MET3220C
Computational Statistics

Programming – AS #5

Dr. Mark Bourassa

Turn in your program and graphics.

UNIX Tip

- If you want to repeat a previous command
  - e.g., f90 as4.f90 correlate.f90 best_fit.f90 guassian.f90
  - You can speed the process up by using the exclamation point
  - E.g., !f
  - This would repeat the last UNIX command that started with f
  - !f90 would repeat the last UNIX command that started with f90

Assignment #5
Histograms

- Read data from a record of unknown length
- Condense data into a histogram
- Compare observation-based histogram with Gaussian and log-normal distributions.

Reading Data Sets of Unknown Length

- How do you read data if you don't know how many data points are in the data set?
- In this example, the file has >800,000 observations of wind speed. These are wind speeds from one day of observations from the SeaWinds instrument on the QuikSCAT satellite.

```
OPEN( 7, status='old', file='/u/a/met3220-02/winds.dat')
i = 0
300 READ(7,end=400, FMT='(F13.5,7X,I1 )') qscat_speed, qscat_flag
   print*, i, qscat_speed, qscat_flag
   i = i + 1
   GOTO 300
400 CONTINUE
CLOSE(7)
```

- The GOTO 300 sends the code to label 300
- The end=400 goes to label 400 when the end of file is reached

Making a Histogram

- We will examine wind speeds from zero to 60m/s,
  - in bins that are 0.5 m/s wide
  - E.g., 0 to 0.5, 0.5 to 1.0, 1.0 to 1.5, and so on to 59.5 to 60.0
- Declare a real array of 120 values (120 = 60 / 0.5)
- When binning the data, we only want data from within this range and that are not flagged as suspect

```
index_spd = qscat_speed / bin_width + 1
IF ( index_spd > 0 .AND. index_spd <= n_bins .AND. qscat_flag = = 1) THEN
    histogram_spd_obs(index_spd) = histogram_spd_obs(index_spd) + 1
    n_good_data = n_good_data + 1
ENDIF
```

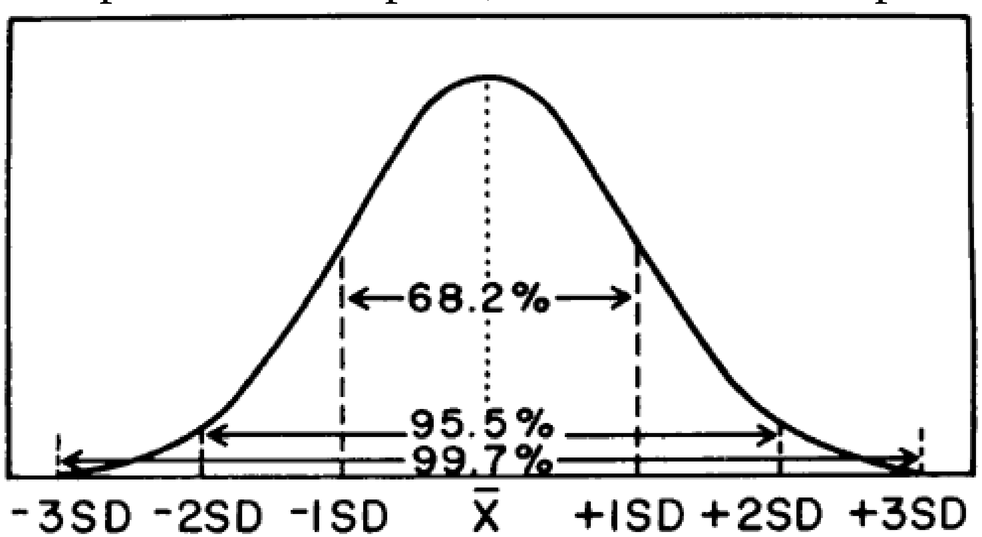
- It is easier to put all these conditions into one IF.

Gaussian Distribution: The Formula

- A normal distribution is described by two parameters: a mean (μ) and a standard deviation (σ).
- A Gaussian distribution (not a pdf) would also have an amplitude.

$$pdf = f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad -\infty < x < \infty$$

- Think about how the the standard deviation influences the shape of $f(x)$.
  - Larger σ implies a wider peak, and a smaller amplitude.



Graphic from http://homepage.univie.ac.at/Franz.Vesely/cp0102/dx/img579.png

Log-Normal Distributions

- There are many occurrences of distributions that have
  - (1) only positive values, and
  - (2) peak is displaced to the left.
- Some of these distributions are log-normal distributions.
  - A transformation of variables is used: $y = \ln(x)$

$$pdf = f(x) = \frac{1}{x\sigma_y\sqrt{2\pi}}\exp\left[-\frac{\left(\ln(x)-\mu_y\right)^2}{2\sigma_y^2}\right], \quad -\infty < y < \infty, \; y = \ln(x)$$

  - Where μ_y and σ_y are the mean and standard deviation of the transformed variable y.

- You will want to calculate the mean and standard deviation of $\log(x)$

Theoretical Distributions
What Points to Plot?

- For visualization purposes, you could plot at any interval that you want.
  - N_pts = (max_value – min_value) / resolution
  - X_value(j) = min_value + (j-1) * resolution   !in a DO loop
- For statistical comparisons (which we will do in later classes) you would want the values at the center of the bins, or better yet the mean of the values in the bin.
  - x_bin_center = (REAL(index_spd) - 0.5) * bin_width.