LAB 2 · CS249B · FIRST STEPS WITH MATLAB

DANIEL BILAR, WELLESLEY COLLEGE

Out: Thursday, March 6^{th}, 2008; In: Thursday, March 13^{th}, 2008

1. Goal

The aim of this lab is to introduce you to the Swiss army knife ("It slices! It dices! It draws!") of scientific computing, *Matlab*. You are going to perform some mathematical analysis of the network of lab 1.

Matlab is very powerful, we are just going to use a minuscule fraction of its power. It has an inbuild tutorial system which is very well done. Type `help` and you will see[1].

2. Fitting powerlaws with Matlab

(1) First, copy the seven m-files from `http://cs.wellesley.edu/~cs249B/m-files/` to the working directory C:\Program Files\MATLAB\R2006a\work. We will need these so-called m-files (you can view them as little programs. CS people: They are scripts) for the exercises in Matlab.

(2) Start Matlab and generate 100,000 random integers from a power law distribution with exponent $\alpha = 2.1$ What is the largest value in your sample? Is it possible for a node in a network to have a degree this high (assuming you don't allow multiple edges between two nodes)?

Procedure. Use the script randompowerlawints.m as follows: At the Matlab prompt, type
`> x = randompowerlawints(numsamples,alpha)`
where `numsamples` is an integer value specifying the number of integer samples you would like, and `alpha` is the power law exponent in the range from 1 to very big, although it gets boring for `alpha` ≥ 3.5

□

(3) Construct a histogram of the frequency of occurrence of each integer in your sample. Turn in both a linear scale plot and a log-log scale plot

Procedure. Use the script binvector.m as follows: At the Matlab prompt, type
`> [binval, binfreq] = binvector(x)`
If `x` is not already an integer array, it will round each value to the nearest integer, and then count how many times an integer occurs. This will be a *histogram* of your data. `binval` is the list of rounded x values, and `binfreq` is the list of corresponding frequencies.

To plot the data on a linear scale:
type `> plot(binval, binfreq,'ro')` where 'ro' marks the points with red circles

[1] For a basic steps tutorial, go to BU; for a tutorial on plotting, go to Bucknell

To plot the data on a log scale:
type `> loglog(binval, binfreq,'ro')` where 'ro' marks the points with red circles

□

(4) Do the steps above(generation and plotting) with distributions from from a power law distribution with exponent $\alpha = 1.7$ and $\alpha = 3.3$. Your computer may hang trying to plot $\alpha = 1.7$ and the plot may be boring for $\alpha = 3.3$. What is the reason for this?

(5) Generate 100,000 random integers from a power law distribution with exponent $\alpha = 2.1$. Now we are going to fit the data twice - in its histogram form (the PMF) and in the preferred form, the CDF. For both fittings, remember to note the fitted α exponent.

(i) Fitting histogram (PMF) data

Procedure. Use the script fitMLE.m as follows: At the Matlab prompt, type
`> alpha = fitMLE(x, xmin)`
where `xmin` is the cut-off point on the left (i.e. we do not consider x values smaller than xmin

□

The function will return the value of the power-law exponent*(-1) and show the fitted line on the log log plot. There is a factor of -1 because we are fitting $x^{-\alpha}$ and the m-file returns the slope, which is -α.

(ii) Fitting the CDF

Procedure. Use the script cumulativecounts.m on your original x data as follows: At the Matlab prompt, type
`> [xlincum,ylincum] = cumulativecounts(x)`

Then, with the data `xlincum` and `ylincum` - similar to the procedure for the 'CDF light' above - compute the power exponent α

□

(iii) Discuss the two alphas you noted. They were produced by fitting the same data (albeit in different forms), produced from a power-law distribution of $\alpha = 2.1$ (if you followed the instructions). Why do you think the alphas are different? (Hint: lecture 2/11/08, Pareto Distribution)

3. Handin

Please write up your solutions *neatly* (I cannot grade what I cannot read) or typeset them and hand them in in class at the due date. You are allowed to discuss the problems and solution approaches with your classmates; copying is not allowed. Please indicate which problems you did with whom.