STA 410/2102, Spring 2003 — Assignment #4

Due at **start** of class on April 11. Worth 17% of the final mark.

Note that this assignment is to be done by each student individually. You may discuss it in general terms with other students, but the work you hand in should be your own.

In this assignment, you will solve a simple one-parameter Bayesian inference problem using the midpoint rule for numerical integration.

Suppose we model the counts $X_1, X_2, \ldots, X_n$ as being independent and identically distributed according to the Poisson distribution with mean $\exp(\theta)$. Based on our prior beliefs about θ, we specify that its prior distribution is normal with mean 0.7 and standard deviation 1.5.

Given a data set, we wish to find three things:

1. The posterior mean of θ.
2. The probability that θ is greater than zero (ie, that the Poisson mean, $\exp(\theta)$, is greater than one.
3. The probability that the next observation, X_{n+1}, will be a zero.

You should write an R function (using good style) that will find these three quantities for any given data set, using the midpoint rule for numerical integration. Your function should take as arguments the vector of data points $(x_1, x_2, \ldots, x_n)$ and the number of points to use for the midpoint rule. You may use the midpoint rule function that was used in the example program given out in class. However, you will need to somehow transform the required integrals to be over a finite interval before you can use this. You should try to do this and to otherwise implement this function so as to produce accurate answers with as few functions evaluations as possible.

You should test your function on at least the following two data sets:

Data set 1: 0 1 1 0 1 0 0 2 0 1

Data set 2: 5 8 5 5 4 8 6 6 4 6 3 8 1 3 7 7 4 8 5 7

You should try solving the problem with varying numbers of points for the midpoint rule, in order to see how many are needed to get accurate answers.

You should hand in your derivation of how you will adapt the midpoint rule for use in this problem, a listing of your R functions, the commands and output for your tests, and a brief discussion of how many points were required for the midpoint rule.