STA 431s23 Assignment Four[1]

For the Quiz on Friday Feb. 10th, please bring printouts of your full R input for Questions 3 and 4. The other problems are not to be handed in. They are practice for the Quiz.

---

1. Independently for $i = 1, \ldots, n$, let $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_1$, where $E(x_{i1}) = \mu_{x1}, E(x_{i2}) = \mu_{x2}, Var(x_{i1}) = \phi_{11}, Var(x_{i2}) = \phi_{22}, Cov(x_{i1}, x_{i2}) = \phi_{12}, Var(\epsilon_i) = \psi$, and ϵ_i is independent of x_{i1} and x_{i2}.
   (a) What is the parameter θ for this model?
   (b) What is the parameter space Θ?
   (c) What is the restricted parameter space Θ_0 under $H_0 : \beta_1 = \beta_2$ and $\phi_{11} = \phi_{22} = \psi = 1$?
   (d) The null hypothesis can be written $H_0 : \mathbf{L}\boldsymbol{\theta} = \mathbf{h}$. Give the $\mathbf{L}$ and $\mathbf{h}$ matrices

2. On the formula sheet, the statement $\widehat{\boldsymbol{\theta}}_n \stackrel{.}{\sim} N(\boldsymbol{\theta}, \mathbf{V}_n)$ is saying that the vector of maximum likelihood estimates is asymptotically normal. That is, for large sample sizes, $\widehat{\boldsymbol{\theta}}_n$ has a probability distribution that is approximately multivariate normal, centered on the vector of true parameter values and with variance-covariance matrix $\mathbf{V}_n$. Generally speaking, the rules (theorems) for exact multivariate normality also apply to asymptotic multivariate normality. It's not rigorous, but you usually arrive at the correct conclusion. Accordingly,
   (a) Let $\boldsymbol{\theta}$ be $m \times 1$, and let $\mathbf{L}$ be an $r \times m$ matrix of constants with linearly independent rows. This part of the question develops the Wald statistic for testing $H_0 : \mathbf{L}\boldsymbol{\theta} = \mathbf{h}$.
      i. What is the asymptotic distribution of $\mathbf{L}\widehat{\boldsymbol{\theta}}_n$? Just write it down.
      ii. What is the asymptotic distribution of $(\mathbf{L}\widehat{\boldsymbol{\theta}}_n - \mathbf{L}\boldsymbol{\theta})^\top \left(\mathbf{L}\mathbf{V}_n\mathbf{L}^\top\right)^{-1} (\mathbf{L}\widehat{\boldsymbol{\theta}}_n - \mathbf{L}\boldsymbol{\theta})$?
      iii. What are the dimensions (number of rows and columns) in the matrix $\left(\mathbf{L}\mathbf{V}_n\mathbf{L}^\top\right)^{-1}$?
      iv. Why is it critical that the rows of $\mathbf{L}$ be linearly independent, so that the rank of $\mathbf{L}$ equals r?
      v. Compare the expression in Question 2(a)ii to the Wald statistic W_n on the formula sheet. There are two differences. Briefly explain them.
   (b) Let $\mathbf{a}$ be an $m \times 1$ non-zero vector of constants. What is the asymptotic distribution of $\mathbf{a}^\top\widehat{\boldsymbol{\theta}}_n$?
   (c) Based on the last result, give a $(1 - \alpha)100\%$ confidence interval for $\mathbf{a}^\top\boldsymbol{\theta}$. Use $z_{\alpha/2}$ to denote the value that cuts off the top $\alpha/2$ of the standard normal distribution (For example for $\alpha = 0.05$, $z_{\alpha/2} = 1.96$). Show some work. Why are you using $\widehat{\mathbf{V}}_n$ instead of $\mathbf{V}_n$?
   (d) Using the same standard error, write down a z statistic for testing $H_0 : \mathbf{a}^\top\boldsymbol{\theta} = h$.
   (e) Show that for $H_0 : \mathbf{a}^\top\boldsymbol{\theta} = h$, $W_n = z^2$.

---

[1]This assignment was prepared by Jerry Brunner, Department of Statistical Sciences, University of Toronto. It is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License. Use any part of it as you like and share the result freely. The LaTeX source code is available from the course website: http://www.utstat.toronto.edu/brunner/oldclass/431s23

3. Let $x_1, \ldots, x_n$ be a random sample from a beta distribution. The density is

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \, x^{\alpha-1} \, (1-x)^{\beta-1}$$

for $0 < x < 1$, where α and β are both greater than zero. Numerical data are available HERE. You can get a copy of the data with

```
x = scan("https://www.utstat.toronto.edu/brunner/openSEM/data/beta24.data.txt")
```

(a) Find the maximum likelihood estimates of α and β.

(b) Test $H_0 : \beta = 2\alpha$ with a large-sample likelihood ratio test. Your output should include the G^2 statistic, the degrees of freedom, and the p-value. Is the null hypothesis rejected at the 0.05 significance level? What, if anything, do you conclude?

(c) Test the same null hypothesis, this time with a Wald test. Your output should include the W_n statistic, the degrees of freedom, and the p-value. Is the null hypothesis rejected at the 0.05 significance level? What, if anything, do you conclude?

(d) Give a 95% confidence interval for the quantity $2\alpha - \beta$. Your answer is a set of two numbers, the lower confidence limit and the upper confidence limit. Hint: Does this remind you of Problem 2c?

Please bring a printout of your full R input and output to the quiz.

---

4. Independently for $i = 1, \ldots, n$, let $y_i = \beta x_i + \epsilon_i$, where $x_i \sim N(\mu_x, \sigma^2_x)$, $\epsilon_i \sim N(0, \sigma^2_\epsilon)$, and x_i and ϵ_i are independent. This is the model of Question 14 in Assignment 3.

(a) Use R to simulate a data set from this model. The true parameter values and the sample size are up to you, but the sample size should be large.

(b) In Question 14 of Assignment 3, you found two method of moments estimators for β. They were

$$\widehat{\beta}_1 = \frac{\overline{y}_n}{\overline{x}_n} \text{ and } \widehat{\beta}_2 = \frac{\sum_{i=1}^n (x_i - \overline{x}_n)(y_i - \overline{y}_n)}{\sum_{i=1}^n (x_i - \overline{x}_n)^2}$$

Calculate $\widehat{\beta}_1$ and $\widehat{\beta}_2$ for your simulated data. Which estimate comes closer to the truth? Of course you would have to carry out this experiment a large number of times to determine whether one of them is better in general.

Please bring a printout of your full R input and output to the quiz.

5. The usual univariate multiple regression model with independent normal errors is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

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. But of course in practice, the explanatory variables are random, not fixed. Clearly, if the model holds *conditionally* upon the values of the explanatory variables, then all the usual results hold, again conditionally upon the particular values of the explanatory variables. The probabilities (for example, p-values) are conditional probabilities, and the F statistic does not have an F distribution, but a conditional F distribution, given $\boldsymbol{\mathcal{X}} = \mathbf{X}$.

   (a) Show that the least-squares estimator $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ is conditionally unbiased.

   (b) Show that $\widehat{\boldsymbol{\beta}}$ is also unbiased unconditionally.

   (c) A similar calculation applies to the significance level of a hypothesis test. Let F be the test statistic (say for an extra-sum-of-squares F-test), and f_c be the critical value. If the null hypothesis is true, then the test is size α, conditionally upon the explanatory variable values. That is, $P(F > f_c | \boldsymbol{\mathcal{X}} = \mathbf{X}) = \alpha$. Find the *unconditional* probability of a Type I error. Assume that the explanatory variables are discrete, so you can write a multiple sum.

6. The point of this question is that under conditions that are fairly common and natural, MLEs and likelihood ratio tests based on a fixed-x regression model are also valid for a random-x model. For notational convenience, suppose that the model parameter $\theta = (\theta_1, \theta_2)$, and that the joint density/probability mass function of the data can be written

$$f_\theta(x, y) = g_{\theta_1}(y|x)\, h_{\theta_2}(x), where$$

   - $f_\theta(x, y)$ is the joint density of x and y. It depends on the entire parameter vector θ.
   - $g_{\theta_1}(y|x)$ is the conditional density of y given x. It depends on θ_1.
   - $h_{\theta_2}(x)$ is the marginal density of x. It depends on θ_2.

   The quantities x, y, θ_1 and θ_2 could all be vectors. There must be no functional connection between θ_1 and θ_2. For example in a regression, we might have $\theta_1 = (\boldsymbol{\beta}, \sigma^2)$, and if $h_{\theta_2}(x)$ is a multivariate normal density, θ_2 would be the unique elements of $\boldsymbol{\mu}_x$ and $\boldsymbol{\Sigma}_x$. The lack of functional connection between θ_1 and θ_2 just means there are no β_j parameters and no σ^2 in $\boldsymbol{\mu}_x$ or $\boldsymbol{\Sigma}_x$. Usually, we only care about the parameters in θ_1.

   (a) Writing the full likelihood as $L(\theta) = \pi_{i=1}^n f_\theta(x_i, y_i)$, show that $\widehat{\theta_1}$ for the random-x model is the same as for the model of y conditional on x. It's easiest to see if you take the log of the likelihood and start differentiating.

   (b) Now consider a likelihood ratio test that *only* restricts θ_1. In regression, it would be about the β_j parameters. Show that the likelihood ratio test statistic

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

   for the random-x model is the same as for the model of y conditional on x.

Please bring printouts of your full R input and output for Questions 3 and 4 to the quiz.