STA 313F2004 Assignment 2

Do this assignment in preparation for the quiz on Friday, Oct. 8. The hand-written parts are practice for the quiz, and are not to be handed in. The computer parts may be handed in, so please begin each computer question on a separate page.

1. Let $Y_1, \ldots, Y_n$ be a random sample from a distribution with density $f(y) = \frac{1}{\theta} e^{-\frac{y}{\theta}}$ for $y > 0$, where the parameter $\theta > 0$. We are interested in testing $H_0 : \theta = \theta_0$.

   (a) What is Θ?

   (b) What is Θ_0?

   (c) What is Θ_1?

   (d) Derive a general expression for the large-sample likelihood ratio statistic $G = -2 \log \frac{\ell(\widehat{\widehat{\theta}})}{\ell(\widehat{\theta})}$.

   (e) A sample of size $n = 100$ yields $\overline{Y} = 1.37$ and $S^2 = 1.42$. One of these quantities is unnecessary and just provided to irritate you. Well, actually it's a mild substitute for reality, which always provides you with a huge pile of information you don't need. Anyway, we want to test $H_0 : \theta = 1$. You can do this with a calculator. When I did it a long time ago I got $G = 11.038$.

   (f) At $\alpha = 0.05$, the critical value of chisquare with one degree of freedom is 3.841459. Do you reject H_0? Answer Yes or No.

2. The label on the peanut butter jar says peanuts, partially hydrogenated peanut oil, salt and sugar. But we all know there is other stuff in there too. In the United States, the Food and Drug administration requires that a shipment of peanut butter be rejected if it contains an average of more than 8 rat hairs per pound (well, I'm not sure if it's exactly 8, but let's pretend). There is very good reason to assume that the number of rat hairs per pound has a Poisson distribution with mean λ (take STA347). We will test $H_0 : \lambda = \lambda_0$.

   (a) What is Θ?

   (b) What is Θ_0?

   (c) What is Θ_1?

   (d) Derive a general expression for the large-sample likelihood ratio statistic.

   (e) We sample 100 1-pound jars, and observe a sample mean of $\overline{Y} = 8.57$. Should we reject the shipment? We want to test $H_0 : \lambda = 8$. What is the value of G? You can do this with a calculator. When I did it a long time ago I got $G = 3.97$.

   (f) Do you reject H_0 at $\alpha = 0.05$? Answer Yes or No.

   (g) Do you reject the shipment of peanut butter? Answer Yes or No.

3. The normal distribution has density

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}.$$

   Find an explicit formula for the MLE of $\theta = (\mu, \sigma^2)$. This example is in every STA257-level textbook.

4. Write an S function that performs a large-sample likelihood ratio test of $H_0 : \sigma^2 = \sigma_0^2$ for data from a single normal random sample. The function should take the sample data and σ_0^2 as input, and return 3 values: G, the degrees of freedom, and the p-value. Run your function on the data in `var.dat`, testing $H_0 : \sigma^2 = 2$; see link to the data on the course web page.

   For this question, you need to bring a printout with a listing of your function (showing how it is defined), and also part of an R session showing execution of the function, and the resulting output.

5. For k samples from independent normal distributions, the usual one-way analysis of variance tests equality of means assuming equal variances. Now you will construct a large-sample likelihood ratio test for equality of means, except that you will *not* assume equal variances. Write an S function to do it.

   Input to the function should be the sample data, in the form of a matrix. The first column should contain group membership (the independent variable). It is okay to assume that the unique values in this column are the integers from 1 to k. The second column should contain values of the normal random variates – the dependent variable.

   The function should return 3 values: G, the degrees of freedom, and the p-value. Run your function on the sample in `kars.dat`; see link to the data on the course web page. This data set shows country of origin and gas mileage for a sample of automobiles.

   For this question, you need to bring a printout with a listing of your function (showing how it is defined), and also part of an R session showing execution of the function, and the resulting output.

Read Chapter 1 in the Structural Equation Models text; see link on the course web page. This is the only part of the text you should print at this point. Observe that x, y and u are random variables and not constants, even though they are not capitalized. You are responsible for the notation in Chapter 1. Please do the following problems.

6. For the model $y = bx + u$ of Chapter 1, find
   (a) σ_{yy}
   (b) σ_{xy}
   (c) σ_{uy}

7. Show $b_{yx} = \frac{\sigma_{xy}}{\sigma_{xx}}$

8. Do the exercise on P. 8. It has 3 parts.

9. What is $E(xv)$?

10. Now let $y_i = bx_i + u_i$, for $i = 1, \ldots, n$; let $x_i \sim N(0, \sigma_{xx})$ and $u_i \sim N(0, \sigma_{uu})$ with x_i and u_i independent. All the random variables are independent for different values of i. We observe pairs (x_i, y_i) for $i = 1, \ldots, n$.

    (a) Write -2 times the log likelihood and simplify. There are three parameters. You may find it easiest to write

    $$\ell(\theta) = \Pi_{i=1}^{n} f_\theta(y_i|x_i) f_\theta(x_i).$$

    (b) Differentiate to find the MLE of θ. Don't forget the second derivative test.

    (c) Give an expression for the large-sample likelihood ratio chisquare statistic G for testing $H_0 : b = 0$. Simplify as much as possible.

    (d) Examining your last answer, would it matter if $x_1, \ldots, x_n$ were constants instead of normal random variables? What if they were random variables but not normal?