

Econ 722 – Advanced Econometrics IV, Part II

Francis J. DiTraglia

University of Pennsylvania

Lecture #1 – AIC-type Information Criteria

Kullback-Leibler (KL) Divergence

Motivation

How well does a given density $f(y)$ approximate an unknown true density $g(y)$? Use this to select between parametric models.

Definition

$$\text{KL}(g; f) = \underbrace{\mathbb{E}_G \left[\log \left\{ \frac{g(Y)}{f(Y)} \right\} \right]}_{\text{True density on top}} = \underbrace{\mathbb{E}_G [\log g(Y)]}_{\substack{\text{Depends only on truth} \\ \text{Fixed across models}}} - \underbrace{\mathbb{E}_G [\log f(Y)]}_{\text{Expected log-likelihood}}$$

Properties

- ▶ Not symmetric: $\text{KL}(g; f) \neq \text{KL}(f; g)$
- ▶ By Jensen's Inequality: $\text{KL}(g; f) \geq 0$ (strict iff $g = f$ a.e.)
- ▶ Minimize KL \iff Maximize Expected log-likelihood

KL Divergence and Mis-specified MLE

Pseudo-true Parameter Value θ_0

$$\hat{\theta}_{MLE} \xrightarrow{P} \theta_0 \equiv \arg \min_{\theta \in \Theta} \text{KL}(g; f_{\theta}) = \arg \max_{\theta \in \Theta} \mathbb{E}_G[\log f(Y|\theta)]$$

What if f_{θ} is correctly specified?

If $g = f_{\theta}$ for some θ then $\text{KL}(g; f_{\theta})$ is minimized at zero.

Goal: Compare Mis-specified Models

$$\mathbb{E}_G [\log f(Y|\theta_0)] \quad \text{versus} \quad \mathbb{E}_G [\log h(Y|\gamma_0)]$$

where θ_0 is the pseudo-true parameter value for f_{θ} and γ_0 is the pseudo-true parameter value for h_{γ} .

How to Estimate Expected Log Likelihood?

For simplicity: $Y_1, \dots, Y_n \sim \text{iid } g(y)$

Unbiased but Infeasible

$$\mathbb{E}_G \left[\frac{1}{T} \ell(\theta_0) \right] = \mathbb{E}_G \left[\frac{1}{T} \sum_{t=1}^T f(Y_t | \theta_0) \right] = \mathbb{E}_G [f(Y | \theta_0)]$$

Biased but Feasible

$T^{-1} \ell(\hat{\theta}_{MLE})$ is a **biased** estimator of $\mathbb{E}_G[\log f(Y | \theta_0)]$.

Intuition for the Bias

$T^{-1} \ell(\hat{\theta}_{MLE}) > T^{-1} \ell(\theta_0)$ unless $\hat{\theta}_{MLE} = \theta_0$. Maximized sample log-like. is an **overly optimistic** estimator of expected log-like.

What to do about this bias?

1. General-purpose asymptotic approximation of “degree of over-optimism” of maximized sample log-likelihood.
 - ▶ Takeuchi’s Information Criterion (TIC)
 - ▶ Akaike’s Information Criterion (AIC)
2. Problem-specific finite sample approach, assuming $g \in f_\theta$.
 - ▶ Corrected AIC (AIC_c) of Hurvich and Tsai (1989)

Tradeoffs

TIC is most general and makes weakest assumptions, but requires very large T to work well. AIC is a good approximation to TIC that requires less data. Both AIC and TIC perform poorly when T is small relative to the number of parameters, hence AIC_c .

Recall: Asymptotics for Mis-specified ML Estimation

Model $f(y|\theta)$, pseudo-true parameter θ_0 . For simplicity $Y_1, \dots, Y_T \sim \text{iid } g(y)$.

Fundamental Expansion

$$\sqrt{T}(\hat{\theta} - \theta_0) = J^{-1} \left(\sqrt{T} \bar{U}_T \right) + o_p(1)$$

$$J = -\mathbb{E}_G \left[\frac{\partial \log f(Y|\theta_0)}{\partial \theta \partial \theta'} \right], \quad \bar{U}_T = \frac{1}{T} \sum_{t=1}^T \frac{\partial \log f(Y_t)}{\partial \theta}$$

Central Limit Theorem

$$\sqrt{T} \bar{U}_T \rightarrow_d U \sim N_p(0, K), \quad K = \text{Var}_G \left[\frac{\partial \log f(Y|\theta_0)}{\partial \theta} \right]$$

$$\sqrt{T}(\hat{\theta} - \theta_0) \rightarrow_d J^{-1} U \sim N_p(0, J^{-1} K J^{-1})$$

Information Matrix Equality

If $g = f_\theta$ for some $\theta \in \Theta$ then $K = J \implies \text{AVAR} = J^{-1}$

Bias Relative to Infeasible Plug-in Estimator

$$B = \frac{1}{T} \ell(\hat{\theta}) - \underbrace{\int g(y) \log f(y|\hat{\theta}) dy}_{\text{uses data only once}}$$