ECOM40006/90013 Econometrics 3

Week 9 Extras (Part 2)

Question 1: The Cramér-Rao lower bound

One of the standard results in maximum likelihood is that of the Cramér-Rao lower bound, or CRLB for short. A good place to start would be the simple univariate case – especially since there's not much point in dealing with general cases unless we understand the simple stuff first! In the univariate case, the Cramér-Rao lower bound can be stated as follows:

Theorem (Cramér-Rao lower bound). *Let $y_1, \dots, y_n$ be n i.i.d. draws from a distribution with pdf $f(y;\theta)$, where θ is the quantity to be estimated. Consider an unbiased estimator $\tilde{\theta}$ of θ. Then the variance of $\tilde{\theta}$ is bounded below as follows:*

$$\text{Var}(\tilde{\theta}) \geq I(\theta)^{-1}.$$

In this question, we'll be proving this theorem in a fashion similar to Silvey (1970).[1]

Getting started. Before we handle this proof, it wouldn't hurt to also deal with some intermediate results.

(a) Result 1. Show that

$$\frac{\partial f(y;\theta)}{\partial\theta} = \frac{\partial \log f(y;\theta)}{\partial\theta} f(y;\theta).$$

(Hint: find something that you can use the Chain Rule on, then rearrange).

(b) Result 2. Consider two random variables X and Y. Show that the squared correlation between the two RVs is always less than or equal to 1:

$$\rho^2 = \frac{[\text{cov}(X,Y)]^2}{\text{Var}(X)\text{Var}(Y)} \leq 1.$$

Note that this also implies a property of correlations that you might be familiar with: correlations are always between -1 and 1. The steps to this are a bit tricky. A rough workflow proceeds as follows:

- Define $V = X - \mathbb{E}(X)$ and $W = Y - \mathbb{E}(Y)$.
- Consider a function $q(t) = \mathbb{E}\left[(V - tW)^2\right]$. Verify that it must be nonnegative (why?).

[1] Silvey, S.D. (1970). *Statistical Inference.*

- Expand and form a quadratic in t.
- The above two points will imply something about the discriminant; use this to complete the question.

In doing so you will also derive the following corollary:

$$[\text{cov}(X, Y)]^2 \leq \text{Var}(X)\text{Var}(Y).$$

Proving the Cramér-Rao lower bound. With these results in hand, further assume that certain 'regularity conditions' hold such that (i) integrals with respect to y and differentiation with respect to θ are interchangeable, and (ii) the Fisher information matrix is always defined.

(c) Suppose that $\tilde{\theta}$ is an unbiased estimator of θ, that is, $\mathbb{E}(\tilde{\theta}) = \theta$. In full, this can be written

$$\begin{aligned} \mathbb{E}(\tilde{\theta}) &= \theta \\ \int_Y \tilde{\theta}(y) f(y; \theta)\, dy &= \theta. \end{aligned}$$

Give a brief explanation as to why the expectation on the left hand side is written in this way – for example, why are we integrating with respect to y? What is the role of $f(y; \theta)$? How come we write $\tilde{\theta}$ as $\tilde{\theta}(y)$?

(d) Show that

$$\mathbb{E}(\tilde{\theta}) = \theta \implies \mathbb{E}\left[\tilde{\theta}(y)\frac{\partial \log f(y; \theta)}{\partial \theta}\right] = 1.$$

(e) Show that the score has zero expectation:

$$\mathbb{E}(S(\theta)) = \mathbb{E}\left[\frac{\partial \log f(y; \theta)}{\partial \theta}\right] = 0.$$

(f) Using your above derivations, now show the Cramér-Rao lower bound holds:

$$\text{Var}(\tilde{\theta}) \geq I(\theta)^{-1}.$$

(Hint: Don't forget that the information matrix is the variance of the score! Result 2 might come in handy too.)

Attaining the Cramér-Rao lower bound. As a lower bound, all we know is that the variance of an unbiased estimator $\tilde{\theta}$ for θ cannot go below the Cramér-Rao lower bound (CRLB). So a natural question would be to ask: under what conditions is it possible for an estimator to actually reach the CRLB? For this we will need a result:

(g) **Result 3.** If equality holds in *Result 2* (i.e. $\rho^2 = 1$), then there exists a number t_0 such that it is possible to write

$$X - \mathbb{E}(X) = t_0[Y - \mathbb{E}(Y)].$$

Note that this also implies that X and Y are linearly related in the sense that one can further rearrange to get $Y = aX + b$ where a and b are constants.

(h) Using Result 3, modify your proof of the CRLB from part (f) to show that an unbiased estimator $\tilde{\theta}$ of θ attains the CRLB if the score can be written as

$$S(\theta) = I(\theta)[\tilde{\theta} - \theta].$$

(Hint: with some relabeling of variables in Result 3, perhaps one can show that the constant t_0 happens to be the information matrix $I(\theta)$. If one can show that, then the working might follow from there...)

The Cramér-Rao lower bound in action.

(i) Attaining the CRLB sounds quite restrictive. Does there actually exist a distribution for which this is possible? The answer is yes, and one such example is the *Bernoulli distribution.* Consider n i.i.d. draws $y_1, y_2, \ldots, y_n$ where $y_i \in \{0, 1\}$ from the Bernoulli distribution

$$p(y_i; \theta) = \theta^{y_i}(1-\theta)^{1-y_i}, \qquad 0 < \theta < 1,$$

where it is known that $\mathbb{E}(y_i) = \theta$ and $\text{Var}(y_i) = \theta(1-\theta)$. Furthermore, you are also aware that the information matrix associated with this distribution is

$$I(\theta) = \frac{n}{\theta(1-\theta)}.$$

Using these results, find an unbiased estimator $\tilde{\theta}$ of θ that attains the Cramér-Rao lower bound.

Question 3: MLE Asymptotics

In this question, we'll explore how the asymptotic theory we've learned in earlier tutorials will apply to maximum likelihood. Firstly: let's establish some notational conventions:

- θ_0 is the true value of the parameters from which our data are drawn from.
- $\hat{\theta}$ is the MLE.

The results we do here generalize readily to multivariate contexts. For now, let θ_0 and $\hat{\theta}$ be scalars.

(a) Show that the true value θ_0 maximizes the *population log-likelihood:*[2]

$$\theta_0 = \arg\max_{\theta \in \Theta} \mathbb{E}(\log L(\theta)).$$

Hint: We want to show that $\mathbb{E}(\log L(\theta)) \leq \mathbb{E}(\log L(\theta_0))$ for any $\theta \in \Theta$ (this is what a maximum represents, after all). To do this, first use *Jensen's Inequality* on the expression

$$\mathbb{E}(\log L(\theta) - \log L(\theta_0)) = \mathbb{E}\left[\log\left(\frac{L(\theta)}{L(\theta_0)}\right)\right]$$

[2] A heuristic explanation for this expression: let me choose any number θ from the 'parameter space' Θ on which θ is allowed to exist. If I evaluate the population log-likelihood at each of those θ values, then the one that gives me the biggest value of the population log-likelihood is the true value θ_0.

since that will give us the inequality that we need. Then, show that the RHS implied by Jensen's inequality is equal to zero (you will need the property that the integral of a marginal density equals one). Your answer will follow from there.

(b) Make the simplifying assumptions that the log-likelihood functions are continuous and have a unique maximizer. Show that under these conditions, the MLE is *consistent*:

$$\hat{\theta} \xrightarrow{p} \theta.$$

(c) Show that the negative expected value of the Hessian evaluated under the true value θ_0 is the variance of the score; that is,

$$-\mathbb{E}(H) = \text{Var}(S).$$

You may assume for simplicity that θ is a scalar so you do not have to use matrix notation (since the univariate version can generalize to multivariate versions readily. Plus, there's a lot of busy algebra anyway).

Question 4 (bonus): Hypothesis Testing in R

This question will make use of the file `Week8ExtraQ_Code.R` *provided alongside these extras.*

In this question, we're going to explore the use of linear regression in R, specifically the use of the auxiliary statistics that come with the `lm()` command. For this exercise, we'll be using the built-in `iris` dataset, which doesn't require the use of any extra files or setting working directories. Furthermore, we're going to use the following packages:

- `AER` for straightforward coefficient extraction. Also allows us to use heteroskedasticity-consistent (HC) standard errors. Also capable of performing IV regressions.
- `stargazer` for auto-formatting of regression tables into LaTeX.

However, in the event that you have to deal with data files, you should look up the `setwd()` and `read.csv` commands. There are plenty of guides on getting started in R (and one should exist for this course should you know where to look).

(a) Load in the `iris` dataset and use `stargazer()` to get some descriptive statistics.

(b) The `iris` dataset has 150 observations on four variables. Since we're only looking at how to perform computations in R, we're not very fussed about any kind of statistical inference. So let's just relabel these variables as follows:

$$\begin{aligned} y_i &= \texttt{Sepal.Length}_i \\ x_{1,i} &= \texttt{Sepal.Width}_i \\ x_{2,i} &= \texttt{Petal.Length}_i \\ x_{3,i} &= \texttt{Petal.Width}_i \end{aligned}$$

and run the following regressions:

- Equation 1. $y_i = \beta_0 + \beta_1 x_{1,i} u_i$
- Equation 2. $y_i = \beta_0 + \beta_1 x_{1.i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + u_i$

Display the results from both equations in a `stargazer()` table.

(c) Let's focus on equation 1 for now. Obtain expressions **to four decimal places** for:

(i.) The variance estimator

$$\tilde{\sigma}_n^2 = \frac{\hat{u}_1' \hat{u}_1}{n-k}$$

where $\hat{u}_1$ is the $n \times 1$ vector of OLS residuals from Equation 1.

(ii.) The coefficient of determination, R^2.

(iii.) The coefficient estimate on $x_{1,i}$, $\hat{\beta}_1$.

(d) The last part of this question involves computation of the *Breusch-Pagan* test for heteroskedasticity, which you might have seen before in previous courses. Even if you're not entirely familiar with it, that's fine; all we're going to do is calculate the appropriate LM test statistic.

(i.) First, run Equation 2 (if you haven't already done so from earlier).

(ii.) Obtain the *squared residuals* from Equation 2, denoted $\hat{u}_i^2$.

(iii.) Run the *auxiliary regression*

$$\hat{u}_i^2 = \alpha_0 + \alpha_1 x_{1.i} + \alpha_2 x_{2,i} + \alpha_3 x_{3,i} + v_i$$

and extract the coefficient of determination R^2 from it.

(iv.) Calculate the test statistic

$$LM = nR^2 \overset{a}{\sim} \chi_3^2,$$

where $m = 3$ is the number of restrictions being tested (and hence the degrees of freedom in the χ^2 distribution). Then, obtain the p-value associated with that test statistic and conclude whether you would reject the null hypothesis at the 5% level of significance. (You don't need to know what the null is in order to reject it. If you're interested, the solutions discuss further.)