Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Statistics & Probability Letters 72 (2005) 237–247

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

Information criterion for Gaussian change-point model

Yoshiyuki Ninomiya*

Graduate School of Mathematics, Kyushu University, 6-10-1, Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan

Received 11 August 2003; accepted 4 October 2004

Abstract

AIC-type information criterion is generally estimated by the bias-corrected maximum log-likelihood. In regular models, the bias can be estimated by p, where p is the number of parameters. The present paper considers the AIC-type information criterion for change-point models which are not regular, the bias of which will not be the same as for regular models. The bias is shown to depend on the expected maximum of a random walk with negative drift. Furthermore, it is shown that by using an approximation to a Brownian motion, the evaluated bias is given by $3m + p_m$ (not $m + p_m$), where m is the number of change-points and p_m is the number of regular parameters, which differs from regular models.

Keywords: Akaike's information criterion; Brownian motion with drift; Maximum of random walk; Number of change-points

1. Introduction

Change-point problems have been applied using data from various fields, such as annual Nile river flow data (Cobb, 1978), coal mining disaster data (Jarrett, 1979) and stock price data (Chen and Gupta, 1997), and several statistical methods for analyzing change-point problems have been proposed and discussed. Consistent estimators of the change-point were reported by Hinkley (1970), Carlstein (1988) and Loader (1996). The asymptotic properties of the likelihood ratio test statistics for detecting the change-point were investigated by James et al. (1987), Horváth (1993),

*Tel.:/fax: +81 92 642 2756.
E-mail address: nino@math.kyushu-u.ac.jp.

0167-7152/$ - see front matter
doi:10.1016/j.spl.2004.10.037

Yao (1993), Davis et al. (1995) and Ninomiya (2004). One difficulty involved in handling the change-point is that the change-point is not a regular parameter, so that the conventional asymptotic normalities of the above estimators do not hold and the test statistics mentioned above do not converge to χ^2 statistics. Because of this difficulty, unlike the Schwarz-type information criterion (Yao, 1988; Lee, 1995; Braun et al., 2000). AIC-type information criterion, which is an information criterion based on the Kullback–Leibler divergence, for the change-point model has not been discussed previously.

This paper presents the AIC-type information criterion for the Gaussian change-point model for the purpose of model selection, including estimation of the number of change-points. The AIC-type information criterion is generally estimated by the bias-corrected maximum log-likelihood. In regular models, the bias can be estimated by the number of parameters, p, and multiplying the bias-corrected maximum log-likelihood by -2 yields the AIC (Akaike, 1973). In change-point models, the bias cannot be obtained by this method, because the change-point is not a regular parameter. This paper shows that the bias depends on the expected value of the maximum of a random walk with a negative drift. Furthermore, we show that by using an approximation to a Brownian motion, the evaluated bias is given by $3m + p_m$ (not $m + p_m$), where m is the number of change-points and p_m is the number of regular parameters, which differs from regular models.

This paper is organized as follows. The information criterion is constructed in Section 2. An approximation of the penalty term to a Brownian motion is discussed in Section 3, and the approximation is supported by numerical experiments in Section 4.

2. Main results

Assuming that $\{X_i,\ 1 \leqslant i \leqslant n\}$ is an independent d-dimensional Gaussian sequence with common variance, that is, the density function of X_i is

$$f(x|\mu_i, \Sigma) = (2\pi)^{-d/2}(\det \Sigma)^{-1/2} \exp\{-(x-\mu_i)^{\mathrm{T}}\Sigma^{-1}(x-\mu_i)\}, \tag{1}$$

where $\mu_i \in \boldsymbol{R}^d$ and Σ is positive definite $d \times d$ matrix. In addition, we assume that the parameters change as follows:

$$\mu_1 = \cdots = \mu_{k^{*(1)}} = \mu^{(1)} \neq \mu_{k^{*(1)}+1} = \cdots = \mu_{k^{*(2)}} = \mu^{(2)} \neq \cdots \neq \mu_{k^{*(m)}+1} = \cdots = \mu_n = \mu^{(m+1)}, \tag{2}$$

that is, $k^{*(1)}, \ldots, k^{*(m)}$ are change-points. $k^* = (k^{*(1)}, \ldots, k^{*(m)})$ is unknown and satisfies the condition that k^*/n is constant as $n \to \infty$. This assumption for asymptotic theory is usual in change-point analysis (see Csörgő and Horváth, 1996, for example). Note that the following theory can be extended to a class of dependent sequences, such as weakly dependent stationary sequences. However, we assume independence based on previous reports regarding change-points and on the ease of extension.

Let $\hat{k}_X$, $\hat{\mu}_X$ and $\hat{\Sigma}_X$ be the maximum likelihood estimator of k^*, μ and Σ based on $X = (X_1, \ldots, X_n)$, and let $f(X|k^*, \mu, \Sigma)$ be the joint density function of X. According to the AIC, and the extended versions thereof, (for example, Konishi and Kitagawa, 1996), model selection can be approached by trying to minimize the Kullback–Leibler divergence of the true model with respect to the fitted model, that is, by trying to maximize the expected log-likelihood