MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Civil and Environmental Engineering

1.017/1.010 Computing and Data Analysis for Environmental Applications/ Uncertainty in Engineering

Quiz 1 (Solutions provided at the end of each problem)
Tuesday, October 7, 2003

Please answer all questions on a separate piece(s) of paper with your name clearly identified:

Problem 1 (55 points)

a: (10 points) A soil sampling experiment classifies soils into 5 discrete porosity values [0.1, 0.2, 0.3, 0.4, 0.5]. The experiment generates 100 samples that can be described with the following histogram (the number of samples corresponding to the 5 porosity values are: [14, 37, 18, 26, 5]):
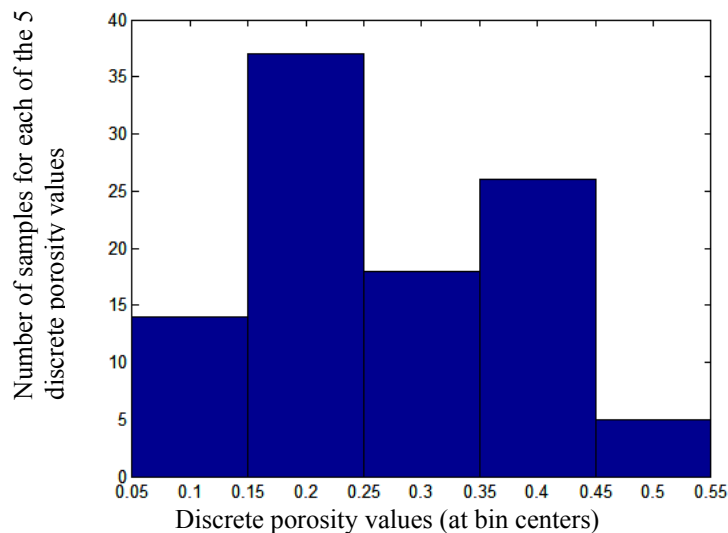
Use this histogram to derive a discrete probability mass function (PMF) and cumulative distribution function (CDF) for the soil porosity. Plot the PMF and CDF on the axes provided below (or on your own if you prefer). Be sure to include a label and scale for each vertical axis.

b: (4 points) What are the outcomes and sample space associated with this experiment?

c: (6 points) Compute the probabilities of the following events:
$A_1 = \{x > 0.35\}$, $A_2 = \{x = 0.1\}$, $A_3 = \{x = 0.2 \text{ or } x = 0.3\}$

d: (6 points) Are these events mutually exclusive? If so, show that they satisfy the third probability axiom (additivity of mutually exclusive event probabilities).

**e: (9 points)** What are the mode, mean, and variance of your PMF?

f: (20 points) Write a computer program in MATLAB or in words (pseudocode) to carry out a virtual experiment that draws sets of 10 samples from the probability distribution constructed in Part a). This experiment should evaluate the probability of obtaining 5 porosities greater than 0.25 from the set of 10 porosity samples. Your program should specify how you would do the following:

1) Use the MATLAB uniform random number generator `rand` to generate 1000 random replicates of 10 porosity values for each replicate. The numbers obtained from `rand` should be transformed to give discrete random porosity values that follow the probability distribution derived in a). Note each the MATLAB statement `rand` produces a single number distributed uniformly over the interval [0,1] while `rand(n,1)` produces a vector of n such numbers.

2) Count the number of replicates yielding the event B_1 = {5 porosities out of 10 are greater than 0.25}

3) Evaluate the probability of this event.

Problem 1 Solution:

a: $p_x(x) = [0.14, 0.37, 0.18, 0.26, 0.05]$
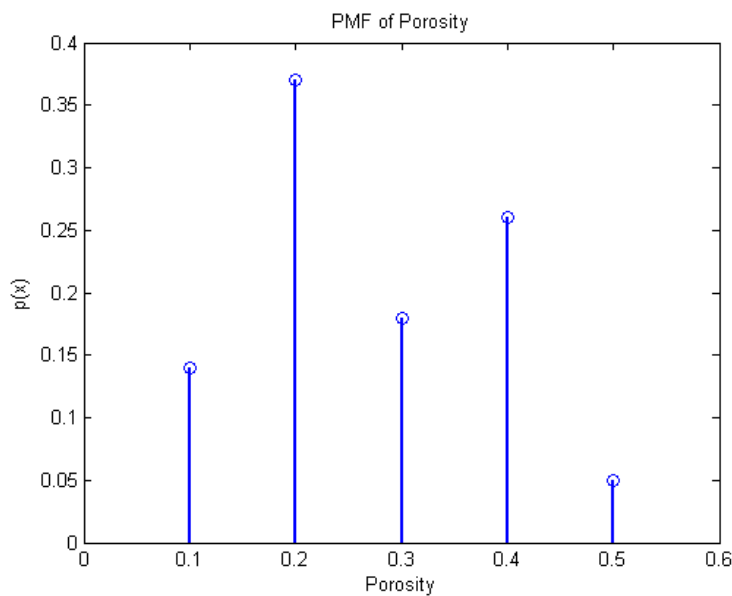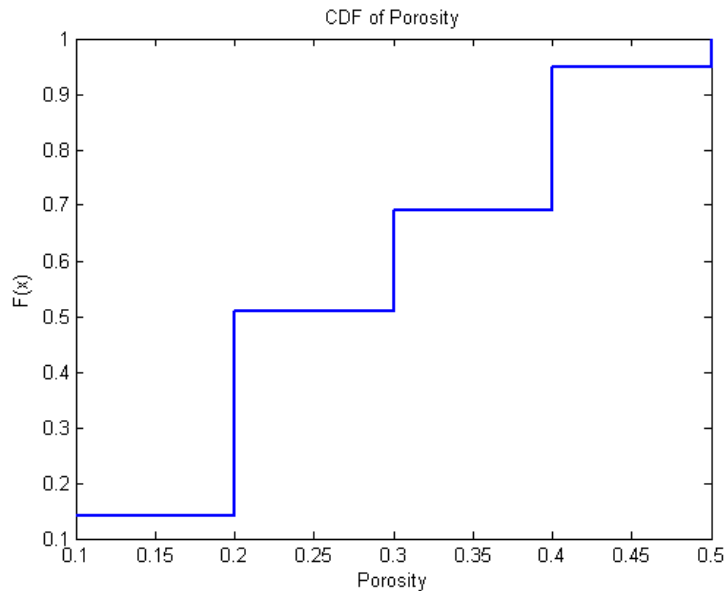
CDF of Porosity

b: Sample space is set of possible outcomes, which is $[.1\ .2\ .3\ .4\ .5]$

c: $P(A1) = .31$ $\qquad P(A2) = .14$ $\quad P(A3) = .55$

d: The events are mutually exclusive since the three events (A1, A2, and A3) do not include any common porosity values, and one soil sample cannot have two values of porosity.
$P(A1+A2+A3) = P(A1)+P(A2)+P(A3) = .31+.14+.55 = 1$

e: Mode $= .2$
Mean $= 1(.14) + 2(.37) + 3(.18) + 4(.26) + 5(.05) = 2.71$
Variance $= (1 - 2.71)^2 (.14) + (2 - 2.71)^2 (.37) + (3 - 2.71)^2 (.18) + (4 - 2.71)^2 (.26)$
$+ (5 - 2.71)^2 (.05) = 1.3059$

f:

```
% Quiz 1 Problem 1(f)

nrep=1000;

%create matrix of random numbers
randoms=rand(nrep,10);
% using CDF(.25), matrix of ones where porosity > 0.25
greater_25=randoms>0.51;
% how many in each set of 10 are greater than .25?
number_greater=sum(greater_25,2);
% Calculate probability:
P5=sum(number_greater==5)/nrep
```

This gives $P5 = 0.245$

Problem 2 (15 points)

A PCB measurement at a low cost sampling station correctly detects a violation (i.e. the measured PCB concentration exceeds a specified standard) for only 80% of the samples that are flagged as violations by a very accurate (and very expensive) instrument. Also, the low cost station measurement incorrectly yields a violation for 10% of the samples that do not yield a violation at the accurate station. 15% of all samples tested at the accurate station are violations. What is the probability that the PCB concentration actually exceeds the standard if the low cost instrument detects a violation.

Problem 2 Solution:

A = exceeds standard (accurate instrument detects violation)
B = violation detected

$P(A)=.15$
$P(B|A)=.8$
$P(B|{\sim}A)=.1$
$P({\sim}A)=0.85$

Looking for $P(A|B) = P(AB)/P(B)$
$P(AB)=P(B|A)P(A)=(0.8)(.15) = 0.12$
$P(B) = P(B|A)P(A)+P(B|{\sim}A)P({\sim}A) =(0.8)(.15) + (.1)(0.85) = 0.205$

Therefore, $P(A|B) = 0.585$

Problem 3 (10 points)

Suppose that the cost of a rail fault detection program is $150,000 per year. The probability that a fault disrupts traffic for more than 1 day during the year is 0.01 with the detection program and 0.05 without the program. The costs arising from a repair shutdown of 1 day or more is $ 4 million. Is the fault detection program worth the investment, based on the 1 day shutdown cost? Why?

Problem 3 Solution:

Calculate the average yearly cost:
    With program: $4E6 * .01 + $150,000 = $40,000 + $150,000 = $190,000
    Without program: $4E6 * .05 = $200,000
Therefore, program is worth it since the expected savings is $10,000/yr

Problem 4 (20 points)

Suppose that, at any given time, 10 oceanographic floats are dispersed at random throughout 100 cells in a square (10 by 10 cell) grid. What is the probability that a ship traveling through a row of 10 cells will intercept more than 1 float; assuming that the floats do not move over the time it takes the ship to cross the 10 cells? Also, assume that a float located anywhere in the same cell as the ship is intercepted. Calculate this probability using a combinatorial approach

Problem 4 Solution:

Use combinations (without replacement, we don't care about order):

P(>1)=1-P(1)-P(0)

P(1) = C(1,10)C(9,90)/ C(10,100) = .41
P(0) = C(0,10)C(10,90)/(10 C 100) = .33

P(>1) = .26