STA 2201s06 Assignment 2

Do this assignment in preparation for Quiz Two on Thursday Jan. 26th. The hand-written parts are preparation for the quiz, and are not to be handed in. Question 4 asks for calculations with R. For this question, please bring your printout to the quiz. It may be handed in.

1. Let X_1 be Normal(μ_1, σ_1^2), and X_2 be Normal(μ_2, σ_2^2), independent of X_1. What is the joint distribution of $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$? What is required for Y_1 and Y_2 to be independent?

2. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ random vectors, and let $\boldsymbol{\Sigma}$ be fixed and *known*. Derive the maximum likelihood estimate of $\boldsymbol{\mu}$ without differentiating. Where do you use the fact that $\boldsymbol{\Sigma}^{-1}$ is positive definite? Indicate this clearly.

3. Let $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{X}$ is an $n \times p$ matrix of known constants, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown constants, and $\boldsymbol{\epsilon}$ is multivariate normal with mean zero and covariance matrix $\sigma^2 \mathbf{I}_n$, with $\sigma^2 > 0$ an unknown constant. This is classical multiple regression.

   (a) What is the distribution of $\mathbf{Y}$?

   (b) The maximum likelihood estimate of $\mathbf{X}$ is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y$. You could get this using the same approach as in Problem 2, but don't bother. What is the distribution of $\hat{\boldsymbol{\beta}}$? Show the calculations.

   (c) The vector of predicted Y values is $\widehat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$. What is the distribution of $\widehat{\mathbf{Y}}$? Show the calculations.

   (d) Let the vector of residuals $\mathbf{e} = (\mathbf{Y} - \widehat{\mathbf{Y}})$. What is the distribution of $\mathbf{e}$? Show the calculations; simplify.

4. Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with $\mu = \sigma^2 = \theta > 0$. Using R's `nlm` function, find the MLE of θ for the data in `normsamp.dat` (see link on course web page). Your final answer is a single number. Bring a printout listing your program and illustrating the run on `normsamp.dat`. On your printout, please circle the MLE.

   By the way, an explicit formula for $\widehat{\theta}$ is possible here. I used it to check my numerical answer; you may do the same if you wish.

Note: In this course, the computer assignments are *not* group projects. Please do them yourself, though of course you may discuss general principles with anybody.