been fitted to the data. Discuss the resulting histograms and the Gumbel fit.
- Create a quantile-quantile plot of the data vs. the Gumbel fits to further examine the goodness of the fit qualitatively. What does the q-q plot suggest about the goodness of fit of your distribution?

2. In this question, you will use data from a 2-D Video disdrometer that was in operation in the Tropical Western Pacific on January 2nd 2015. This instrument is able to capture the fall speed, size, and shape of individual raindrops. There is a known relationship between the number density of raindrops and rain drop diameter called the Marshall-Palmer Index. The Marshall-Palmer index is a variation on a Gamma distribution. The original and most commonly used version is an exponential ($\alpha = 1$ in the gamma distribution) multiplied by a constant defined as follows:

$$N(D) = N_0 e^{-\lambda D} \tag{1}$$

where D is the drop diameter, $N(D)$ is the number density of raindrops of size D, N_0 is called the intercept parameter, and λ is called the slope parameter

The data file for this question is located in the class data folder and is called `twpvdisC3.b1.20150102.000000.cdf`. It is in netcdf format, although the file suffix is .cdf. Within this datafile are variables for the number density (num_density) and drop diameter (drop_diameter) for each minute of the day. Also within the file are estimates of the intercept parameter (intercept_parameter) and slope parameter (slope_parameter) for each minute of data.

- Choose one example minute from the data file to examine its drop size distribution. Choose this specific minute such that you have some non-zero number density for higher drop diameters (ie. you have some drops of sizes 2-3mm). Make sure to indicated the chosen time in your subsequent figure. Create a semi-log plot of the number density vs. drop diameter for each of these days, where the y-axis is a logarithmic scale. Add the function $N(D)$ produced using the estimated λ and N_0 values from the file to this plot.
- A more complex form of the Marshall Palmer index with three parameters can be used to provide a better estimation of the drop size distribution. The equation is as follows:

$$N(D) = N_0 D^{\mu} e^{-\lambda D} \tag{2}$$

  Add an additional line to your plots above that is $N(D)$ computed using this 3-parameter version. Test out various values for N_0, λ, and μ and choose a final set that provide a better approximation of the drop size distribution than the original 2-parameter formulation.