MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Civil and Environmental Engineering

1.017 Computing and Data Analysis for Environmental Applications

Quiz 3
Tuesday December 12, 2000

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

Problem 1 (20 points)

Consider a random sample consisting of n pairs of concentrations $(X_1, Y_1), (X_2, Y_2),..., (X_n, Y_n)$ for two solutes X and Y. Suppose that you wish to fit the following trend function to the data:

$$y(x) = a_1 x$$

Assume that the measurements can be described by:

$$Y_i = a_1 X_i + V_i \ \ ; i = 1,... \ n$$

where the V_i are a set of independent, identically distributed random variables with mean 0 and variance 1.0.

a) Derive an estimator for the parameter a_1.

b) Show that this estimator is unbiased

Problem 2 (20 points)

Derive a two-sided 90% confidence interval for the mean of a normally distributed random variable X given the following random sample of X. State any assumptions that you need to make.

$$[x_1, x_2 , ... \ x_n] = [\ 2, 1, -8, 3, 6, 4, 2, 0, 5, -4]$$

Problem 3 (20 points)

The results of a water quality screening test are summarized by assigning $X = 0$ if the sample fails and $X = 1$ if it passes. A series of tests on a set of $n = 100$ independent samples drawn from the same population yields 80 passes and 20 fails. Define a test statistic and provide the p value for a two-sided test of the hypothesis that the fraction of passes in the population is equal to 0.7. Use a sketch to illustrate the test statistic distribution and the p value.

Problem 4 (20 points)

Write a MATLAB program which uses a Monte Carlo approach to evaluate and plot a sample density function of the **maximum** annual streamflow observed in a random sample of 20 values. Assume that the annual streamflow has the following exponential probability density:

$$f_X(x) = \frac{1}{a}\exp\left[-\frac{x}{a}\right] \quad ; \quad x \geq 0$$

$$f_X(x) = 0 \quad ; \quad \text{otherwise}$$

where a is the mean annual streamflow. The probability distribution of the maximum for this case is an example of an **extreme value distribution**. Suppose that the mean and maximum streamflows obtained from 20 actual annual stream flow measurements are 10 m^3/sec and 50 m^3/sec. How would you use your program to estimate the probability of obtaining, in another 20 year sample, a maximum greater than 50 m^3/sec.?

Problem 5 (20 points)

Suppose that you are given the MATLAB function `ahatcdf(aval,a)` which returns the cumulative distribution function (CDF) value of a test statistic `ahat` for any value `aval` of this statistic. The test statistic is used to test hypotheses about the distributional property a of a sampled random variable X (e.g a could be the mean of X). Note that the CDF of `ahat` depends on the parameter a. You are also given a function `ahat(x)` that returns the `ahat` (test statistic) value obtained from a particular random sample of X. The n measurements in the sample are assembled in an n vector `x`.

a) Write a MATLAB function `pvalue` which returns the p value for a one-sided test of the hypothesis **H$_0$:** `a = 1` vs. the alternative **H$_1$:** `a > 1` The data vector `x` should be passed to `pvalue` through the function argument.

b) Suppose that `ahat` is a positive continuous random variable (e.g. a sum of squares). Sketch a hypothetical probability density function of `ahat` showing the **H$_0$** acceptance region (**A$_0$**) for the hypothesis test in Part a). Indicate on your sketch the relationship between the p value and the value of `ahat` derived from the sample `x`.

Problem 6 (20 points)

Suppose that you wish to test the hypothesis **H$_0$:** $\theta = 0$ vs. the alternative **H$_1$:** $\theta = 1$. You have derived that the probability distribution function of your test statistic **given H$_0$:** $\theta = 0$ is normal with mean 0.0 and variance 1. Also, you have derived that the probability distribution function of your test statistic **given H$_1$:** $\theta = 1$ is normal with mean 1.0 and variance 4. What are the Type I and Type II error probabilities if your **H$_0$** acceptance region is A0 = [-∞, 0.5]? Provide a sketch which identifies each of these probabilities.