Econometrics 2
Lecture 19: Consistency of the Maximum Likelihood Estimator

AUEB | Spring Semester 2026
Taught by: Prof. S. Arvanitis
Notes digitized by: T. Kourtalis

1 Initial Assumptions and Framework

In this lecture, we investigate the **consistency** of the Maximum Likelihood Estimator (MLE). We establish the following foundational assumptions:

- **Correct Specification:** We assume the statistical model is well-specified, meaning there exists a parameter $\beta_0 \in \Theta$ such that the density $p(z, \beta_0)$ represents the true distribution of the population.
- **Independence:** For simplicity, we assume independence (or conditional independence, as in the Probit model) between sample elements.

2 Representation of the MLE as an Optimization Problem

Under independence, the sample average log-likelihood function is:

$$l_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \ln p(z_{(i)}, \beta) \tag{1}$$

Note: In the case of conditioning on variables $X_{(i)}$, the form is $l_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \ln p(z_{(i)} | X_{(i)}, \beta)$.

For investigating consistency, it is equivalent and useful to represent the MLE, $\hat{\beta}_n$, as the solution to a minimization problem involving the difference from the true log-likelihood:

$$\hat{\beta}_n \in \arg\min_{\beta \in \Theta} [l_n(\beta_0) - l_n(\beta)] \tag{2}$$

Using logarithmic properties, this difference is expressed as:

$$l_n(\beta_0) - l_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \ln \frac{p(z_{(i)}, \beta_0)}{p(z_{(i)}, \beta)} \tag{3}$$

3 Convergence to the Kullback-Leibler Divergence

In order to investigate the consistency of l_n, we specialize our general theory for **M-type estimators** to the likelihood framework.

Under regularity conditions (homogeneity of z, stationarity of conditioning variables, and analytical properties of $p(z, \beta)$), the Law of Large Numbers (LLN) ensures that the sample average converges in probability to its expectation:

$$l_n(\beta_0) - l_n(\beta) \xrightarrow{p} \mathbb{E}_{\beta_0}\left(\ln \frac{p(z_{(1)}, \beta_0)}{p(z_{(1)}, \beta)}\right) = KL\left(p(z_{(1)}, \beta_0), p(z_{(1)}, \beta)\right) \tag{4}$$

This limit is the **Kullback-Leibler (KL) Divergence**. For the proof to hold, this convergence is typically required to be **locally uniform** with respect to β.

In conditional models, the limit involves the expectation over the distribution of $X_{(1)}$:

$$\mathbb{E}_{X_{(1)}}\left[KL\left(p(z_{(1)}|X_{(1)}, \beta_0), p(z_{(1)}|X_{(1)}, \beta)\right)\right] \tag{5}$$

4 Properties of the Asymptotic Criterion and Consistency

The KL divergence satisfies:

1. $KL(p, q) \geq 0$ for any densities p, q.
2. $KL(p, q) = 0 \iff p = q$.

The asymptotic criterion is minimized when $p(z, \beta) = p(z, \beta_0)$. If the **identification condition** holds ($p(z, \beta) = p(z, \beta_0) \implies \beta = \beta_0$), then β_0 is the unique global minimizer.

Under further regularity conditions—specifically the **compactness** of Θ and **continuity** of $p(z, \beta)$ (or **convexity** of the criterion and Θ)—the parameter β_0 is identifiable/separable. Thus:

$$\hat{\beta}_n \xrightarrow{p} \arg\min_{\beta \in \Theta} KL\left(p(z_{(1)}, \beta_0), p(z_{(1)}, \beta)\right) = \beta_0 \tag{6}$$

5 Specialization to the Probit Model

In a well-specified Probit model with conditional independence:

$$l_n(\beta_0) - l_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left[Y_{(i)} \ln \frac{\Phi(X_{(i)}\beta_0)}{\Phi(X_{(i)}\beta)} + (1 - Y_{(i)}) \ln \frac{1 - \Phi(X_{(i)}\beta_0)}{1 - \Phi(X_{(i)}\beta)}\right] \tag{7}$$

Under appropriate conditions (locally uniform convergence w.r.t. β):

$$l_n(\beta_0) - l_n(\beta) \xrightarrow{p} \mathbb{E}\underbrace{\left(Y_{(1)} \ln \frac{\Phi(X_{(1)}\beta_0)}{\Phi(X_{(1)}\beta)} + (1 - Y_{(1)}) \ln \frac{1 - \Phi(X_{(1)}\beta_0)}{1 - \Phi(X_{(1)}\beta)}\right)}_{A} \tag{*}$$

By the **Law of Iterated Expectations (L.I.E.)**:

$$(*) = \mathbb{E}(A) = \mathbb{E}\left(\mathbb{E}(A|X_{(1)})\right) \tag{8}$$

The conditional expectation, given $Y_{(1)}|X_{(1)} \sim Ber(\Phi(X_{(1)}\beta_0))$, is:

$$\Phi(X_{(1)}\beta_0) \ln \frac{\Phi(X_{(1)}\beta_0)}{\Phi(X_{(1)}\beta)} + (1 - \Phi(X_{(1)}\beta_0)) \ln \frac{1 - \Phi(X_{(1)}\beta_0)}{1 - \Phi(X_{(1)}\beta)} \tag{9}$$

This is the KL divergence between Bernoulli distributions: $KL(Ber(\beta_0; X_{(1)}), Ber(\beta; X_{(1)}))$. The final asymptotic criterion is:

$$\mathbb{E}\left(KL\left(Ber(\beta_0; X_{(1)}), Ber(\beta; X_{(1)})\right)\right) \tag{10}$$

Consistency is established as β_0 is the unique minimizer of this expected KL divergence.