Econometrics 2

Lecture 16: Elements of Likelihood Theory

AUEB | Spring Semester 2026

Taught by: Prof. S. Arvanitis

Notes transcribed & digitized by: T. Kourtalis

1. Elements of Likelihood Theory

Likelihood theory applies to parametric models. To each $\theta \in \Theta$ corresponds a unique distribution. Let P_θ be the distribution for the sample. This can be represented by:

- a probability mass function (when it is discrete).
- a probability density function (when it is continuous or smooth).

For the following, we assume that the statistical model is correctly specified, meaning $\exists \theta_0 \in \Theta$ and P_{θ_0} is the true distribution of the sample.

In a parametric model, for each value of the parameter, β is a well-defined function. Let:

$$\begin{aligned} p(\beta, z_n) =& p(\beta, z_{(n)} | z_{(n-1)}, z_{(n-2)}, \dots, z_{(1)}) \cdot \\ & p(\beta, z_{(n-1)} | z_{(n-2)}, \dots, z_{(1)}) \cdots \\ & p(\beta, z_{(1)}) \end{aligned}$$

In the case of independence, the factors are the marginal distributions of $z_{(i)}$ with respect to the parameter β.

2. Linear Model

Let the linear model be:

$$Y_n = X_n \beta + \varepsilon_n, \quad \varepsilon_n | X_n \sim N(0_{n\times 1}, I_{n\times n})$$

Therefore:

$$Y_n | X_n \sim N(X_n \beta, I_{n\times n})$$

For each observation i:

$$Y_{(i)} | X_n \sim N(X_{(i)} \beta, 1)$$

and is independent of $Y_{(j)}$ for all $i \neq j$, given X_n. We set $Z_n = Y_n$ and $z_{(i)} = Y_{(i)}$.

The joint distribution is given by:

$$p(\beta, z_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}(Y_n - X_n\beta)'(Y_n - X_n\beta)\right)$$

Due to the (conditional on X_n) independence, we have:

$$p(\beta, Y_{(i)}|Y_{(i-1)}, Y_{(i-2)}, \ldots, Y_{(1)}) = p(\beta, Y_{(i)}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(Y_{(i)} - X_{(i)}\beta)^2\right)$$

Thus:

$$p(\beta, z_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(Y_{(i)} - X_{(i)}\beta)^2\right)$$

3. Kullback-Leibler Divergence

The Kullback-Leibler divergence (relative entropy) is analyzed as follows. The assumption of correct specification is equivalent to $p(\beta_0, z_n)$ being the true (and unknown) distribution of our sample. Expected values with respect to this will be denoted by $\mathbb{E}_{\beta_0}$.

We have that:

$$KL(p(\beta_0, z_n), p(\beta, z_n)) := \mathbb{E}_{\beta_0}(\log p(\beta_0, z_n)) - \mathbb{E}_{\beta_0}(\log p(\beta, z_n))$$

Analyzing the second term:

$$\begin{aligned}\mathbb{E}_{\beta_0}(\log p(\beta, z_n)) &= \mathbb{E}_{\beta_0}\left(\log \prod_{i=1}^{n} p(\beta, z_{(i)}|z_{(i-1)}, z_{(i-2)}, \ldots, z_{(1)})\right) \\ &= \mathbb{E}_{\beta_0}\left(\sum_{i=1}^{n} \log p(\beta, z_{(i)}|z_{(i-1)}, z_{(i-2)}, \ldots, z_{(1)})\right)\end{aligned}$$

The above quantity is uniquely maximized at β_0. However, it is not observable because it is derived as an expected value with respect to $p(\beta_0, z_n)$, which is unknown to us. We approximate it by its sample analogue, which is:

$$\frac{1}{n}\sum_{i=1}^{n} \log p(\beta, z_{(i)}|z_{(i-1)}, \ldots, z_{(1)})$$

4. Maximum Likelihood Estimator

The Maximum Likelihood Estimator of β_0 within the framework of the parametric model is ultimately defined by the optimization problem:

$$\hat{\beta}_n = \arg\max_{\beta \in \Theta} \frac{1}{n}\sum_{i=1}^{n} \log p(\beta, z_{(i)}|z_{(i-1)}, \ldots, z_{(1)})$$

For the linear model, the logarithm of the likelihood function simplifies as follows:

$$\begin{aligned}
\frac{1}{n}\log p(\beta, z_n) &= \frac{1}{n}\sum_{i=1}^{n}\left(\log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}(Y_{(i)} - X_{(i)}\beta)^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(-\frac{1}{2}\ln(2\pi) - \frac{1}{2}(Y_{(i)} - X_{(i)}\beta)^2\right) \\
&= -\frac{1}{2}\ln(2\pi) - \frac{1}{2n}\sum_{i=1}^{n}(Y_{(i)} - X_{(i)}\beta)^2
\end{aligned}$$

5. MLE Equals OLS in the Linear Model

Maximizing $\frac{1}{n}\log p(\beta, z_n)$ over $\beta \in \Theta$ is equivalent to:

$$\hat{\beta}_n = \arg\min_{\beta\in\Theta} \frac{1}{n}\sum_{i=1}^{n}(Y_{(i)} - X_{(i)}\beta)^2 = \arg\min_{\beta\in\Theta} \frac{1}{n}(Y_n - X_n\beta)'(Y_n - X_n\beta)$$

since $-\frac{1}{2}\ln(2\pi)$ is a constant with respect to β. This is precisely the **Ordinary Least Squares** estimator, so $\hat{\beta}_n^{MLE} = \hat{\beta}_n^{OLS} = (X_n' X_n)^{-1} X_n' Y_n$.