

## **ASSIGNMENT 2. PROBABILITY DISTRIBUTION**

1. Run geosa2.m, selecting one series from either V1 or V2 for analysis. Run the script on either the full series length or some sub-period. You will turn in figure windows 1, 2 and 3 with answers in the figure captions.
2. (Caption to Fig. 1) Time series plot. Describe your plotted time series, commenting on a couple of prominent features related to spread.
3. (Caption to Fig. 2) Box plot, normal probability plot and histogram with superposed normal density plot. Is the selected time series segment normally distributed? Comment on how each of the three graphics and the Lilliefors test support your contention
4. (Caption to Fig. 3) Cumulative distribution function and Lilliefors test. At what standardized value does the cdf of your time series differ most from the cdf of the normal distribution. What is the non-exceedance probability of your time series at that standardized value? What is the non-exceedance probability of the normal distribution at that standardized value? How is this maximum difference in cdfs relevant to the Lilliefors test for normality? Would your conclusion change if the time series were first converted to “z-scores” (subtract mean, divide by standard deviation) before running the Lilliefors test?
5. Put your figures with captions in a single word processing document. Save or print the document as pdf and email it to me as an attachment. Give your pdf file the same name as your last name, with “2” appended to indicate assignment # 2 (e.g., “smith2.pdf”)

Running geosa2.m

1. >geosa2
2. Message box: message introducing geosa2.m; click OK to remove message and move on
3. Respond to input dialog with the name of your data file; click OK
4. Menu: Select the data set the time series is to come from. Can be V1, V2 or V3.
5. Menu: Select the time series. An "\*" appears opposite the selection. If OK, click to accept.
6. Message box: summary information on selected series. Click OK.
7. Edit dialog: enter start and end year for analysis and click OK

The default initially in box is the full period of coverage of the time series

The three figure windows appear, along with a message box telling how you can extract some statistics after running the script. Click OK to close message box

Fig 1. Time series plot, with horizontal line at analysis-period mean

Fig 2. UL: Box plot. UR: Histogram with normal density superposed. LL: normal-probability plot

Fig 3. Lilliefors test results

## PROGRAMMING NOTES

Empirical non-exceedance probabilities in plots in figure 3, top, computed from  $j/(N+1)$ , where  $j$  is the rank and  $N$  is the sample size

Selected Matlab functions called by geosa2:

**mean** – sample mean

**median** – sample median

**std** – sample standard deviation

**boxplot** – box plot

**histfit** – plot histogram with overlaid normal probability pdf

**lillietest** - compute Lilliefors's statistic to test for normality

**nanmean** – compute sample mean ignoring missing data

**normcdf** – statistics related to cdf of normal distribution

**normplot** – normal probability plot, to graphically assess whether sample from normal distribution

**skewness** – compute skewness of a sample

**zscore** – convert data to standardized departures from the mean