



HAL
open science

Optimal Penalty Term for EDC Markov Chain Order Estimator

Chang C. Y. Dorea

► **To cite this version:**

Chang C. Y. Dorea. Optimal Penalty Term for EDC Markov Chain Order Estimator. Annales de l'ISUP, 2008, LII (3), pp.15-25. hal-03633210

HAL Id: hal-03633210

<https://hal.science/hal-03633210v1>

Submitted on 6 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimal Penalty Term for EDC Markov Chain Order Estimator

Chang C.Y. Dorea¹
 Universidade de Brasília

Abstract

The Efficient Determination Criterion (EDC) generalizes the AIC and BIC criteria and provides a class of consistent estimators for the order of a Markov chain. Several authors have addressed the question of choosing between AIC or BIC estimates. The first tends to overestimate the order and the second, though consistent, may lead to underestimation. All these estimators are based on penalized maximum log-likelihood functions. In this work, we study the choice of the optimal penalty term and show that corresponds to neither AIC nor BIC estimators.

AMS 2000 Subject Classification: 60J10, 62F12, 62M05.

Key Words and Phrases: Efficient Determination Criterion; Markov chain order; penalty term; AIC; BIC.

1 Introduction

Let $X_1^n = (X_1, \dots, X_n)$ be a sample from a multiple Markov chain $X = \{X_n\}_{n \geq 1}$ of unknown order r . Assume that X takes value on a finite state space E and that the transition probabilities are given by

$$p(a_{r+1}|a_1^r) = P(X_{n+1} = a_{r+1} | X_{n-r+1}^n = a_1^r) \quad (1.1)$$

where $a_1^r = a_1^k a_{k+1}^r = (a_1, \dots, a_r) \in E^r$.

The approximation of Kullback-Leibler information measure by Neyman-Pearson statistics along with the asymptotic χ^2 -distribution of the maximum log-likelihood ratio form

¹Partially supported by CNPq, FAPDF/PRONEX, CAPES/PROCAD and FINATEC/UnB.
 Depto. de Matemática, Universidade de Brasília, 70910-900, Brasília-DF, Brazil; cdorea@mat.unb.br.

the basis to derive the Akaike's information criterion (AIC, Akaike (1974)),

$$\text{AIC}(k) = -2\log \hat{L}(k) + 2\gamma(k), \quad \gamma(k) = |E|^k(|E| - 1). \quad (1.2)$$

Where $|E|$ denotes the cardinality of the set E ,

$$\begin{aligned} \log \hat{L}(k) &= \sum_{a_1^{k+1}} N(a_1^{k+1}|X_1^n) \log \frac{N(a_1^{k+1}|X_1^n)}{N(a_1^k|X_1^n)}, \\ \log \hat{L}(0) &= \sum_a N(a|X_1^n) \log \frac{N(a|X_1^n)}{n} \end{aligned} \quad (1.3)$$

and

$$N(a_1^k|X_1^n) = \sum_{j=1}^{n-k+1} 1(X_j = a_1, \dots, X_{j+k-1} = a_k)$$

that is, the number of occurrences of a_1^k in X_1^n . If $k = 0$ we take $N(\cdot|X_1^n) = n$. The sums are taken over positive terms $N(a_1^{k+1}|X_1^n) > 0$, or else, we convention $0/0$ or $0 \cdot \infty$ as 0. The function $\gamma(k)$ in the penalty term is just the number of degrees of freedom if all transition probabilities are strictly positive. In the framework of nested model selection r is estimated by

$$\hat{r}_n = \arg \min_{0 \leq k \leq K} \text{AIC}(k)$$

where K is an upper bound for r , assumed to be finite.

The BIC estimator is an alternative to the AIC, proposed by Tong (1975) and Schwarz (1978), who showed its consistency,

$$\text{BIC}(k) = -2\log \hat{L}(k) + \gamma(k) \log n, \quad 0 \leq k \leq K. \quad (1.4)$$

Katz (1981) pointed out the inconsistency of AIC estimates. Other authors have addressed the consistency problem for BIC, see, for example, Finesso et al. (1996). More recently, Csiszár and Shields (2000) established the strong consistency for BIC, which strengthens earlier results by removing the assumption that sets up an a priori bound on the order.

Zhao et al. (2001) proposed the EDC criterion that is strongly consistent and from which we can derive the AIC and the BIC criteria by appropriately choosing the penalty function,

$$\text{EDC}(k) = -2\log \hat{L}(k) + \gamma(k)c_n. \quad (1.5)$$

It is shown that $\hat{r}_n = \arg \min_{0 \leq k \leq K} \text{EDC}(k)$ is strongly consistent if $c_n > 0$ satisfies

$$\frac{c_n}{\log \log n} \xrightarrow{n} \infty \quad \text{and} \quad \frac{c_n}{n} \xrightarrow{n} 0. \quad (1.6)$$

Our Corollary 2 shows that strong consistency for EDC can be obtained under weaker conditions and without assumption of $K < \infty$. An interesting problem is the best choice for the penalty term $\gamma(k)c_n$. In fact, Csiszár and Shields (2000) posed the problem relative to BIC: "it remains open whether smaller penalty terms suffice for consistency in the absence of a prior bound on the order". In section 3, we justify why AIC tends to overestimate r and indicates how BIC may lead to possible underestimation. Corollary 1 shows that the best choice is given by

$$\text{EDC}_{opt}(k) = -2 \log \hat{L}(k) + 2|E|^{k+1} \log \log n.$$

2 Auxiliary Results

First, observe that for $k \geq r$ the transition probabilities for the first order Markov chain $Y^{(k)} = \{Y_n^{(k)}\}_{n \geq 1}$, $Y_n^{(k)} = (X_n, \dots, X_{n+k-1})$, satisfy

$$\begin{aligned} p(a_2^{k+1}|a_1^k) &= P(Y_{n+1}^{(k)} = a_2^{k+1} | Y_n^{(k)} = a_1^k) \\ &= P(X_{n+k} = a_{k+1} | X_n^{n+k-1} = a_1^k) \\ &= p(a_{k+1}|a_1^k) = p(a_{k+1}|a_{k-r+1}^k). \end{aligned} \quad (2.1)$$

We will assume that $Y^{(r)}$ is an ergodic Markov chain, that is, there exists an equilibrium (stationary) distribution $\pi(\cdot)$ satisfying

$$\lim_{n \rightarrow \infty} |P_\nu(Y_n^{(r)} = a_1^r) - \pi(a_1^r)| = 0, \quad \forall a_1^r \in E^r,$$

where ν is any initial distribution. Denoting $aa_1^{r-1} = (a, a_1, \dots, a_{r-1})$ we have

$$\pi(a_1^r) = \sum_{b_1^r} \pi(b_1^r) p(a_r | b_1^r) = \sum_a \pi(aa_1^{r-1}) p(a_r | aa_1^{r-1}).$$

Using the notation

$$\pi(a_1^k) = \pi(a_1^r) p(a_{r+1} | a_1^r) \dots p(a_k | a_{k-r}^{k-1}), \quad k > r \quad (2.2)$$

we obtain

$$\pi(a_1^k) = \sum_a \pi(aa_1^{k-1}) p(a_k | a_{k-r}^{k-1}). \quad (2.3)$$

Which shows that $\pi(\cdot)$ defined by (2.2) is a stationary distribution for $Y^{(k)}$.

Proposition 1. Assume that the derived Markov chain $Y^{(r)}$ is ergodic then for $k > r$ the process $Y^{(k)}$ is ergodic with stationary distribution given by (2.2). Moreover, if g is a bounded function then there exist $\alpha_g > 0$ and $0 < \rho < 1$ such that,

$$|E\{g(Y_{j+l}^{(k)}) | \mathcal{F}_j\} - E\{g(Y_{j+l}^{(k)})\}| \leq \alpha_g \rho^l, \quad (2.4)$$

where

$$\mathcal{F}_j = \sigma(X_1, X_2, \dots, X_j). \quad (2.5)$$

The proof of ergodicity is immediate and (2.4) is just Proposition 3.1 from Roussas and Ioannides (1987). Next, we state a version of the Law of Iterated Logarithm from Meyn and Tweedie (1993, theorems 17.0.1 and 17.2.2), adapted for our needs.

Lemma 1. Let $\{Z_n\}$ be an ergodic Markov chain with finite state E and stationary distribution π . Let g be any function on E and $S_n(g) = \sum_{j=1}^n g(Z_j)$. Then, for

$$\sigma_g^2 = E_\pi\{g^2(Z_1)\} + 2 \sum_{j=2}^n E_\pi\{g(Z_1)g(Z_j)\} \quad (2.6)$$

we have almost surely

$$\lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} [S_n(g) - E_\pi\{S_n(g)\}] = 0, \quad \text{if } \sigma_g^2 = 0,$$

and for $\sigma_g^2 > 0$

$$\liminf_{n \rightarrow \infty} \frac{S_n(g) - E_\pi\{S_n(g)\}}{\sqrt{2\sigma_g^2 n \log \log n}} = -1 \quad \text{and} \quad \limsup_{n \rightarrow \infty} \frac{S_n(g) - E_\pi\{S_n(g)\}}{\sqrt{2\sigma_g^2 n \log \log n}} = 1,$$

(E_π : expectation with initial distribution π ; a.s. : almost surely).

Lemma 2. If $Y^{(r)}$ is ergodic then for $k \geq r$ and $a_1^{k+1} \in E^{k+1}$ we have a.s.

$$\limsup_{n \rightarrow \infty} \frac{[N(a_1^{k+1}|X_1^n) - N(a_1^k|X_1^n)p(a_{k+1}|a_1^k)]^2}{n \log \log n} = 2\pi(a_1^{k+1})(1 - p(a_{k+1}|a_1^k)). \quad (2.7)$$

Proof. Define

$$g(Y_j^{(k+1)}) = 1(Y_j^{(k+1)} = a_1^{k+1}) - 1(Y_j^{(k)} = a_1^k)p(a_{k+1}|a_1^k)$$

and

$$\begin{aligned} S_n(g) &= \sum_{j=1}^{n-k} g(Y_j^{(k+1)}) \\ &= N(a_1^{k+1}|X_1^n) - N(a_1^k|X_1^n)p(a_{k+1}|a_1^k) + o(\delta_n). \end{aligned}$$

The term $o(\delta_n)$ stands for

$$A_n = o(\delta_n) \quad \text{if} \quad \frac{A_n}{\delta_n} \xrightarrow{n} \infty \quad \text{for any} \quad \delta_n \xrightarrow{n} \infty. \quad (2.8)$$

Since $E_\pi\{g(Y_j^{(k+1)})\} = 0$ we have

$$S_n(g) - E_\pi\{S_n(g)\} = N(a_1^{k+1}|X_1^n) - N(a_1^k|X_1^n)p(a_{k+1}|a_1^k) + o(\delta_n).$$

Also, since π is the stationary distribution we have by (2.2)

$$E_\pi\{g^2(Y_1^{(k+1)})\} = \pi(a_1^{k+1})(1 - p(a_{k+1}|a_1^k))$$

and for \mathcal{F}_j defined by (2.5) we have for $j > 1$

$$\begin{aligned} E\{1(Y_j^{(k+1)} = a_1^{k+1})|\mathcal{F}_{j+k-1}\} &= E\{1(Y_j^{(k)} = a_1^k)1(X_{j+k} = a_{k+1})|\mathcal{F}_{j+k-1}\} \\ &= 1(Y_j^{(k)} = a_1^k)p(a_{k+1}|a_1^k). \end{aligned}$$

Since $Y_1^{(k+1)}$ is \mathcal{F}_{j+k-1} -measurable we have

$$\begin{aligned} E\{g(Y_1^{(k+1)})g(Y_j^{(k+1)})|\mathcal{F}_{j+k-1}\} &= \\ &= g(Y_1^{(k+1)})[E\{1(Y_j^{(k+1)} = a_1^{k+1})|\mathcal{F}_{j+k-1}\} - 1(Y_j^{(k)} = a_1^k)p(a_{k+1}|a_1^k)] = 0. \end{aligned}$$

Let $\sigma_g^2 = E_\pi\{g^2(Y_1^{(k+1)})\} = \pi(a_1^{k+1})[1 - p(a_{k+1}|a_1^k)]$. Applying Lemma 1 we have (2.7) for $\sigma_g^2 \geq 0$. \square

3 Optimal Penalty Term

For $k \geq r$ the log-likelihood function satisfies

$$\begin{aligned} \log L(k) &= \sum_{a_1^{k+1}} N(a_1^{k+1}|X_1^n) \log p(a_{k+1}|a_1^k) \\ &= \sum_{a_1^{k+1}} N(a_1^{k+1}|X_1^n) \log p(a_{k+1}|a_{k-r+1}^k) \\ &= \log L(r) + o(\delta_n) \end{aligned} \quad (3.1)$$

where $o(\delta_n)$ is defined by (2.8). Similarly, for $0 \leq k < r$ we can write

$$\log \hat{L}(k) = \sum_{a_1^{r+1}} N(a_1^{r+1}|X_1^n) \log \frac{N(a_{r+1-k}^{r+1}|X_1^n)}{N(a_{r+1-k}^r|X_1^n)} + o(\delta_n). \quad (3.2)$$

For $k \geq r$ define

$$E_+^{k+1} = \{a_1^{k+1} : \pi(a_1^{k+1}) > 0\} \quad \text{and} \quad \gamma_+(k) = \sum_{a_1^{k+1} \in E_+^{k+1}} (1 - p(a_{k+1}|a_1^k)) \quad (3.3)$$

and for $0 \leq k < r$ let

$$\delta(k) = \sum_{a_1^{r+1}} \pi(a_1^r) p(a_{r+1}|a_1^r) \log \left\{ \frac{p(a_{r+1}|a_1^r)}{q(a_{r+1}|a_{r-k+1}^r)} \right\} \quad (3.4)$$

where $q(a_{r+1}|a_{r+1}^r) = \sum_{a_2^r} \pi(a_2^{r+1})$ and

$$q(a_{r+1}|a_{r-k+1}^r) = \frac{\sum_{a_2^{r-k}} \pi(a_2^{r+1})}{\sum_{a_2^{r-k}} \sum_{a_{r+1}} \pi(a_2^{r+1})}, \quad 0 < k < r.$$

Lemma 3. Assume that $Y^{(r)}$ is ergodic.

(i) If $k \geq r$ then

$$\log \hat{L}(k) - \log L(k) \geq o(\delta_n) \quad \text{and} \quad \limsup_{n \rightarrow \infty} \frac{\log \hat{L}(k) - \log L(k)}{\log \log n} = \gamma_+(k) \quad \text{a.s.} \quad (3.5)$$

(ii) If $0 \leq k < r$ then $\delta(k) > 0$, $\delta(k) \geq \delta(k+1)$ and

$$\lim_{n \rightarrow \infty} \frac{\log \hat{L}(r) - \log \hat{L}(k)}{n} = \delta(k) \quad \text{a.s.} \quad (3.6)$$

Proof. (a) For $k \geq 0$ we have

$$\log \hat{L}(k+1) - \log \hat{L}(k) \geq o(\delta_n). \quad (3.7)$$

From (1.3) and Jensen's inequality,

$$\begin{aligned} \log \hat{L}(k+1) - \log \hat{L}(k) &= \sum_{a_1^{k+2}} N(a_1^{k+2}|X_1^n) \log \left\{ \frac{N(a_1^{k+2}|X_1^n)}{N(a_1^{k+1}|X_1^n)} \cdot \frac{N(a_2^{k+1}|X_1^n)}{N(a_2^{k+2}|X_1^n)} \right\} + o(\delta_n) \\ &\geq - \sum_{a_1^{k+1}} N(a_1^{k+1}|X_1^n) \log \left\{ \sum_{a_{k+2}} \frac{N(a_2^{k+2}|X_1^n)}{N(a_2^{k+1}|X_1^n)} \right\} + o(\delta_n) = o(\delta_n). \end{aligned}$$

(b) From (3.1) we can write for $k \geq r$

$$\log \hat{L}(k) - \log L(k) = - \sum_{a_1^{k+1}} N(a_1^{k+1}|X_1^n) \log(1 + z_n(a_1^{k+1}))$$

where

$$z_n(a_1^{k+1}) = \frac{N(a_1^k|X_1^n) p(a_{k+1}|a_1^k) - N(a_1^{k+1}|X_1^n)}{N(a_1^{k+1}|X_1^n)}.$$

Using Large Numbers for Markov chains (see, for example, Dacunha-Castelle and Duflo (1986)) we have

$$\frac{N(a_1^k | X_1^n) p(a_{k+1} | a_1^k)}{N(a_1^{k+1} | X_1^n)} \xrightarrow{a.s.} 1 \quad \text{and} \quad z_n(a_1^{k+1}) \xrightarrow{a.s.} 0.$$

Now, we use the inequality

$$-z + \frac{\beta - 1}{2\beta} z^2 \leq -\log(1 + z) \leq -z + \frac{\beta + 1}{2\beta} z^2,$$

where $\beta > 1$ and $-\frac{1}{\beta + 1} < z < \frac{1}{\beta - 1}$. Since for all $a_1^{k+1} \in E^{k+1}$ and n large we have $z_n(a_1^{k+1})$ small,

$$\sum_{a_1^{k+1}} N(a_1^{k+1} | X_1^n) z_n(a_1^{k+1}) = o(\delta_n)$$

and we can write as $\beta \rightarrow \infty$

$$\log \hat{L}(k) - \log L(k) = \frac{1}{2} \sum_{a_1^{k+1}} N(a_1^{k+1} | X_1^n) z_n^2(a_1^{k+1}) + o(\delta_n).$$

Since $\frac{N(a_1^{k+1} | X_1^n)}{n} \xrightarrow{a.s.} \pi(a_1^{k+1})$ we have from Lemma 2,

$$\limsup_{n \rightarrow \infty} \frac{\log \hat{L}(k) - \log L(k)}{\log \log n} = \sum_{a_1^{k+1} \in E_+^{k+1}} (1 - p(a_{k+1} | a_1^k)) = \gamma_+(k). \quad \text{a.s.}$$

(c) Let $0 \leq k < r$. Since r is the true order we cannot have $p(a_{r+1} | a_1^r) = q(a_{r+1} | a_{r-k+1}^r)$ for all $a_1^{r+1} \in E^{r+1}$. From (3.4) we have

$$\delta(k) = - \sum_{a_1^r} \pi(a_1^r) \sum_{a_{r+1}} p(a_{r+1} | a_1^r) \log \left\{ \frac{q(a_{r+1} | a_{r-k+1}^r)}{p(a_{r+1} | a_1^r)} \right\}$$

and from Jensen's inequality

$$\delta(k) > - \sum_{a_1^r} \pi(a_1^r) \log \left\{ \sum_{a_{r+1}} q(a_{r+1} | a_{r-k+1}^r) \right\} = 0.$$

From (3.2) we have

$$\log \hat{L}(k) - \log \hat{L}(r) = \sum_{a_1^{r+1}} N(a_1^{r+1} | X_1^n) \log \left\{ \frac{N(a_1^r | X_1^n)}{N(a_1^{r+1} | X_1^n)} \cdot \frac{N(a_{r-k+1}^{r+1} | X_1^n)}{N(a_{r-k+1}^r | X_1^n)} \right\} + o(\delta_n).$$

By the Law of Large Numbers,

$$\frac{N(a_{r-k+1}^{r+1}|X_1^n)}{N(a_{r-k+1}^r|X_1^n)} \xrightarrow{a.s.} \frac{\sum_{a_2^{r-k}} \pi(a_2^{r+1})}{\sum_{a_{r+1}} \sum_{a_2^{r-k}} \pi(a_2^{r+1})} = q(a_{r+1}|a_{r-k+1}^r)$$

and since E is finite we have

$$\lim_{n \rightarrow \infty} \frac{\log \hat{L}(k) - \log \hat{L}(r)}{n} = -\delta(k) \text{ a.s.}$$

Using (3.7) and writing

$$\frac{\log \hat{L}(k+1) - \log \hat{L}(r)}{n} = \frac{\log \hat{L}(k+1) - \log \hat{L}(k)}{n} + \frac{\log \hat{L}(k) - \log \hat{L}(r)}{n}$$

we conclude that $\delta(k) \geq \delta(k+1)$. \square

Theorem 1. Assume that E contains at least two elements and that for all $a_1^{r+1} \in E^{r+1}$ we have $p(a_{r+1}|a_1^r) > 0$. Moreover, assume that c_n satisfies

$$\liminf_{n \rightarrow \infty} \frac{c_n}{\log \log n} \geq \frac{2|E|}{|E|-1} \text{ and } \limsup_{n \rightarrow \infty} \frac{c_n}{n} = 0. \quad (3.8)$$

Then,

$$\limsup_{n \rightarrow \infty} \frac{\text{EDC}(k) - \text{EDC}(r)}{\log \log n} \geq 2\gamma(r) \text{ a.s., } k > r \quad (3.9)$$

and the lim sup is increasing on k . Also,

$$\lim_{n \rightarrow \infty} \frac{\text{EDC}(k) - \text{EDC}(r)}{n} = 2\delta(k) \text{ a.s., } 0 \leq k < r \quad (3.10)$$

and the limit is decreasing on k .

Proof. (a) Since all entries of the transition probabilities are strictly positive we have $Y^{(r)}$ ergodic.

(b) Let $k > r$. From (3.3) we have

$$\gamma_+(k) = \gamma(k) = |E|^k(|E|-1) > 0.$$

From (1.5) and (3.1) we can write

$$\begin{aligned} \text{EDC}(r+1) - \text{EDC}(r) &= -2[\log \hat{L}(r+1) - \log L(r+1)] + 2[\log \hat{L}(r) - \log L(r)] \\ &\quad + 2\gamma(r)(|E|-1)c_n + o(\delta_n). \end{aligned}$$

If c_n satisfies (3.8) we have from Lemma 3

$$0 \leq \liminf_{n \rightarrow \infty} \frac{\log \hat{L}(r+1) - \log L(r+1)}{\log \log n} \leq 2\gamma(r+1) \text{ a.s.}$$

and

$$\limsup_{n \rightarrow \infty} \frac{\text{EDC}(r+1) - \text{EDC}(r)}{\log \log n} \geq 2\gamma(r) \text{ a.s.}$$

Similarly, for $k \geq r+1$ we have

$$\limsup_{n \rightarrow \infty} \frac{\text{EDC}(k+1) - \text{EDC}(k)}{\log \log n} \geq 2\gamma(k) \text{ a.s.}$$

Thus, the monotonicity and (3.9) follows.

(c) Let $0 \leq k < r$. Since $\frac{c_n}{n} \xrightarrow{n} 0$ we have from (3.10)

$$\lim_{n \rightarrow \infty} \frac{\text{EDC}(k) - \text{EDC}(r)}{n} = 2\delta(k) \text{ a.s.}$$

with $\delta(k) \geq \delta(k+1) > 0$. \square

Note that for the AIC estimate we have $c_n = 2$ and

$$\liminf_{n \rightarrow \infty} \frac{\text{AIC}(k) - \text{AIC}(r)}{\log \log n} \leq -2\gamma(k) + 2\gamma(r) < 0, \quad k > r. \quad (3.11)$$

So that AIC is not consistent and overestimation of the true order could occur.

As for the BIC estimate we have $c_n = \log n$ and the contribution of the penalty term is larger than necessary. Let $k < r$, then, if n is not large enough we could have

$$(\gamma(r) - \gamma(k)) \log n > (-2 \log \hat{L}(k) + 2 \log \hat{L}(r))$$

and since $(\gamma(k) - \gamma(r)) \log n < 0$ this may lead to possible underestimation. To avoid preponderancy of the penalty term one needs to take c_n smaller than $\log n$ but sufficiently large to guarantee the consistency of the estimator. Our Theorem 1 indicates that c_n can be taken of the form $c_n = \alpha \log \log n$, where the constant α satisfies $\alpha \geq \frac{2|E|}{|E| - 1}$. In this case, for $k < r$ we have

$$\text{EDC}(k) - \text{EDC}(r) > \text{BIC}(k) - \text{BIC}(r). \quad (3.12)$$

It remains to determine the optimal choice of α . Note that, for $c_n = \alpha \log \log n$ we have

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{\text{EDC}(r+1) - \text{EDC}(r)}{\log \log n} &= -2\gamma(r+1) + 2 \liminf_{n \rightarrow \infty} \frac{\log \hat{L}(r) - \log L(r)}{\log \log n} \\ &+ \alpha(\gamma(r+1) - \gamma(r)). \end{aligned}$$

Since $2(\log \hat{L}(r) - \log L(r))$ has asymptotic χ^2 distribution (cf. Billingsley (1961)) we necessarily have $\liminf_{n \rightarrow \infty} \frac{\log \hat{L}(r) - \log L(r)}{\log \log n} < \gamma(r)$. Now, to assure

$$-2\gamma(r+1) + \alpha(\gamma(r+1)) - \gamma(r) \geq 0$$

we need $\alpha \geq \frac{2|E|}{|E|-1}$. This leads to the optimal choice $c_n = \frac{2|E|}{|E|-1} \log \log n$ and the EDC takes a particularly simple form

$$\text{EDC}_{opt}(k) = -2 \log \hat{L}(k) + 2|E|^{k+1} \log \log n. \quad (3.13)$$

Corollary 1. Under conditions of Theorem 1, with $c_n = \alpha \log \log n$ for some $\alpha > 0$, the optimal penalty term is given by $2|E|^{k+1} \log \log n$ and the optimal estimator is given by (3.13).

Our Theorem 1 also shows that the consistency of EDC criterion without the finite boundness assumption for the true order and under weaker conditions than (1.6).

Corollary 2. If $Y^{(r)}$ is ergodic and (3.8) holds then the BIC and EDC estimates are consistent. Moreover, (3.9) and (3.10) hold with $\gamma_+(r)$ in place of $\gamma(r)$.

References

- [1] Akaike, H. (1974) - *A new look at the statistical model identification*, IEEE Trans. Automation and Control, vol. 19, 716-723.
- [2] Billingsley, P. (1961) - *Statistical Methods in Markov Chains*, Annals of Math. Statistics, vol. 32, 12-40.
- [3] Csiszár, I. and Shields, P.C. (2000) - *The consistency of the BIC Markov order estimator*, Annals of Statistics, vol. 28, 1601-1619.
- [4] Dacunha-Castelle, D. and Duflo, M. (1986) - *Probability and Statistics*, vol. 2, Springer-Verlag, New York.
- [5] Dorea, C.C.Y. and Zhao, L.C. (2006) - *Exponential bounds for the probability of wrong determination of the order of a Markov chain by using the EDC criterion*, Journal of Statistical Planning and Inference (to appear).
- [6] Finesso, L.; Liu, C.C. and Narayan, P. (1996) - *The optimal error exponent for Markov order estimation*, IEEE Trans. Information Theory, vol. 42, 1488-1497.
- [7] Katz, R.W. (1981) - *On some criteria for estimating the order of a Markov chain*, Technometrics, vol. 23, 243-249.

- [8] Meyn, S.P. and Tweedie, R.L. (1993) - *Markov Chains and Stochastic Stability*, Springer-Verlag, London.
- [9] Roussas, G.G. and Ioannides, D. (1987) - *Moment inequalities for mixing sequences of random variables*, Stochastic Analysis and Applications, vol. 5, 61-120.
- [10] Schwarz, G. (1978) - *Estimating the dimension of a model*, Annals of Statistics, vol. 6, 461-464.
- [11] Tong, H. (1975) - *Determination of the order of a Markov chain by Akaike's information criterion*, Jour. of Applied Probability, vol. 12, 488-497.
- [12] Zhao, L. C., Dorea, C.C.Y. and Gonçalves, C.R. (2001) - *On determination of the order of a Markov Chain*, Statistical Inference for Stochastic Processes, vol. 4, 273-282.

Università degli Studi di Torino, Italy

e-mail: ignaccolobescon@unito.it

²Dipartimento di Statistica

Università degli Studi di Bari, Italy

e-mail: rlabacco@dis.uniba.it

Abstract

The distribution of a variable observed over a domain depends on the underlying process and also on the geographical location at which the variable has been measured. In this paper, we fit a model to the distribution supposing that the observations are generated by a stationary strong-mixing random field. Indeed, after estimating the density of the considered variable, we construct a test statistic in order to verify the goodness of fit of the observed spatial data. The proposed class of tests is a generalization of the classical chi-square test and of the Neyman smooth test. In the framework of increasing domain asymptotics, we analyse the large sample behaviour of the test. The limiting distribution is a linear combination of χ^2 c.v.s where the coefficients are the eigenvalues of a matrix Σ essentially related to the spectral density of the random field. Finally some indications about the implementation are provided.

Keyword: goodness of fit, correlated data, spatial process, mixing random field