STA 410/2102, Spring 2002 — Assignment #4

Due at **start** of class on April 12. Worth 18% 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.

A vintner is interested in modeling the problem that bottles of wine sometimes turn into vinegar, either soon after being produced, or when the bottles are left to age. Data on 50 bottles of wine is available. For each bottle, the age, x, is recorded (in years), along with an indicator, y, of whether the bottle has turned into vinegar ($+1$ if it has, -1 if it hasn't). The vintner decides to use a logistic regression model for this data, as follows:

$$P(y_i \,|\, x_i) \;\; = \;\; [1 + \exp(-y_i(\beta_0 + \beta_1 x_i))]^{-1}$$

The vintner knows that while wine can turn into vinegar, vinegar never turns back into wine. He therefore believes that the β_1 cannot possibly be negative. He also believes that wine has only a small chance of turning into vinegar immediately, so that β_0 should have a fairly large negative value.

The vintner formalizes these and other beliefs in terms of a prior distribution for the parameters in which β_0 and β_1 are independent, with β_0 having a normal distribution with mean -5 and variance 1, and β_1 having an exponential distribution with mean $1/2$.

The vinter is interested in the posterior distribution of β_0 and β_1 given the 50 observed data points, which can be obtained from `/u/radford/ass4-data`. In particular, the vintner would like to know the posterior means and standard deviations for β_0 and β_1.

For this assignment, you should use numerical quadrature to solve this problem. You should use the mid-point rule, for which you can use the procedures for doing two-dimensional integrals with the mid-point rule that are available from the course web page. Since the ranges of the parameters are $(-\infty, +\infty)$ and $(0, +\infty)$, you will have to transform the parameters to a finite interval such as $(0, 1)$ in order to solve the problem. You should use the transformations $1/(1 + \exp(-\beta_0))$ and $1 - \exp(-\beta_1)$ for this. You should write the program to use a specified number of points for the integrations, but you should obtain results with varying numbers of points in order to judge how accurate the results are.

You should hand in your program, the results you obtain on the data, and a brief discussion of how these results compare with those obtained by fitting the logistic regression model by maximum likelihood, using `glm`.