EXERCISE SET 1: ANALYZING WIND DATA

JULIAN HAIMOVICH, JIEHUA CHEN AND MATT ZEBIAK
ASSIGNMENTS WEBSITE

The purpose of this problem set is to gain familiarity with some basic MATLAB array and plotting tools in order to analyze hourly wind speed data. To begin the exercise, first locate your computer's MATLAB folder (default) or your own custom working directory in MATLAB. Paste Iowa's hourly wind speed data (mdatahw1.xls, from the Input Data/Files folder) into that directory, and develop a MATLAB script to:

1. Load the data set (mdatahw1.xls) into MATLAB using the xlsread function, and find any "suspicious" outliers in the dataset. Comment on their characteristics. What makes them "suspicious"? Note that the first column is a time stamp in the format of year/month/hour/minute (for example June 15 2011 at 2:50 would be 201106150250), and the second column contains the wind speeds in mph. Remove these outliers from the dataset.

2. Draw a histogram of the wind speed without the outliers. The wind speed should be in m/s (you should convert the windspeed from mph to m/s). Title the plot and label the axes. There should be approximately 8600 data points (8760 minus the "suspicious" points).

3. Draw histograms of the summer (June through August) and winter (December through February) wind speed separately, but on the same figure (Hint: use the subplot function). Do not forget to include titles and axis labels for the plots. What is the relationship between the summer and winter wind?
   In general, this is best done by examining what part of the time stamp corresponds to the times you are looking for. You can cut the time stamps into individual year/month/etc. components by first converting it to a string using the num2str function, and selecting the needed information. For example, the minute on the time stamp can be selected by first converting the entire date into a string, and then selecting columns 11 and 12 from each piece of date data. You can then convert the minute string back to a double for simple comparison using if statements. This process is shown below:

   ```
   date_string = num2str(dates);
   minute_string = date_string(:, 11:12);

   for i = 1:length(minute_string)
       minute_double(i) = str2double(minute_string(i, :));
   end
   ```

   The end result of this process is an array of doubles, minute_double that contains only the minute of each piece of recorded data. The method used to generate the array will help you separate the winter and summer times, as well as the day and night times.

4. Draw histograms of the night (hours 19 to 6) and day (hours 7 to 18) time wind speed separately, but on the same figure. Again, do not forget a title and axes labels. What is the relationship between the day and night wind?

Date: June 29, 2012.

5. Draw a scatterplot of the average hourly wind speed (in other words, average the annual wind speed data for each individual hour and draw the scatterplot of those averages - average the wind speeds for the 1st hour of each day, and repeat the process for the other 23 hours). Also scatterplot the average hourly wind speed separately for winter and summer. The three plots should be done on the same figure. The x-axis will be hours and the y-axis wind speed (in m/s).

6. Plot the empirical density function of the wind speed data. This can be done via the following code:

```
 %get the counts and values of bin centers from the histogram
[N, xout] = hist(wspd_ms);
 % calculate the frequencies of each bin
freq_wspd_hist = N/8671;
 %calculate bin sizes
binsize = diff(xout);
 %since the bin sizes are all the same,
 %we can just take the first value of the calculated bin sizes
binsize = binsize(1);
 % calculate the empirical densities of wind speed for each bin
dist_wspd_hist = freq_wspd_hist./binsize;
 %plot the bar plot of the empirical densities
bar(xout, dist_wspd_hist)
```

Compare the distribution of the wind speed data with a Weibull Distribution and a Rayleigh Distribution with the mean equal to the sample mean of the data: add the estimated Rayleigh distribution curve to the histogram plot. Do not forget to include a legend.

To plot the Rayleigh and Weibull Distributions: First, compute the parameters:

```
phat = mean(wspd_ms)/sqrt(pi/2); %scaling parameter
[parmhat, parmci] = wblfit(wspd_ms(wspd_ms>0));
```

phat will be the scaling parameter equal to the mean of the sample mean of the data. parmhat and parmci contain the lower bounds of the confidence intervals for the parameters, and the upper bounds of the confidence intervals, respectively. Note that the wind speed values used must be positive (non-zero) values, or an error will be thrown.

The following code will plot the Rayleigh and Weibull Distributions:

```
plot(0:0.1:30, wblpdf(0:0.1:30, parmhat(1), parmhat(2)),
0:0.1:30, raylpdf(0:0.1:30, phat));
legend('Fitted Weilbull Density (a=5.32, b=1.94)',
'Fitted Rayleigh Density (a=3.38, b=2)');
xlabel('Wind Speed (m/s)');
ylabel('Density');
grid on;
```

7. Estimate Weilbull distribution parameters from this dataset using only positive (non-zero) values; and estimate Weilbull distribution parameters for winter data and summer data separately. Plot the distributions for the entirety of the dataset, the winter wind, and summer wind on the same figure. To do this, first calculate the parmhat and parmci variables for the non-zero summer and winter wind values and then plot them using the distributions using the following code:

```
plot(0:0.1:30, wblpdf(0:0.1:30, parmhat(1), parmhat(2)),
```

```
0:0.1:30, wblpdf(0:0.1:30, parmhats(1), parmhats(2)),
0:0.1:30, wblpdf(0:0.1:30, parmhatw(1), parmhatw(2)) );

legend('Fitted Weilbull Density (a=5.32, b=1.94)',
'Fitted Summer Weilbull Density (a=4.26, b= 2.29)',
'Fitted Winter Weilbull Density (a=6.06, b= 1.92)');
xlabel('Wind Speed (m/s)');
ylabel('Density');
grid on;
```

Note: If you wish to run the solutions directly, it helps to first paste the file, HW1_solution.m, into your present working directory in MATLAB. This code is located in the solutions folder, and the PDF with explanations is also included. Final figures may be found in the figures folder.

E-mail address: jc3288@columbia.edu, jsh2146@columbia.edu and msz2112@columbia.edu