MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Civil and Environmental Engineering

1.017 Computing and Data Analysis for Environmental Applications

Practice Quiz 1, 2003

Problem 1 (Descriptive Statistics)

During the manufacture of links in a chain, they are sampled at random and the tensile strength measured. Characterize the set of 12 maximum load values given below by plotting or computing the following descriptive statistics:
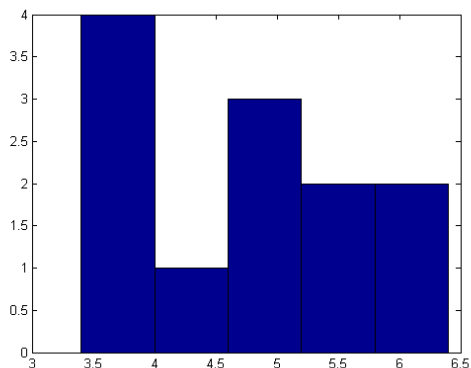
Histogram
Sample cumulative distribution function (CDF)
Sample mean
Sample standard deviation
Sample median

6.2	3.8	5.0	4.8
3.4	5.3	3.9	6.4
4.2	5.5	5.2	4.0

Use the sample CDF to estimate the probability of a link having a tensile strength less than or equal to 4.
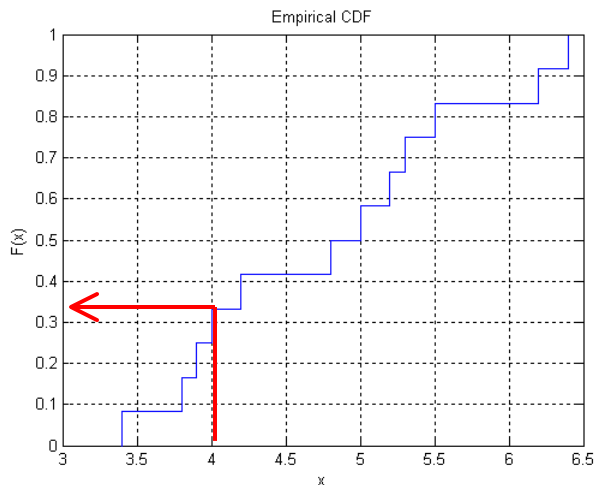
Solution:

Histogram:

Mean = 4.8
Standard Deviation = 0.96
Median = 4.9

CDF:

Empirical CDF

P(strength<=4) = 4/12 = 0..33

Problem 2 (Combinations and Virtual Experiments)

A geologic formation consists of 9 layers: 4 shale, 3 sandstone, and 2 limestone. We presume that these layers were laid down at random. Now suppose that we drill a borehole through the top 3 layers. What is the probability that these sampled layers are all shale?

If you were to write a Matlab program that estimates the probability that the top three layers are shale using a Monte Carlo technique, how would you do it? Describe in both words and code the steps you would take and the logical order of the program (write a pseudocode). You will not lose points for small syntax errors, but the steps should be thoroughly explained and correct.

Solution:

Analytically:
P(top 3 shale) = (4/9)(3/8)(2/7) = .0476

In MATLAB:
```
nrep=10000;
count=0;

% 9 layers: 1-4 = shale, 5-7=sandstone, 8-9=limestone
% create a for loop that goes from 1 to the number of
%repetitions. Create a (nrep x 9) matrix 'formation' of the
%9 different layers in each rep using the randperm function
%in matlab. Randperm(9) randomly arranges the integers 1-9,
%returning a (1 x 9) vector.
```

```
for i=1:nrep
    formation(i,:)=randperm(9);

    % Create an if statement that counts the number of
%times the top three layers are shale, ie how many times
%are the first three numbers in each row of formation less
%than 5? Each time this occurs, add one to the count.

    if formation(i,1)<5 formation(i,2)<5&formation(i,3)<5
            count=count+1;
    end
end

% calculate the probability as the final count divided by
%the total number of replicates.
p=count/nrep
```
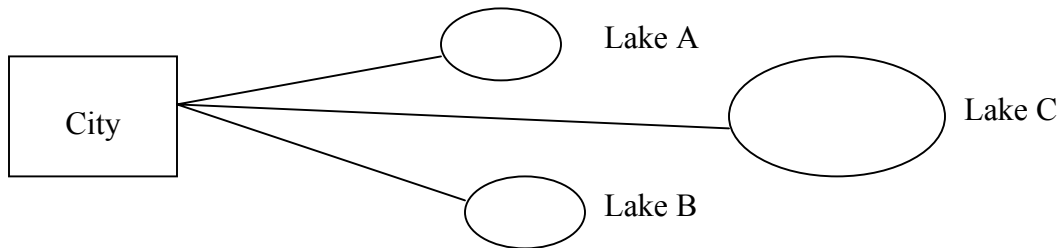
Problem 3 (Conditional Probability and Bayes Theorem)

There are three lakes, A, B, and C, in a city and their water is used as water supply. Lake C is located very far from the city, and transporting water to the city is expensive, but the environment there is clean. Lakes A and B are more likely to be polluted. Water from any one lake is adequate to meet the demand of the city. Every day, a test is conducted for each lake, and if the lake is found to be polluted, the water in that lake cannot be used for the water supply. We know from past record that lake A is unpolluted with a probability 0.3, lake B is unpolluted with probability 0.2, and lake C unpolluted with probability 0.8. We also know that if lake A is polluted, the probability that lake B is unpolluted is 0.1. If lakes A and B are both polluted, the probability that lake C is unpolluted is 0.8.

a.) What is the probability that we cannot get unpolluted water from lake A or B?
$$P(\sim A \sim B) = P(\sim B|\sim A)*P(\sim A)$$
$$P(\sim B|\sim A) = 1-P(B|\sim A) = 0.9$$
$$\therefore P(\sim A \sim B)=(0.9)(0.7)=0.63$$

b.) What is the probability that the city can get unpolluted water?

$$P(A \cup B \cup C) = 1 - P(\sim A \sim B \sim C)$$
$$P(\sim A \sim B \sim C) = P(\sim C | \sim A \sim B) * P(\sim A \sim B)$$
$$P(\sim C | \sim A \sim B) = 1 - P(C | \sim A \sim B) = 1 - .8 = .2$$
From above, $P(\sim A \sim B) = 0.63$

$$\therefore P(A \cup B \cup C) = 1 - (.2)(.6) = 0.87$$

Problem 4 (Binomial Distribution)

The probability that a concrete cylinder has a crushing strength below the specified minimum is 0.2. The cylinders come in shipments of 6. What is the mean number of cylinders per shipment that are below the minimum?

Solution:

For the binomial distribution, $E(X) = np$. $\therefore E(x) = (6)(.2) = 1.2$

Problem 5 (Probability Distributions)

Streamflow in a particular river on a given day is approximately an exponentially distributed random variable. It's PDF is given by:

$$f(\text{streamflow}) = 5e^{-5x} \quad x \geq 0; \quad = 0 \text{ otherwise}$$

Calculate the mean streamflow in this river.

Solution:

$$\text{Mean} = \int xf(x)dx = \int 5xe^{-5x}dx = 1/5$$