VARIABLE SELECTION WITH AKAIKE INFORMATION CRITERIA: A COMPARATIVE STUDY

Meral Candan Çetin* and Aydin Erar*

Received 20.06.2002

Abstract

In this paper, the problem of variable selection in linear regression is considered. This problem involves choosing the most appropriate model from the candidate models. Variable selection criteria based on estimates of the Kullback-Leibler information are most common. Akaike’s AIC and bias corrected AIC belong to this group of criteria. The reduction of the bias in estimating the Kullback-Leibler information can lead to better variable selection. In this study we have compared the Akaike Criterion based on Fisher Information and AIC criteria based on Kullback-Leibler.

Key Words: Akaike information criteria, robust selection, Kullback information, variable selection

1. Introduction

The Akaike information criterion and the corrected Akaike information criterion are based on estimators of expected Kullback-Leibler information. For the benefit of the reader we briefly explain Kullback-Leibler and Fisher information.

Kullback-Leibler Information:

Kullback-Leibler (K-L) information is used as a means of discriminating between the true model and the candidate model. Suppose $X$ is a continuous random vector and $f(x/\theta)$ a probability density function of $x$, where $\theta$ is a $p$-dimensional parameter vector $\theta = (\theta_1, \theta_2, \ldots, \theta_p)$, $\theta \in \mathbb{R}^p$.

Let $\theta^*$ be the true parameter of $\theta$ with density function $f(x/\theta^*)$. Kullback-Leibler information, or the generalized entropy $B$ of Boltzmann, measure the closeness of $f(x/\theta^*)$ to $f(x/\theta)$:

$$B(\theta^*; \theta) = -\text{I}(\theta^*; \theta) = E[\log f(x/\theta) - \log f(x/\theta^*)]$$

*Hacettepe University, Department of Statistics, Beytepe, Ankara, Turkey.