INFORMATION CRITERIA BASED EDGE DETECTION

Frédéric Jouzel [1] *Christian Olivier* [2] *Abdelaziz El Matouat* [3]

[1] PSI-La3i,
University of Rouen, France
Frederic.Jouzel@univ-rouen.fr

[2] SIC-IRCOM, UMR CNRS 6615,
University of Poitiers, France
Olivier@sic.univ-poitiers.fr

[3] Ecole Normale Supérieure, Fès
BP 5206, Bensouda, Fès, Morocco
& Faculté des Affaires Internationales,
university of Le Havre, France

ABSTRACT

We propose an algorithm to detect ruptures in a signal by means of information criteria along with its application to edge detection on grey-level images. Information criteria are made of two terms : a likelihood term and an original penalization factor. They enable us to find in an optimal way autoregressive models and their number to model the signal. The model changes represent the edges of the objects in the image.

1 INTRODUCTION

Let us consider a parameterized model characterized by the probability density function (PDF) $f(\cdot \mid \theta^*)$ in which the "true" parameter vector θ^* and its dimension k^*, called model order, are not known. The well-known techniques such as maximum likelihood (ML) allow the estimation of θ^*, when the order is known, but the problem of order estimation remains more difficult. The ML principle generally leads to an over-parameterization of the model. A penalization of the log-likelihood can palliate this drawback. The most current information criterion is Akaike's criterion AIC [1], though it is not satisfactory: it improves the ML principle but leads to a strict overparameterization of the model order. AIC is based on the minimization of the Kullback-Leibler information between $f(\cdot \mid \theta^*)$ and the PDF $f(\cdot \mid \theta_k)$ specified by the parameter vector θ_k with dimension k. Hannan and Quinn's criterion [5] makes a further improvement by an order estimation which is convergent in probability. An original generalization of the latter criterion based on *stochastic complexity* [7] is introduced in [4]. It leads to criteria noted as φ_β. These criteria ensure an almost sure convergence. The behaviour of these criteria is compared with the classical strongly consistent BIC (Bayesian Information Criterion) [8], also called MDL for Minimum Description Length [6], on an edge detection problem in time series under the *a priori* hypothesis of an autoregressive (AR) modelisation.

In the following part, we sum up all the forms of the different information criteria mentioned above. In a second part, we give their particular expression in the AR case and we show how they behave on Monte-Carlo simulations. In the third part, given a signal which evolution corresponds to a succession of unknown models, we seek to segment this signal in c blocks, each block $i, i = 1, ..., c$ being modelized by an AR model of order k_i. The matter is therefore to determine the number of model changes $c-1$ and their position, as well as the optimal orders k_i estimated by the criteria presented above. The detection method proposed with AR models is an extension of classical techniques generally used for gaussian mixture models (see [10]).

2 MODEL SELECTION CRITERIA

For model order selection, the most known criterion is surely Akaike's criterion, written as follows:

$$AIC(k) = -2\sum_{i=1}^{N} \log f\left(X_i \mid \hat{\theta}_k\right) + 2k \qquad (1)$$

where $\hat{\theta}_k$ is the ML estimator for the unknown parameter θ_k based on the sample $X_1, ..., X_N$, and the order choice is such that $\hat{k} = \arg\min_k AIC(k)$. Theoretical research in the Akaike's criterion has helped to specify the asymptotic behaviour of AIC. Akaike's criterion is then unsatisfactory since it asymptotically leads to a strictly positive overparametrization probability of the model order [9].

In order to palliate the inconsistency of Akaike's criterion, G. Schwarz (1978) proposed a new criterion for an exponential family founded on a bayesian justification. He suggested the BIC (Bayesian Information Criterion) [8]:

$$BIC(k) = -2\sum_{i=1}^{N} \log f\left(X_i \mid \hat{\theta}_k\right) + k \log N \qquad (2)$$

In a different way, J. Rissanen (1978) came up with an equivalent criterion using a coding technique (minimising the codelength in relation to the observations) for a parametrized density, which is referred as MDL (Minimum Description Length) principle. This criterion is asymptotically convergent in that it helps in finding the appropriate model when $N \to \infty$ (strong consistency). Note that the latter criterion penalises more stringently the log-likelihood as the number of observations increases in comparison with AIC.

A third criterion was introduced by E. J. Hannan and B. G. Quinn [5] in the case of an autoregressive process. It substitutes $k \log\log N$ for the preceding penalty and leads to convergence in probability of the order estimator (weak consistency); This criterion is written as φ and stands as a compromise between AIC and BIC.

Finally, let us give A. El Matouat and M. Hallin's generalisation [4] drawn on Rissanen's works [7] ending up in a criterion written as φ_β in the general case of parametrized PDF. For a sequence of observations $X_1, \ldots, X_N$, the criterion φ_β is written:

$$\varphi_\beta(k) = -2\sum_{i=1}^{N} \log f(X_i \mid \hat{\theta}_k) + kN^\beta \log\log N, 0 < \beta < 1 \tag{3}$$

We can observe that Hannan and Quinn's criterion φ is present in a more general form. Indeed, as $\varphi_\beta(k) - \varphi(k)$ leads to zero in probability when β does so, the φ criterion can be seen as a limit case of φ_β ($\beta = 0$).

Then, the general form of Information Criterion (IC) can be summed up as:

$$IC(k) = -2\log ML(k) + \alpha(k)c_N \tag{4}$$

where $ML(.)$ denotes the maximized likelihood of the model of dimension k and $\alpha(k)$ denotes the number of free parameters in the model we are considering. The possible penalities are $c_N = \log N$ for BIC, $c_N = \log\log N$ for φ (Hannan and Quinn's criterion) and $c_N = N^\beta \log\log N, \beta \in]0,1[$ for φ_β. We can notice that some authors introduce the MMDL criterion (for Modified MDL) without theoretical justification. For this criterion, $c_N = 5\log N$.

3 EXPERIMENTAL COMPARISON ON AR MODELS

Now, we propose to apply the several model selection criteria mentioned in the preceding section to the estimation of the order of AR models. There are many studies involving the traditional criteria AIC, BIC and φ in the AR model case [3, 11]. Let be a time series $X_1, ..., X_N$ modelized by an AR model of order k:

$$\begin{cases} X_t = -\sum_{i=1}^{k} a_i X_{t-i} + e_t \\ E(e_t) = 0, E(e_s e_t) = \sigma_e^2 \delta_{st} \end{cases}$$

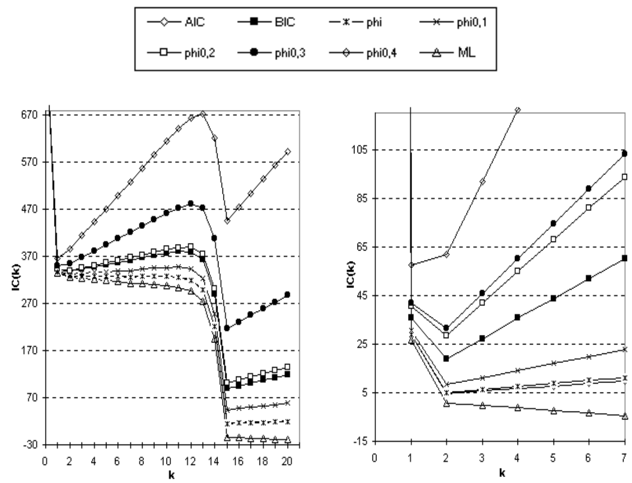
where δ_{st} is the Kronecker's symbol and $e^N = e_1, ..., e_N$ is a gaussian white noise process with variance σ_e^2. Using notations of part 1, $\theta_k = (a_1 \ldots a_k)^t$. Omitting terms that do not depend on k, it is well-known that the likelihood term in (4) becomes $N \log \widehat{\sigma}_e^2$, where $\widehat{\sigma}_e$ is the ML estimate of σ_e. Thus, we obtain:

$$IC(k) = N\log\widehat{\sigma}_e^2 + c_N k \tag{5}$$

The selected order $\hat{k}$ verifies $\hat{k} = \arg\min_k IC(k)$. We verify the criteria behaviour on synthetic signals using Monte-Carlo simulations. The two chosen models for signal synthesis are as follows[1]:

- AR(2) model: $a_1 = -0.55$ and $a_2 = -0.05$;
- AR(15) model: $a_1 = -0.50, a_2 = -0.06, a_{15} = -0.45$ and $a_i = 0$ for the remaining parameters.

The search for the order is repeated on 100 replications of the experiment, and the tested orders range from 0 to 20. For $N = 1000$ observations and $\sigma_e = 1$, we obtain curves averaged on 100 experiments (see Fig. 1).



(a) (b)

Figure 1: Averaged criteria values for the AR(15) model (1(a)) and the AR(2) model (1(b))

Figure 1 clearly shows the best quality of the order selection allowed by BIC and φ_β. Table 1 gives the frequencies of the obtained orders, $0 \leq k \leq 20$, by the different criteria for the AR(15) process.

4 EDGE DETECTION

Consider the signal $X_1, ..., X_N$. Let c be an *a priori* fixed number of models (adjacent blocks), and let us consider the following conditional mixture model:

$$ML(c) = \prod_{i=1}^{c} f(X_{t_{i-1}+1}, ..., X_{t_i} \mid \hat{\theta}_i) \tag{6}$$

where $t_0 = 0$ and $t_c = N$. The edges positions $t_i, i = 1, ..., c-1$ are estimated using a *dynamic programming* algorithm described in [10]. For an AR model,

[1] These AR models are often considered in litterature, for instance in [2, 11].

Table 1: Results for AR(15), $\sigma_e = 1$

$N = 1000$

order	AIC	BIC	φ	$\varphi_{0.2}$	$\varphi_{0.5}$
0	0	0	0	0	0
1	0	0	0	0	100
...	...	...	...	...	...
15	**50**	**82**	**49**	**84**	**0**
16	9	5	9	3	0
17	9	7	7	7	0
18	12	3	14	4	0
19	14	3	15	2	0
$\geq$20	6	0	6	0	0

$N = 10000$

order	AIC	BIC	φ	$\varphi_{0.2}$	$\varphi_{0.5}$
0	0	0	0	0	0
1	0	0	0	0	0
...	...	...	...	...	...
15	**65**	**100**	**74**	**100**	**100**
16	13	0	11	0	0
17	7	0	6	0	0
18	6	0	5	0	0
19	2	0	1	0	0
$\geq$20	7	0	3	0	0

$f(X_{t_{i-1}+1}, ..., X_{t_i} \mid \hat{\theta}_i) = (2\pi e \widehat{\sigma}_i^2)^{-\frac{n_i}{2}}$ where n_i is the length of block i and $\widehat{\sigma}_i^2$ is the ML estimate of the prediction error variance of the model with k_i parameters. Finally, the criteria to minimize are the following:

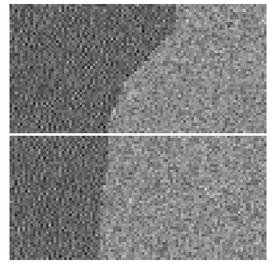
$$IC(c) = -2 \log ML(c) + \alpha(c) c_N \qquad (7)$$

and the optimal number of models is estimated by $\hat{c} = \arg \min_{t_1,...,t_c-1,k_1,...,k_c,c} IC(c)$. From processing time arguments, we chose to implement the following approximation : $\hat{c} = \arg \min_{t_1,...,t_c-1,k_1=\hat{k}_1,...,k_c=\hat{k}_c,c} IC(c)$, where the AR model orders $\widehat{k}_i, i = 1, ..., c$ are obtained by minimizing the IC given in (5) on each block. Thus, we have : $\alpha(c) = \sum_{i=1}^{c} \widehat{k}_i + c - 1$. Figure 2(a) presents a synthetised image[2]. Figure 2(b) shows the behaviour of the criteria given in (7) on a line of the original 256 grey-level image (Fig. 2(c)). The better minimum visualization of the criteria $\varphi_\beta, \beta = 0.3, 0.4, 0.5$, justifies their benefit.

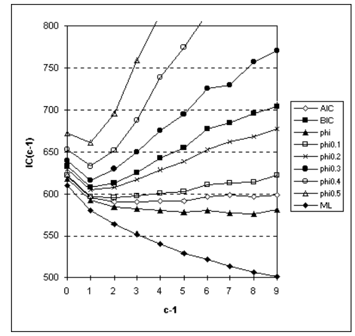
Figure 3 shows the edges estimated by the several criteria, the algorithm being repeated for each line. We can notice that BIC and $\varphi_\beta, \beta = 0.2, 0.5$ allow us to obtain edges in a precise way (good edge localization), avoiding edge overdetection. AIC leads to edge overdetection.

Figure 4 shows a natural image (cross section of a heart), and the corresponding edge detection by the proposed method, the algorithm being repeated for each line and each column. A comparative study of the proposed algorithm with classical edge detection operators (Canny-Deriche, Sobel, Roberts, ...) shows the better detections obtained by the proposed approach.

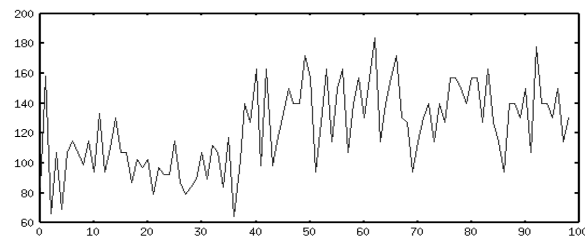
[2] each line of the image is made of a AR(2) model and a AR(4) model with different prediction error variance



(a)



(b)



(c)

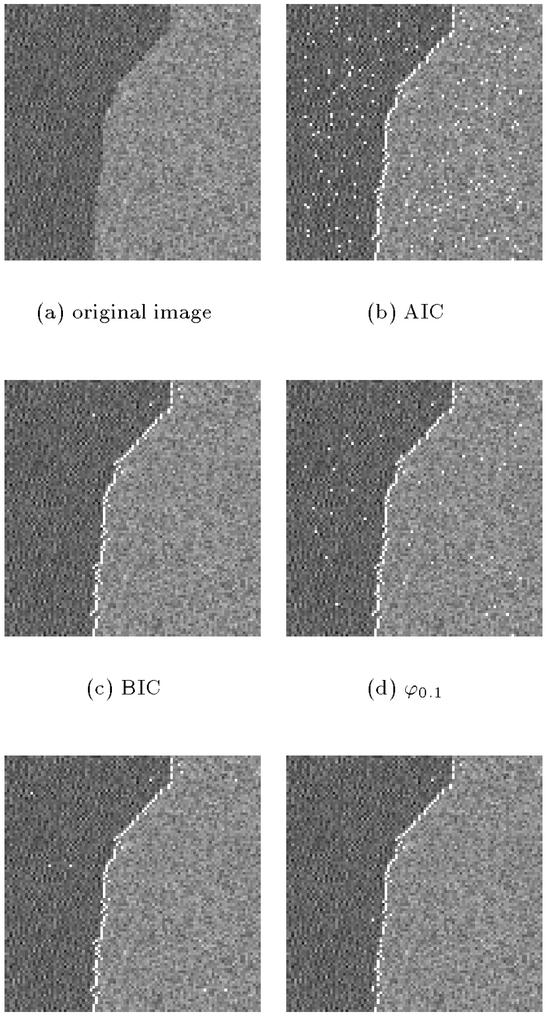
Figure 2: number of edges detected by the different criteria (2(b)) for a line of the original image (2(c))

5 CONCLUSION

In this paper, we verified the asymptotic behaviours of three information criteria in the AR model case with aim of edge detection on digital images, and we proposed a new criterion to achieve this task. We showed the good behaviour of BIC and $\varphi_\beta, \beta \leq 0.5$ for simulated and real data when N is large enough. The approach that we propose can also be extended to more complex models such as ARMA models or multivariate AR processes (multi-spectral images or color images for example). Finally, the criteria developped here can be used in various mixture model contexts for classification tasks involving the EM algorithm.

References

[1] H. Akaike. Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Cáski, editors, *Second International Symposium on Information Theory*, pages 267–281, Budapest, 1973. Akademiai Kaidó.



(a) original image (b) AIC

(c) BIC (d) $\varphi_{0.1}$

(e) $\varphi_{0.2}$ (f) $\varphi_{0.5}$

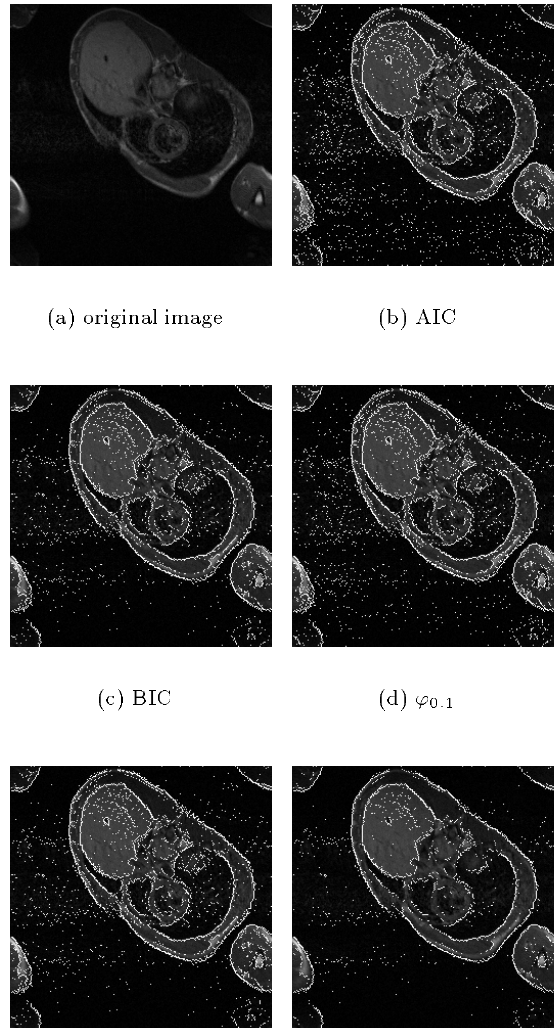
Figure 3: Edge detection on a synthetic image

[2] R.J. Bhansali and D.Y. Downham. Some properties of the order of an autoregressive model selected by a generalization of Akaike's FPE criterion. *Biometrika*, 64:547–551, 1977.

[3] J.R. Dickie and A.K. Nandi. A comparative study of AR order selection methods. *Signal Processing*, 40(2):239–256, November 1994.

[4] A. El Matouat and M. Hallin. Order selection, stochastic complexity and Kullback-Leibler information. In P.M. Robinson and M. Rosenblatt, editors, *Time Series Analysis*, volume 2, in memory of E.J. Hannan, pages 291–299. Springer Verlag, New York, 1996.

[5] E .J. Hannan and B. G. Quinn. The determination of the order of an autoregression. *Journal of the Roy. Stat. Soc. B*, 41(2):190–195, 1979.



(a) original image (b) AIC

(c) BIC (d) $\varphi_{0.1}$

(e) $\varphi_{0.2}$ (f) $\varphi_{0.3}$

Figure 4: Edge detection on a real image

[6] J. Rissanen. Modeling by shortest data description. *Automatica*, 14:465–471, 1978.

[7] J. Rissanen. *Stochastic Complexity in Statistical Inquiry.* World Scientific, N.J., 1989.

[8] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6:461–464, 1978.

[9] R. Shibata. Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika*, 63:117–126, 1976.

[10] M. Thune, B. Olstad, and N. Thune. Edge detection in noisy data using finite mixture distribution analysis. *Pattern Recognition*, 30(5):685–699, 1997.

[11] T. Van Eck. On objective autoregressive model testing. *Signal Processing*, 10:185–191, 1986.