MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Civil and Environmental Engineering

1.017 Computing and Data Analysis for Environmental Applications

---

Quiz 2
Thursday, November 9, 2000

Please answer any 5 of the following 6 problems (maximum score = 100 points):

Problem 1 (20 points)

Suppose that a remote sensing signal z (measured in watts/m^2) is the sum of two components x and y which represent emissions from two adjacent regions on the ground (i..e. $z = x + y$). The independent discrete random variables x and y can have the following values $x = \{1 ,2 \}$, $y = \{ 1 , 2\}$, depending on land use. The probabilities of these x and y values are:

$p_x(1) = 0.3 \quad p_x(2) = 0.7$

$p_y(1) = 0.8 \quad p_y(2) = 0.2$

a) What are the possible values for z (present as a table)?

b) Derive the probability mass function $p_z(z)$, assuming that x and y are independent (i.e. $p_{XY}(x,y) = p_X(x)p_Y(y)$). Plot the probability mass function and cumulative distribution of z vs z.

c) What is the mean of z?

Problem 2 (20 points)

Suppose that we want to have a representative set of air samples to evaluate air quality in a city. We suspect that ozone levels exceed air quality standards 20% of the days. These high ozone days are distributed randomly throughout the year. We collect 12 samples, one on the first of each month. Use a normal approximation to evaluate the probability that we obtain 4 or more samples that violate the standard (tables attached) if the ozone levels do, in fact, exceed the standard on 20% of the days?

Problem 3 (20 points)

a) A geological formation consists of 9 layers -- 4 shale layers, 3 sandstone layers, and 2 limestone layers. 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?

b) Write a MATLAB program which estimates the probability in Problem 3 using a Monte Carlo technique.

Problem 4 (20 points)

(Try to answer Part b for 1/2 credit even if you cannot answer Part a)

a) Consider the following simple model of runoff production, where Q = runoff (mm/hr), R = rainfall (mm/hr), and I = infiltration capacity:

$Q = 0$ if $R < I$
$Q = R - I$ if $R \geq I$

Suppose that R has an exponential probability density function $f_R(\mathrm{R}) = a\exp(-aR)$ with distributional parameter a = 0.2 $(\mathrm{mm/hr})^{-1}$. Plot the cumulative distribution function of Q when I = 3 mm/hr. Make your plot as accurate as possible. {HINT: The random variable Q is a mixed discrete/continuous random variable).

b) Write a MATLAB program which uses a Monte Carlo approach to evaluate the probability that $Q = 0$. Assume that you have available the internal MATLAB function `exprnd(mu,m,n)`, which provides an `m` by `n` array of independent exponentially distributed random variables with parameter `mu` = $1/a$.

Problem 5 (20 points)

A few days after a heavy rain storm a particular Boston harbor beach can be affected by combined sewer discharges at three points (A1, A2, and A3) with probabilities of overflowing of 0.2, 0.5, and 0.3, respectively (one and only one of the points will discharge in any given storm). The probability that post-discharge coliform levels are high enough to require a beach closing is 0.3, 0.4, 0.2, for a discharge from A1, A2, and A3, respectively. What is the probability that A1 discharged during a particular storm if the beach was subsequently required to close? What is the probability that the beach will have to be closed after a storm?

Problem 6 (20 points)

Find the mean and variance of a random variable x with the following probability density function:

$f_X(x) = 6x(1 - x)$ for $0 < x \leq 1$

Plot the density function and the cumulative distribution function of x.