STA 431s09 Assignment 4

Do this assignment in preparation for the quiz on Friday, Feb. 6th. Answers are practice for the the quiz, and are not to be handed in. You should bring a calculator to the quiz.

1. This will be useful later. Show that for any collection of numbers $x_1, \ldots, x_n$,
$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2.$$

2. Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2.
   (a) What is the parameter θ?
   (b) What is the parameter space Θ?
   (c) For $H_0 : \mu = 0$, what is the restricted parameter space Θ_0?
   (d) For $H_0 : \sigma^2 = \sigma_0^2$, what is the restricted parameter space Θ_0?
   (e) For $H_0 : \sigma^2 = \sigma_0^2$ **and** $\mu = \mu_0$, what is the restricted parameter space Θ_0?

3. Again, let $X_1, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2.
   (a) Find the two-dimensional MLE $\widehat{\theta}$. Show all your work.
   (b) Write $-2 \ln L(\widehat{\theta})$ and simplify. Show your work.
   (c) For $H_0 : \mu = 0$, find the restricted MLE $\widehat{\theta}_0$. Show your work. Remember that $\widehat{\theta}_0$ consists of two quantities.
   (d) Write $-2 \ln L(\widehat{\theta}_0)$ and simplify. Show your work.
   (e) Write and simplify a formula for the large-sample likelihood ratio test statistic
   $$G = -2 \ln \frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})}.$$
   (f) What are the degrees of freedom for this test? Your answer is a single number.
   (g) For a sample of size $n = 100$, we obtain $\sum_{i=1}^{n} x_i = 1.59$ and $\sum_{i=1}^{n} x_i^2 = 108.4265$. What is the numerical value of G? Your answer is a single number.
   (h) The critical value of chisquare at $\alpha = 0.05$ is
   ```
   > qchisq(0.95,df=1)
   [1] 3.841459
   ```
   Do you reject H_0? Answer Yes or No.

4. We continue working with the same random sample: $n = 100$, $\sum_{i=1}^{n} x_i = 1.59$ and $\sum_{i=1}^{n} x_i^2 = 108.4265$.

   (a) For $H_0 : \sigma^2 = 1.5$, find the restricted MLE $\widehat{\theta}_0$. Show your work. Remember that $\widehat{\theta}_0$ consists of two quantities.

   (b) Write $-2 \ln L(\widehat{\theta}_0)$ and simplify. Show your work.

   (c) Write and simplify a formula for the large-sample likelihood ratio test statistic

$$G = -2 \ln \frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})}.$$

   (d) What are the degrees of freedom for this test? Your answer is a single number.

   (e) What is the numerical value of G? Your answer is a single number.

   (f) Do you reject H_0? Answer Yes or No.

   (g) Is σ^2 equal to 1.5, greater than 1.5 or less than 1.5? Make your best guess.

5. We will now test whether these data come from a standard normal distribution.

   (a) For $H_0 : \mu = 0$ and $\sigma^2 = 1$, give the restricted MLE $\widehat{\theta}_0$. Remember that $\widehat{\theta}_0$ consists of two quantities.

   (b) Write $-2 \ln L(\widehat{\theta}_0)$ and simplify. Show your work.

   (c) Write and simplify a formula for the large-sample likelihood ratio test statistic

$$G = -2 \ln \frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})}.$$

   (d) What are the degrees of freedom for this test? Your answer is a single number.

   (e) For a sample of size What is the numerical value of G? Your answer is a single number.

   (f) Do you reject H_0? Answer Yes or No.

   (g) Do the data come from a standard normal distribution? If you had to guess, would you answer Yes or No?

6. The k-dimensional multivariate normal density is

$$f(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}(2\pi)^{\frac{k}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right],$$

where $\boldsymbol{\Sigma}$ is symmetric and positive definite.

The idea here is that the data vectors represent k pieces of information from each of n individuals — say, tastiness ratings of k potato chip brands. Are these ratings related to each other? For example, if a participant in the study gives high ratings to Brand One, will she also tend to give high ratings to Brand Three? But we want a test for all the correlations (or equivalently, the covariances) at the same time. So, the null hypothesis of interest is that $\boldsymbol{\Sigma}$ is diagonal; *all the covariances are zero.*

(a) You already know that the unrestricted MLE is $\widehat{\theta} = (\overline{\mathbf{x}}, \widehat{\boldsymbol{\Sigma}})$, where

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i-\overline{\mathbf{x}})(\mathbf{x}_i-\overline{\mathbf{x}})'.$$

Write $-2\ln L(\widehat{\theta})$ and simplify. Show your work.

(b) Write the log likelihood when H_0 is true; simplify. Because zero covariance implies independence for the multivariate normal, the joint density is a product of marginals. Hint: There are nk data values, so your answer will involve $x_{i,j}$ for $i = 1,\ldots,n$ and $j = 1,\ldots,k$.

(c) Now start differentiating with respect to μ_1. Stop as soon as you see the pattern. Is $\widehat{\boldsymbol{\mu}}_0$ still $\overline{\mathbf{x}}$? Answer Yes or No.

(d) What is $\widehat{\boldsymbol{\Sigma}}_0$? If you think about what happened on the preceding part of this question, you can confidently write down the answer without doing any calculations. Or, you can start differentiating with respect to σ_1^2 and continue until you see what is happening.

(e) Write $-2\ln L(\widehat{\theta}_0)$ and simplify. Show your work.

(f) Now write down the large-sample likelihood test statistic G and simplify. What are the degrees of freedom (a function of k)? You have derived something useful.

(g) Here are some calculations on the $n = 200$ SAT data from Assignment Three.

```
> sat <- read.table("sat.data")
> dim(sat)
[1] 200   3
> # So we see n = 200
> SigmaHat <- (n-1) * var(sat) / n  # The unrestricted MLE of Sigma
> print(SigmaHat)

           verbal        math        gpa
verbal 5306.08907 1320.630648 13.5625025
math   1320.63065 4357.919406  7.4042050
gpa      13.56250    7.404205  0.3334161

> prod(diag(SigmaHat))
[1] 7709750
> det(SigmaHat) # Determinant
[1] 6300991
> qchisq(0.95,df=3)  # Critical value
[1] 7.814728
```

Calculate the test statistic G. Your answer is a single number.

(h) Do you reject H_0? Answer Yes or No.

(i) Are the three variables all independent of one another? Answer Yes or No.