CSE383M and CS395T Surprise Quiz

April 1, 2013

This test is closed notes, books, and computers.

1. (Only one problem!) You have N data points of the usual form (x_i, y_i, σ_i) for $i = 1, \ldots, N$, where x_i are the independent variables, y_i are the measured values, and σ_i are the measurement errors (assumed to be normally distributed).

Your model is an arbitrary linear combination of three known functions, $f_1(x)$, $f_2(x)$, and $f_3(x)$. That is,

$$y = a_1 f_1(x) + a_2 f_2(x) + a_3 f_3(x) + e$$

where $e \sim N(0, \sigma)$.

(a) (10 points) Write an explicit expression for χ^2, given the data. It should be a explicit function of the three unknown parameters (a_1, a_2, a_3).

(b) (15 points) Suppose that an oracle told you the true population values of (a_1, a_2, a_3). For these fixed true values, if you drew many data sets (a frequentist concept!) each of size N, how would you expect the resulting values of χ^2 (your answer in part (a)) to be distributed? If your answer has symbols like ν or words like "degrees of freedom," say explicitly what is their value.

(c) (25 points) Write down a set of linear equations that can be solved to find maximum likelihood values for the parameters (a_1, a_2, a_3). In particular, write explicit expressions for the matrix coefficients M_{jk} and the right-hand side vector B_j so that the equations have the form

$$\sum_{k=1}^{3} M_{jk} a_k = B_j, \quad j = 1, \ldots, 3$$

Quiz 4/1/2013

Bill's Solution Set

1.

10 (a) $\chi^2(a_1, a_2, a_3) =$

$$\sum_{i=1}^{N} \left(\frac{y_i - a_1 f_1(x_i) - a_2 f_2(x_i) - a_3 f_3(x_i)}{\sigma_i} \right)^2$$

15 (b) It is the sum of N independent t-values, squared, so it is distributed as Chisquare(N). This has some explicit p.d.f. that you are not required to memorize.

25 (c) $y = \sum_k a_k \underbrace{f_k(x_i)}_{\text{call this } f_{ki}}$

$$\chi^2 = \sum_i \left(\frac{y_i - \sum_k a_k f_{ki}}{\sigma_i} \right)^2$$

for a minimum,

$$0 = \frac{\partial \chi^2}{\partial a_j} = 2 \sum_i \left(\frac{y_i - \sum_k a_k f_{ki}}{\sigma_i} \right) \left(- \frac{f_{ji}}{\sigma_i} \right)$$

so $0 = \displaystyle\sum_i \frac{y_i f_{ji}}{\sigma_i^2} - \sum_k \sum_i \frac{f_{ki} f_{ji}}{\sigma_i^2} a_k$

and $\displaystyle\sum_k \underbrace{\left(\sum_i \frac{f_{ki} f_{ji}}{\sigma_i^2} \right)}_{\equiv M_{jk}} a_k = \underbrace{\left(\sum_i \frac{f_{ji} y_i}{\sigma_i^2} \right)}_{\equiv B_j}$