Homework Version 3

1. Men and women are calling a technical support line according to independent Poisson processes with rates λ_1 and λ_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 Poisson($\lambda_1 + \lambda_2$). Here are the data, which are also available in the file `poisson.data` on the Shimmering Portal website:

```
12  9 25 11 22 16 17  8 14 17 14 17  8 14 18 16 13 17 13 11  9 15 18
14 16 17 16 19 13 18 12 12 13 10  8 15 13 11 15 15  6 10 13 13 11 13
 9 15 16  9  5 10  8 18 13 17  7 13 13 17 12 17 14 16  6 12 17 10  9
14 11 19 13 17 15 20 14 10 13 14 17  9 13 14  7 16 16  9 25 10 10  9
17  7 15 12 14 21 14 18 14 12 13 15 12 11 16 14 15 16  8 19 13 17 15
11 18 13 12 11 19 14 16 17 13 13 19 19 11 19 10 12  9 18 11 14  9 14
14 14 13  9 13 18
```

   (a) The parameter in this problem is $\boldsymbol{\theta} = (\lambda_1, \lambda_2)'$. Try to find the MLE analytically. Show your work.

   (b) Now try to find the MLE numerically by minimizing the minus log likelihood with R's `nlm` function. The Hessian is interesting, becuase it's the observed Fisher information in the sample evaluated at the MLE; ask for it. Try two different starting values. What seems to be happening here?

   (c) Try inverting the Hessian to get the asymptotic covariance matrix. Any comments?

   (d) To better understand what happened in the last item, calculate the Fisher information in a single observation from the definition. That is, letting $\ell = \log f(Y; \boldsymbol{\theta})$, calculate the elements of the 2×2 matrix whose (i, j) element is

$$-E\left(\frac{\partial^2 \ell}{\partial\theta_i \partial\theta_j}\right)$$

   (e) The Fisher information in the sample is just n times the Fisher information in a single observation. Using the numerical MLEs from one of your `nlm` runs, estimate this quantity (a 2×2 matrix). Compare it to the Hessian. Now do you see what happened when you tried to calculate the asymptotic covariance matrix?

   (f) What does the (log) likelihood look like geometrically?

Please bring your R input and output as well as the paper and pencil work.

2. Independently for $i = 1, \ldots, n$, let

$$\begin{aligned} Y_i &= \beta X_i + \epsilon_i \\ W_i &= X_i + e_i \end{aligned} \tag{1}$$

where $E(X_i) = E(\epsilon_i) = E(e_i) = 0$, $Var(X_i) = \sigma_x^2$, $Var(\epsilon_i) = \sigma_\epsilon^2$, $Var(e_i) = \sigma_e^2$, and X_i, ϵ_i and e_i are all independent. Notice that W_i is just X_i plus a piece of random noise. This is a simple additive model of measurement error.

Unfortunately, we cannot observe the X_i values. All we can see are the pairs (W_i, Y_i) for $i = 1, \ldots, n$. So we do what everybody does, and fit the *naive* (mis-specified, wrong) model

$$Y_i = W_i \beta + \epsilon_i$$

and estimate β with

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

(a) The file `wy.data` on the Shimmering Portal website has a set of data generated from the true model. Calculate $\widehat{\beta}_n$ based on these data. Your answer is a single number. Assuming the naive model with a normal distribution for ϵ_i, calculate a 95% confidence interval for β. Don't forget that W_i is a random variable, and not a fixed constant, so this does require some thought. Your answer is a pair of real numbers.

(b) Where does $\widehat{\beta}_n$ go as $n \to \infty$? Show your work.

(c) But $\widehat{\beta}_n$ was based on an incorrect model. The correct model is (1). For the correct model, assume in addition that X_i, ϵ_i and e_i are normally distributed.

   i. What is the distribution of the observable data? Express the parameters of the distribution in terms of the (Greek letter) constants in Model (1)

   ii. So, the distribution of the observable data depends upon a vector of four parameters. Are the parameters identifiable? Answer Yes or No, and give a *complete* proof of your answer. The answer should start with the definition of identifiablility that you are using.

   iii. For what set of values is the parameter β identifiable?

   iv. Write each of the other parameters in terms of σ_x^2 and the three covariances σ_{11}, σ_{12} and σ_{22}, in such a way that the covariance matrix

$$V \begin{pmatrix} W_i \\ Y_i \end{pmatrix} = \mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

   remains constant as σ_x^2 varies.

   v. For fixed σ_{11}, σ_{12} and σ_{22}, over what interval can σ_x^2 range?

   vi. Can the data provide any information at all about the parameter? Give details.

3. Here is a model for measurement error just in the response (dependent) variable. Independently for $i = 1, \ldots, n$, let

$$\begin{aligned} Y_i &= \beta X_i + \epsilon_i \qquad (3) \\ V_i &= Y_i + e_i \end{aligned}$$

where $E(X_i) = E(\epsilon_i) = E(e_i) = 0$, $Var(X_i) = \sigma_x^2$, $Var(\epsilon_i) = \sigma_\epsilon^2$, $Var(e_i) = \sigma_e^2$, and X_i, ϵ_i and e_i are all independent. Notice that V_i is just Y_i plus a piece of random noise.

This time, we cannot observe the Y_i values. All we can see are the pairs (X_i, V_i) for $i = 1, \ldots, n$. So again, we fit the *naive* (mis-specified, wrong) model

$$V_i = X_i\beta + \epsilon_i$$

and estimate β with

$$\widehat{\beta}_n = \frac{\sum_{i=1}^n X_i V_i}{\sum_{i=1}^n X_i^2}. \qquad (4)$$

(a) Is the estimator (4) a consistent estimator of β? Answer Yes or No and show the calculation.

(b) What is the parameter vector for this problem? (Careful now!)

(c) What is the covariance matrix of the observable data pairs? Express the your answer in terms of the (Greek letter) constants in Model (3)

(d) So, the covariance matrix of the observable data depends upon a vector of four parameters. Are the parameters identifiable from the covariance matrix? Answer Yes or No, and prove your answer.

(e) For what set of values is the parameter β identifiable?

(f) Would you trust the usual tests and confidence intervals based on the naive model? Answer Yes or No, and explain.

4. In this example, there are two explanatory variables measured with error. Independently for $i = 1, \ldots, n$,

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \epsilon_i \\ W_{i,1} &= X_{i,1} + e_{i,1} \\ W_{i,2} &= X_{i,2} + e_{i,2}, \end{aligned}$$

where where $E(X_{i,1}) = \mu_1$, $E(X_{i,2}) = \mu_2$, $E(\epsilon_i) = E(e_{i,1}) = E(e_{i,2}) = 0$, $Var(\epsilon_i) = \sigma^2$, $Var(e_{i,1}) = \omega_1$, $Var(e_{i,2}) = \omega_2$, the errors $\epsilon_i, e_{i,1}$ and $e_{i,2}$ are all independent, $X_{i,1}$ is independent of $\epsilon_i, e_{i,1}$ and $e_{i,2}$, $X_{i,2}$ is independent of $\epsilon_i, e_{i,1}$ and $e_{i,2}$, and

$$Var \begin{bmatrix} X_{i,1} \\ X_{i,1} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{bmatrix}.$$

Again, because the actual explanatory varibles $X_{i,1}$ and $X_{i,2}$ are latent variables that cannot be observed, $W_{i,1}$ and $W_{i,2}$ are used in their place. The data analyst fits the naive model

$$Y_i = \beta_0 + \beta_1 W_{i,1} + \beta_2 W_{i,2} + \epsilon_i.$$

The usual least-squares estimator of β_1 based on the naive model is

$$\widehat{\beta}_1 = \frac{\left(\sum_{i=1}^n (W_{i,2}-\overline{W}_2)^2\right)\left(\sum_{i=1}^n (W_{i,1}-\overline{W}_1)(Y_i-\overline{Y})\right) - \left(\sum_{i=1}^n (W_{i,1}-\overline{W}_1)(W_{i,2}-\overline{W}_2)\right)\left(\sum_{i=1}^n (W_{i,2}-\overline{W}_2)(Y_i-\overline{Y})\right)}{\left(\sum_{i=1}^n (W_{i,1}-\overline{W}_1)^2 \sum_{i=1}^n (W_{i,2}-\overline{W}_2)^2\right) - \left(\sum_{i=1}^n (W_{i,1}-\overline{W}_1)(W_{i,2}-\overline{W}_2)\right)^2}$$

Let

$$\mathbf{D}_i = \left(\begin{array}{c} W_{i,1} \\ W_{i,2} \\ Y_i \end{array} \right).$$

(a) Calculate the mean of $\mathbf{D}_i$

(b) Calculate the variance-covariance matrix of $\mathbf{D}_i$.

(c) Using the fact that sample variances are strongly consistent estimators of the corresponding population quantities, find where $\widehat{\beta}_1$ goes as $n \to \infty$. *Simplify!* Is $\widehat{\beta}_1$ consistent for β_1?

(d) Would you trust the usual tests and confidence intervals for β_1 based on the naive model? Answer Yes or No, and explain.

(e) What is the parameter vector for this problem?

(f) Is β_1 identifiable from the covariance matrix of the observable data? Answer Yes or No and try to prove it.

Notice that we are doing the same trick again and again. We estimate the parameters of a mis-specified model; the estimator is a function of the sample moments of the observable data. Then we calculate the population moments under the true model. Using the fact that sample moments converge to population moments, we see what happens to the naive estimators as $n \to \infty$ when the true model holds. Sometimes it's okay, and sometimes not