ECOM40006/90013 Econometrics 3

Week 7 Extras

Question 1: Intuition for the likelihood function

Suppose that you have a list of random variables $x_1, x_2, \ldots, x_n$ that are i.i.d. with marginal density f that is determined by a vector of parameters θ. In this question, we are going to lay down the foundations for maximum likelihood estimation.

Our first object of interest is going to be the *joint density* $f(x_1, x_2, \ldots, x_n)$. This can be thought of as the 'probability' that a sample of random draws will be the same as our dataset. The goal of maximum likelihood estimation is, in principle, to maximize this probability.

(a) Show that under the assumptions above, the *likelihood function* $L(\theta; x_i)$ can be written

$$L(\theta; x_i) = f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i).$$

Explain which assumptions you use and where you use them.

(b) Show that

$$\log L(\theta; x_i) = \sum_{i=1}^{n} \log f(x_i)$$

and hence (you don't actually have to show this part)

$$\arg\max_{\theta} L(\theta; x_i) = \arg\max_{\theta} \log L(\theta; x_i).$$

Note that this particular result is derived from the fact that a maximum is invariant to positive monotone transformations.

(c) Why is it that we are trying to find the values of θ that maximize the likelihood function and not x_i?

Question 2: Maximum Likelihood, the univariate case

Consider a dataset of i.i.d. random draws $x_1, x_2, \ldots, x_n$ from the exponential distribution

$$f(x_i; \theta) = \frac{1}{\theta} \exp\left(-\frac{x_i}{\theta}\right)$$

which is defined for $\theta > 0$.

(a) Construct the likelihood function for this distribution, and hence obtain the log-likelihood function corresponding to the exponential distribution.

(b) What is the *maximum likelihood estimator* (MLE)? That is: what value of θ maximizes the log-likelihood above?

(c) Verify that the second-order condition corresponding to the MLE is consistent with a maximum.

(d) Repeat parts (a)-(c) above for the *geometric distribution*

$$f(x_i;\theta) = \theta(1-\theta)^{x_i}, \qquad x_i = 0, 1, 2, \ldots$$

where $\theta \in (0,1)$.

Question 3: Bivariate maximum likelihood

The entryway into bivariate maximum likelihood is the classic *normal distribution*. To remind you, the density of the normal distribution is

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

where $\theta = \{\mu, \sigma^2\}$. You'll have seen this in lectures but it never hurts to do it again.[1]

(a) Write out the log-likelihood function. It is not necessary to obtain the likelihood function in this case.

(b) Use the log-likelihood function to derive the *gradient vector* $G(\theta)$.

(c) Find the MLEs corresponding to the normal distribution in this case. Do the MLEs look familiar?

(d) Derive the Hessian $H(\theta)$ corresponding to the normal distribution.

(e) Evaluate the Hessian at $\hat{\theta}$ to verify that the MLEs are consistent with a maximum. Recall that in order to do this, you will need to show that the Hessian is *negative definite*. In particular, appropriate use of your MLEs will verify that the off-diagonals of the Hessian are zero,[2] from which the method of leading principal minors[3] will be sufficient.

[1] It also doesn't help that the normal distribution is one of the very rare few distributions that actually HAS an analytic solution for the MLE that we can derive. Most bivariate distributions need to have their MLEs numerically solved using some kind of iterative algorithm!

[2] You might want to go back a few steps from the MLE; the first few lines of the FOCs might be more useful to you in this case.

[3] Specifically: what you need to show is (i) the top left entry of the Hessian is negative and (ii) the determinant of the Hessian is positive at the value of the MLE $\hat{\theta}$.

Question 4: More univariate maximum likelihood

A common distribution in probability (and also econometrics) is the *Bernoulli distribution*, which is useful for modelling binary variables in practice. Let's set up a maximum likelihood using this distribution as our foundation. Let $y_1, y_2, \ldots, y_n$ be n i.i.d observations from the Bernoulli distribution with probability mass function

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

where it is known that $\mathbb{E}(y_i) = \theta_0$ and $\text{Var}(y_i) = \theta_0(1-\theta_0)$,[4] where θ_0 is the true value of θ that generates the data y_i.

(a) Obtain the likelihood function $L(\theta|y_i)$.

(b) Obtain the *log*-likelihood function $\log L(\theta|y_i)$.

(c) Derive the gradient, or score, of the log-likelihood function, denoted $S(\theta)$.

(d) Derive the Hessian of the log-likelihood function, $H(\theta)$.

(e) Solve for the *maximum likelihood estimator* (MLE) of θ, and call it $\hat{\theta}$.

(f) Calculate the expected score, $\mathbb{E}(S)$.

(g) Find the Fisher information matrix, which in this case is the variance of the score, $\text{Var}(S)$.

(h) Verify that the following equality holds:

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

[4] If you are curious about where the mean and variance comes from, some of the earliest extra questions might give you a few details.