

## STA 431s15 Assignment Five<sup>1</sup>

The non-computer questions on this assignment are for practice, and will not be handed in. For the SAS part of this assignment (Question 4) please bring your log file and your output file to the quiz. There may be one or more questions about them, and you may be asked to hand the printouts in with the quiz.

1. Men and women are calling a technical support line according to independent Poisson processes with rates  $\lambda_1$  and  $\lambda_2$  per hour. Data for 144 hours are available, but unfortunately the sex of the caller was not recorded. All we have is the number of callers for each hour, which is distributed  $\text{Poisson}(\lambda_1 + \lambda_2)$ .
  - (a) The parameter in this problem is  $\boldsymbol{\theta} = (\lambda_1, \lambda_2)'$ . Try to find the MLE by differentiating. Show your work. Are there any points in the parameter space where both partial derivatives are zero?
  - (b) Is the parameter identifiable? Answer Yes or No and prove your answer. If the answer is no, all you need to prove your answer is a simple numerical example of two different parameter vectors that yield the same probability distribution for the sample data.
  - (c) Give one identifiable *function* of  $\boldsymbol{\theta}$ . Show it is identifiable by expressing it as a function of the probability distribution of the observable data.
2. Independently for  $i = 1, \dots, n$ , let

$$\begin{aligned} Y_i &= \beta X_i + \epsilon_i \\ W_i &= X_i + e_i \end{aligned}$$

where all random variables are normal with expected value zero,  $\text{Var}(X_i) = \phi > 0$ ,  $\text{Var}(\epsilon_i) = \psi > 0$ ,  $\text{Var}(e_i) = \omega > 0$  and  $\epsilon_i$ ,  $e_i$  and  $X_i$  are all independent. The variables  $W_i$  and  $Y_i$  are observable, while  $Y_i$  is latent. Error terms are never observable.

- (a) What is the parameter vector  $\boldsymbol{\theta}$  for this model?
- (b) Calculate the covariance matrix  $\boldsymbol{\Sigma}$  of the observable variables, expressed as a function of the model parameters.
- (c) Does this model pass the test of the parameter count rule? Answer Yes or No and give the numbers.
- (d) Are there any points in the parameter space where the parameter  $\beta$  is identifiable? Are there infinitely many, or just one point?
- (e) The usual estimator of  $\beta$  is

$$\hat{\beta}_n = \frac{\sum_{i=1}^n W_i Y_i}{\sum_{i=1}^n W_i^2}.$$

Is  $\hat{\beta}_n$  a consistent estimator of  $\beta$ ? Answer Yes or No and prove your answer.

---

<sup>1</sup>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 L<sup>A</sup>T<sub>E</sub>X source code is available from the course website: <http://www.utstat.toronto.edu/~brunner/oldclass/431s15>

3. Independently for  $i = 1, \dots, n$ , let

$$\begin{aligned} Y_i &= \beta X_i + \epsilon_i \\ V_i &= Y_i + e_i \\ W_{i,1} &= X_i + e_{i,1} \\ W_{i,2} &= X_i + e_{i,2} \end{aligned}$$

where

- $Y_i$  is a latent variable.
  - $V_i$ ,  $W_{i,1}$  and  $W_{i,2}$  are all observable variables.
  - $X_i$  is a normally distributed *latent* variable with mean zero and variance  $\phi > 0$ .
  - $\epsilon_i$  is normally distributed with mean zero and variance  $\psi > 0$ .
  - $e_i$  is normally distributed with mean zero and variance  $\omega > 0$ .
  - $e_{i,1}$  is normally distributed with mean zero and variance  $\omega_1 > 0$ .
  - $e_{i,2}$  is normally distributed with mean zero and variance  $\omega_2 > 0$ .
  - $X_i$ ,  $\epsilon_i$ ,  $e_i$ ,  $e_{i,1}$  and  $e_{i,2}$  are all independent of one another.
- (a) Make a path diagram of this model.
  - (b) What is the parameter vector  $\theta$  for this model?
  - (c) Does this problem pass the test of the Parameter Count Rule? Answer Yes or No and give the numbers.
  - (d) Calculate the variance-covariance matrix of the observable variables as a function of the model parameters. Show your work.
  - (e) Is the parameter vector identifiable at every point in the parameter space? Answer Yes or No and prove your answer.
  - (f) Some parameters are identifiable, while others are not. Which ones are identifiable?
  - (g) If  $\beta$  (the parameter of main interest) is identifiable, propose a Method of Moments estimator for it and prove that your proposed estimator is consistent.
  - (h) Suppose the sample variance-covariance matrix  $\widehat{\Sigma}$  is

	W1	W2	V
W1	38.53	21.39	19.85
W2	21.39	35.50	19.00
V	19.85	19.00	28.81

Give a reasonable estimate of  $\beta$ . There is more than one right answer. The answer is a number. (Is this the Method of Moments estimate you proposed? It does not have to be.) **Circle your answer.**

4. This is a slight extension of the SAS part of Assignment Four. Again, students in a first-year Calculus class took a diagnostic test with two parts: Pre-calculus and Calculus. Their High School Calculus marks and their marks in University Calculus were also available. In order, the variables in the data file are: Identification code, Mark in High School Calculus, Score on the Pre-calculus portion of the diagnostic test, Score on the Calculus portion of the diagnostic test, and mark in University Calculus<sup>2</sup>. Data are available in the file [mathtest.data.txt](#).
- (a) Using SAS `proc calis`, first carry out an unconditional regression in which the explanatory variables are Mark in High School Calculus, Score on the Pre-calculus portion of the diagnostic test and Score on the Calculus portion of the diagnostic test. The response variable is mark in University Calculus. All the variables are observable. This is your full, unrestricted model, and it is saturated.
  - (b) You want a single test for whether the two parts of the diagnostic test are useful for predicting university calculus mark. Carry out a likelihood ratio test to do this. What is the null hypothesis? It doesn't matter whether your test statistic is  $G^2$  or  $\frac{n-1}{n} G^2$ . Be able to state your conclusions from this test in simple, non-technical language. Your conclusion is something about the usefulness of the diagnostic test. You should be guided by the  $\alpha = 0.05$  significance level, but *do not mention the significance level in your conclusion*. It's too technical for your boss, who will fire you if you start showing off.
  - (c) In addition to the requested likelihood ratio test, your output file has lots of “*t*-tests” (actually, *Z*-tests). Know what null hypotheses they all are testing, but pay attention *only* to the ones based on the unrestricted model. This is a general rule.
  - (d) What null hypothesis is being tested by the “Baseline Model Chi-Square?” Is it the same for the restricted and unrestricted model? Why are the degrees of freedom what they are?

Bring your log file and your list file to the quiz. You may be asked for numbers from your printouts, and you may be asked to hand them in. There are lots of **There must be no error messages, and no notes or warnings about invalid data on your log file.**

---

<sup>2</sup>Thanks to Cleo Boyd for permission to use these original data.