Exercise due Tuesday, 10/16*

A LOGIT MODEL

Our data consist of an indicator variable y_i that, for each individual i is 1 if that person rides the subway and 0 if not, plus the distance d_i from that person's residence to the nearest subway station, in miles. We postulate the model

$$P[y_i = 1 \mid c, \gamma, d_i] = \frac{e^{c+\gamma/d_i}}{1 + e^{c+\gamma/d_i}} .$$

The distances are available only grouped by tenths of a mile, and they are, together with the number sampled at each distance and the number of those who ride the subway,

distance	Σy	number sampled
.1	6	10
.2	8	15
.3	5	11
.4	5	13
.5	4	20
.6	1	5
.7	0	1
.8	2	7

(a) Assuming the choices of the 82 people in the sample are independent (so the sample pdf is the product of their individual pdf's), find an expression for the likelihood function for this sample.

(b) Display a contour plot of the likelihood as a function of c and γ. [If my own calculations are right, using (-3,0) as the range for c and (0,4) as the range for γ should work. It is probably best to calculate the log likelihood, and then before exponentiating it, subtract off the maximum value among those you calculate. Otherwise the log likelihood is likely to emerge as, e.g., all below -60, and exponentiating it can cause numerical problems.]

(c) Plot the marginal pdf's of c and γ under a flat prior.

(d) The population of the city is now uniformly distributed over the distances $.1, \ldots, .8$ from subway stops. The city contemplates building new subway stops so that instead everyone is uniformly distributed over the distances .1, .2, .3., .4. Plot the pdf and show a minimum-length 95% probability interval for the proportional increase in subway ridership the new stops will produce. Here again use a flat prior. [Note that what you are asked to compute is a function of c and γ alone (not the sample data), so you can calculate it from the likelihood (which is also the posterior pdf under a flat prior).]