Journal of Statistical Planning and Inference 170 (2016) 106–116

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Consistent model selection in segmented line regression

Jeankyung Kim [a], Hyune-Ju Kim [b,*]

[a] Department of Statistics, Inha University, 253 Yonghyundong, Namgu, Incheon, 402-751, Republic of Korea
[b] Department of Mathematics, Syracuse University, 215 Carnegie Building, Syracuse, NY 13244-1150, USA

ARTICLE INFO

Article history:
Received 19 September 2014
Received in revised form 28 May 2015
Accepted 23 September 2015
Available online 13 October 2015

Keywords:
Bayes Information Criterion
Segmented line regression
Model selection

ABSTRACT

The Schwarz criterion or Bayes Information Criterion (BIC) is often used to select a model dimension, and some variations of the BIC have been proposed in the context of change-point problems. In this paper, we consider a segmented line regression model with an unknown number of change-points and study asymptotic properties of Schwarz type criteria in selecting the number of change-points. Noticing the over-estimating tendency of the traditional BIC observed in some empirical studies and being motivated by asymptotic behavior of the modified BIC proposed by Zhang and Siegmund (2007), we consider a variation of the Schwarz type criterion that applies a harsher penalty equivalent to the model with one additional unknown parameter per segment. For the segmented line regression model without the continuity constraint, we prove the consistency of the number of change-points selected by the criterion with such type of a modification and summarize the simulation results that support the consistency. Further simulations are conducted for the model with the continuity constraint, and we empirically observe that the asymptotic behavior of this modified version of BIC is comparable to that of the criterion proposed by Liu et al. (1997).

1. Introduction

A main concern in regression model selection is how to select the "best" set of independent variables, and two major approaches of the model selection are hypothesis testing and information criteria approaches. In the context of change-point problems, both approaches have been applied to select the number of change-points, and their analytic and empirical properties have been investigated by many researchers. One of widely used information criteria is the Bayes Information Criterion (BIC) proposed by Schwarz (1978). This Schwarz criterion selects the model dimension by finding the Bayes solution that maximizes a posterior probability of the model, and Schwarz (1978) derived the following criterion by evaluating the leading terms of its asymptotic expansion:

$$\mathrm{SC}(p) = \sup_{\boldsymbol{\theta}_p} \log(\mathrm{lik}(\boldsymbol{\theta}_p)) - \frac{p}{2}\log n = \log(\mathrm{lik}(\hat{\boldsymbol{\theta}}_p)) - \frac{p}{2}\log n,$$

where $\mathrm{lik}(\boldsymbol{\theta}_p)$ is the likelihood function of $\boldsymbol{\theta}_p$ for the model with dimension p and $\hat{\boldsymbol{\theta}}_p$ is the maximum likelihood estimator of $\boldsymbol{\theta}_p$. As in general information criteria, the Schwarz criterion has two parts, the log of the maximized likelihood function and the penalty function that penalizes for the model dimension, and the method selects the model that maximizes $\mathrm{SC}(p)$.

* Corresponding author.
E-mail address: hjkim@syr.edu (H.-J. Kim).

http://dx.doi.org/10.1016/j.jspi.2015.09.008
0378-3758/

Note that its validity is established in Schwarz (1978) for "the case of independent, identically distributed observations, and linear models".

Yao (1988) studied the problem to select the number of change-points in means of normally distributed random variables, where the total number of unknown parameters for the model with k change-points is $p = 2(k+1)$. For the number of change-points estimated by minimizing $-SC(p)$ with $p = 2k$, Yao (1988) proved its consistency. Lee (1997) considered a similar type of a criterion to select the number of change-points in a sequence of random variables from an exponential family distribution. Under some mild conditions on spacings of successive change-points, Lee (1997) proved the consistency of the number of change-points estimated by the Schwarz type criterion whose penalty term is greater than $2k(1+\epsilon_0)\log n$ for some $\epsilon_0 > 0$. Zhang and Siegmund (2007) noted that the usage of the Schwarz criterion "is not theoretically justified" in their situation due to irregularities in the likelihood function and proposed a modified BIC derived as an asymptotic approximation of the Bayes factor to determine the number of change-points in means of normally distributed random variables. For other types of modifications and applications for detecting mean changes, see Ninomiya (2005), Pan and Chen (2006), and Hannart and Naveau (2012).

In the context of segmented line regression, similar approaches have been proposed to select the number of change-points. Kim et al. (2000) proposed the permutation test to select the number of change-points in the segmented line regression model where segments are assumed to be continuous at change-points, called the joinpoint regression model in their paper. Kim et al. (2009) considered the traditional BIC,

$$\mathrm{BIC}(k) = -\frac{2}{n}\mathrm{SC}(2k) = \log(\mathrm{RSS}_k/n) + 2k\frac{\log n}{n},$$

where RSS_k is the residual sum of squares for the model with k change-points, and compared its performance with those of the permutation test procedure of Kim et al. (2000) and the method based on generalized cross validation used in MARS of Friedman (1991). Note that the penalty term of $2k\frac{\log n}{n}$ is chosen based on $2k+3$ unknown parameters for the joinpoint regression model with k change-points. Liu et al. (1997) considered a general segmented line regression model allowing a discontinuity at the change-point and non-Gaussian errors, proposed a penalty term with a bigger order than that of $\mathrm{BIC}(k)$, and proved the consistency of the dimension selected by minimizing their criterion:

$$\mathrm{MIC}(k) = \log(\mathrm{RSS}_k/(n-p^*)) + p^*\frac{c_0(\log n)^{2+\delta_0}}{n},$$

where $p^* = p^*(k) = (k+1)p + k$ for the model with k change-points and p covariates and c_0 and δ_0 are positive constants. Two Bayesian model selection methods based on the Bayes factor and a Bayesian version of BIC were developed in Tiwari et al. (2005) who investigated their empirical properties via simulations and compared their performances with that of the permutation procedure of Kim et al. (2000). Martinez-Beneito et al. (2011) also proposed a Bayesian model selection method that provides posterior probabilities and is flexible to work with Poisson count data.

This paper is motivated from empirical results where the traditional BIC indicated a tendency to over-estimate the number of change-points (See Table 1 of Kim et al., 2009, Table 1 of Zhang and Siegmund, 2007). When the argument of Zhang and Siegmund (2007) is applied to segmented line regression, the penalty of the modified BIC is harsher than that of the traditional BIC, asymptotically corresponding to one additional unknown parameter per segment under some conditions, and this motivated us to consider a BIC type criterion whose penalty is $4k\frac{\log n}{n}$ for the segmented line regression model without the continuity constraint and $3k\frac{\log n}{n}$ for the model with the continuity constraint. Note that for segmented line regression with k change-points, the number of unknown parameters is $3k+3$ for the model without the continuity constraint and $2k+3$ for the model with the continuity constraint. Let

$$\mathrm{BIC}_d(k) = \log(\mathrm{RSS}_k/n) + PE_d(k) = \log(\mathrm{RSS}_k/n) + dk\frac{\log n}{n}, \tag{1}$$

for penalty coefficients, d. Then the traditional BICs are BIC_3 for the unconstrained model and BIC_2 for the constrained model.

Our interest in this paper is on asymptotic behavior of BIC_d, a simple model selection criterion whose penalty term has the same order as that of the traditional BIC, and we focus on asymptotic properties of BIC_4 for the model without the continuity constraint and BIC_3 for the model with the continuity constraint. In Section 2, we formally introduce the unconstrained model and prove the consistency of the model dimension selected by BIC_4 for the unconstrained model with Gaussian errors. This result provides a consistent model selection criterion that imposes an asymptotically milder penalty than MIC does. In Section 3, we present the results of a simulation study where we compare the performance of BIC_4 with those of BIC_3 and MIC. Section 4 includes empirical results and discussion on the constrained case where the segments are constrained to be continuous at the change-points. Further discussion is presented in Section 5.

2. Selection methods and consistency: Unconstrained model

Suppose that we observe $(x_1, y_1), \ldots, (x_n, y_n)$ and consider a segmented line regression model such that

$$y_i = \beta_{j,0} + \beta_{j,1}x_i + \epsilon_i, \quad \text{if } \tau_{j-1} < x_i \le \tau_j \ (j = 1, \ldots, \kappa+1), \tag{2}$$

where κ is the unknown number of change-points, the τ's are unknown change-points with $\tau_0 = \min_i x_i - \frac{1}{n}$ and $\tau_{\kappa+1} = \max_i x_i$, and the ϵ_i are independent $N(0, \sigma^2)$.