STA 410/2102 — Practice Questions for First Test

Please note that these questions do **not** cover all the topics that may be on the test.

1. For each of the following true statements about arithmetic on real numbers, give an example showing that it is not necessarily true for computer arithmetic on floating-point numbers in which the results are rounded to three decimal digits (ie, the mantissa consists of three decimal digits). In your examples, assume that overflow and underflow do not occur. (In other words, these should be examples of the effects of round-off error, not of the effects of overflow or underflow.)

   (a) If $x \neq y$ then $x + z \neq y + z$.

   (b) $x + (y + z) = (x + y) + z$.

   (c) If $x \neq 0$ then $x \times (1/x) = 1$.

2. When writing a program, it is sometimes necessary to determine whether two floating-point numbers, held in variables `x` and `y`, have the same sign, or different signs, and do different things accordingly. (Assume that it makes no difference which thing is done when one or both variables are zero.) Here is a way that you might consider testing for this:

   `if (x*y>0)`

   Do things appropriate for when the signs are the same

   `else`

   Do things appropriate for when the signs are different

   Why is this not a reliable way of accomplishing this task?

3. Write an R function to sum the first n terms of the following series:

   $$1 - \frac{1}{4} + \frac{1}{9} - \frac{1}{16} + \frac{1}{25} - \cdots$$

   The ith term of this series is $(-1)^{i+1}/i^2$.

   The function should take n as an argument, and should return the sum of the first n terms of the series as its value. Write the function so that it returns as accurate an answer as can be easily obtained. Although it might be most efficient to use the `sum` function built in to R, for this question, do *not* use `sum`, but instead write a loop. Otherwise, make the function as efficient as you reasonably can.

4. Consider the distribution on $(1, \infty)$ with density function $f(x) = 1/x^2$. Write an R function that takes as its argument a positive integer and which returns a vector of that length containing independent pseudo-random variates generated from this distribution. Show how you derived the formula or formulas required.

5. Some people discard as an "outlier" any data point that seems like it is too far from the sample mean, in comparison with the sample standard deviation.

   (a) Write a function called `trim.data` that takes a data vector and a value k as its arguments, and returns the data vector in which all points more than k times the sample standard deviation from the sample mean have been discarded.

   (b) Write a function that evaluates how trimming data affects the validity of a t test, by simulating many normally-distributed data sets (of some specified size) in which the null hypothesis that the mean is zero is true, trims each data set using `trim.data`, and then applies a t test to each trimmed data set. Try this with $k = 2$, $k = 2.5$, and $k = 3$. How uniform is the distribution of p-values in each case?