Likelihood 2011, Homework 5

Logistic regression—The excel file "CAD_data.xls" contains data from the logistic regression homework assignment from Biometry I. To the right of the data is the analysis conducted in Minitab v14. The maximum likelihood estimates for the parameters of the linear model are a = -5.309 (the intercept) and b = 0.111 (the slope). The likelihood ratio tests statistic for the null hypothesis that the slope = 0 is reported as G = 29.31.

Part I.
Determine the log-likelihood function for the entire dataset under the logistic regression model. Place this function into the python template and use the built in numerical optimizer to find the ML for the data under both the null hypothesis (b = 0) and the alternative hypothesis (b not constrained). If you succeed your output should match the output from Minitab (MLE: a = -5.309, b = 0.111; LRT = 29.31).

Part II.
The Minitab output includes a standard error for the intercept (1.13) and the slope (0.024). These values were calculated from the information matrix and their validity is based on asymptotic properties of maximum likelihood. A direct and perhaps more robust method for interval estimation is the parametric bootstrap. In homework 4, you used this procedure to establish the critical value for a likelihood ratio test. Then you were simulating data under the null hypothesis. Here, you want to simulate data using the unconstrained maximum likelihood parameter estimates (alternative model). From each replicate, obtain MLE for parameters and then record these to a file. The distribution of estimates (across bootstrap replicates) is your estimate for the sampling distribution of the estimators (the standard error is the standard deviation of the sampling distribution).