Assignment #3

Due in class on november 24. Worth 16% of the final mark.

Suppose that a vintner is interested in the amount of sugar in a particular kind of grape juice, and in the amount of sugar in the wine produced when this grape juice is fermented in a certain way. Four measurements are made of the amount of sugar in the grape juice, and four measurements are also made of the amount of sugar in the wine produced. On the basis of these measurements, the vintner wishes to estimate the amount of sugar in the grape juice (θ_1) and in the wine (θ_2).

Suppose that the vintner knows that the amount of sugar in juice and in wine must be between 0 and 10 (in appropriate units). The vinter also knows that fermentation can only reduce the sugar content, not increase it, so the amount of sugar in the wine cannot be greater than the amount of sugar in the juice (ie, $\theta_2 \leq \theta_1$). The vinter doesn't know much else about what these values might be, and therefore choses to use a prior distribution for θ_1 and θ_2 that is jointly uniform over the two-dimensional region allowed by these constraints.

The measurements of sugar concentration are made with a device that is known to produce normally-distributed measurements with the true value as the mean and with a standard deviation of one. The errors in the eight measurements are all independent.

Calculate the posterior means of θ_1 and θ_2 based on the prior above and the following data:

Measurements of sugar in juice: $6.1, 7.2, 6.9, 5.4$
Measurements of sugar in wine: $6.2, 5.7, 6.4, 5.2$

Compare these posterior means with the sample means of the data above. Why do they differ as they do?

You should do the two-dimensional integrals required for this problem using nested applications of the trapezoidal rule. (You can try Simpson's rule as well if you like, but this isn't required.) To keep down the effort on this assignment, you don't need to have your program automatically figure out how many points are needed to achieve sufficient accuracy — instead you can just use some fixed number of points for all the applications of the trapezoidal rule. However, you should check to see how much the answer changes when you use half that number of points, to get some idea of the error (and adjust the number upward if the error is large).

What problems do you think would arise if you tried to use your program to solve the same problem, but with many more than four measurements for the juice and for the wine?