Order Estimation of Superimposed Nonlinear Complex Cisoid Model Using Adaptively Penalizing Likelihood Rule: Consistency Results

Sharmishtha Mitra[*] and Anupreet Porwal

Departnment of Mathematics and Statistics,Indian Institute of Technology Kanpur,
Kanpur 208016, India

*Corresponding author

Keywords: Bayesian information criterion, Consistency, Order estimation, Penalizing adaptively the likelihood.

Abstract. Recently a novel approach of model order selection based on penalizing adaptively the likelihood (PAL) function was introduced in [1]. In this paper, we use the PAL method for order estimation of complex valued nonlinear exponential (cisoid) model and study its asymptotic statistical properties. We investigate the asymptotic statistical properties for the 1-dimensional cisoid model under the assumption of circularly symmetric gaussian error distribution and establish that the PAL estimator is consistent. We also present simulation examples to compare the performance of PAL rule with the commonly used information criteria based rules.

Introduction

Model order selection is a fundamental task in time series analysis and signal processing as once the integer valued parameter model order is estimated, we know the complexity of the model. The parameters characterizing the model can only be estimated after the model order is estimated. The most widely used approach for order estimation in such a situation is the one based on information theoretic criteria, the AKAIKE Information Criterion (AIC) and the Bayesian Information Criterion (BIC) being the most prominent ones. For a review of the various order estimation methods, see [1]-[3] and the references cited therein.

Under a general model order selection framework, many of the popular rules used for estimation of model order have the following form:

$$-2\ln f_m(y,\hat{\theta}_m) + \rho(n)\,m \tag{1}$$

where $\rho(n)\,m$ denotes the penalty associated with model order which may depend on sample size n and m, where m denotes the model order and $f_m(y,\hat{\theta}_m)$ denotes the probability density function under the hypothesis that y is generated from a model with dimension of the signal parameter vector being m. $\hat{\theta}_m \in \mathbb{R}^{m \times 1}$ being the maximum likelihood estimate of the parameter vector given y. The different penalty terms gives rise to different order selection rules: $AIC: \rho(n) = 2;$ $BIC: \rho(n) = \ln n.$

In this paper, we propose to use the novel method introduced in [1] using a data adaptive penalty and having oracle-like properties for cisoid models and prove that the estimators of model order using PAL rule is consistent. We also present some simulation results for a 1-d cisoid model and compare the performance of PAL with other widely used information criterion rules.

PAL Rule and Its Consistency for 1-d Complex Cisoid Model

We consider the problem of estimating the number of components of the following complex cisoid model:

$$y_t = \sum_{k=1}^{m} \alpha_k e^{it\omega_k} + \varepsilon_t, \quad t = 1, 2, ..., n.$$

(2)

$\theta_m = (\alpha_{1_R}, \alpha_{1_C}, \omega_1, ..., \alpha_{m_R}, \alpha_{m_C}, \omega_m)'$ is a $3m$x1 vector of unknown signal parameters; α_{j_R} and α_{j_C} denote the real and imaginary parts of $\alpha_j, j=1,..., m$.

Let m_o be the true number components in the the observed signal. Given a sample of size n, $y = (y_1, y_2, ..., y_n)'$, the model order estimation problem is to estimate m_o. For establishing the consistency results of the paper, we make the following assumptions:

Assumption A1: $\{\varepsilon_t\}$ is a sequence of i.i.d complex valued gaussian random variables with zero mean such that $\varepsilon_t = \varepsilon_{t_R} + i\varepsilon_{t_C}; \varepsilon_{t_R} \sim N(0, \sigma^2/2), \varepsilon_{t_C} \sim N(0, \sigma^2/2)$ and are independent.

Assumption A2: $\forall k = 1, 2, ..., m_o : \omega_k \in (0, 2\pi) \quad ; \quad \omega_j \neq \omega_k, \forall j \neq k$. Furthermore, $\forall k = 1, 2..., m_o : \alpha_k$'s are bounded.

Assumption A3: The true model parameter vector θ_{m_o} is an interior point in the parameter space $\Theta \subset \mathbb{R}^{3m_o}$.

Let $y = (y_1, y_2, ..., y_n)'$, $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)'$ and for an m-component model, let $\theta_m^* = (\theta_m, \sigma_m^2)'$ denote the vector containing the underlying signal and noise parameters, then the p.d.f. of y under these assumptions can be written as

$$f_m(y, \theta_m^*) = (2\pi\sigma_m^2/2)^n \exp\left(-\frac{(y - f(\theta_m))^H (y - f(\theta_m))}{\sigma_m^2}\right).$$

(3)

$f(\theta_m) = (f(1, \theta_m), f(2, \theta_m), ..., f(n, \theta_m))'$ and $f(t, \theta_m) = \sum_{k=1}^{m} \alpha_k e^{it\omega_k}$. We consider the set of $\tilde{m}$ nested models given by $\{M_m\}_{m=1}^{\tilde{m}}$ where M_m is the m component cisoid model with parameter vector θ_m^*. We assume that the true model M_{m_o} is contained in this set .i.e., $m_o \leq \tilde{m}$ and y is not completely a white noise process, i.e., $M_0 \neq M_{m_o}$. Consider the two generalized likelihood ratios,

$$r_m = 2\ln\left(\frac{f_{m-1}(y, \hat{\theta}_{m-1}^*)}{f_0(y, \hat{\theta}_0^*)}\right) \quad \text{and} \quad \rho_m = 2\ln\left(\frac{f_{\tilde{m}}(y, \hat{\theta}_{\tilde{m}}^*)}{f_{m-1}(y, \hat{\theta}_{m-1}^*)}\right),$$

(4)

where $\hat{\theta}_k^*$ is the MLE of underlying signal and noise parameter vector θ_k^* and $f_0(y, \hat{\theta}_0^*)$ denotes the p.d.f of y when M_0 is the model, i.e. $f(\theta_0) = 0$. The PAL function and the PAL rule based estimator can then be defined using the GLR ratios as follows:

$$PAL(m) = -2\ln\left(f_m(y, \hat{\theta}_m^*)\right) + (3m+1)\ln(3\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)}$$

(5)

$$\hat{m} = \arg\min_{m \in \{1, 2, ..., \tilde{m}\}} PAL(m).$$

(6)

Remark 1: Realize that the ratio $\dfrac{\ln(r_m+1)}{\ln(\rho_m+1)}$ in the PAL(m) function is such that (i) it is an increasing function of m, (ii) at $m = 1$, $r_1 = 0$ hence the ratio is 0 and (iii) the ratio is $> 0 \quad \forall m \geq 2$.

To prove the consistency of the PAL rule for complex cisoid model we need the following

lemmas.   The lemmas can be easily proved using the results in [3,4,5]

Lemma 1: Under the assumptions A1-A3, $\forall m \leq m_o$,

$$\hat{\sigma}_m^2 = \sigma^2 + \sum_{j=1}^{m_o} \alpha_j^H \alpha_j - \sum_{j=1}^{m} \hat{\alpha}_j^H \hat{\alpha}_j + o(1) \text{ a.s. as } n \to \infty.$$

Lemma 2: Under A1-A3, for any integer $k \geq 1$

$$\hat{\sigma}_{m_o+k}^2 = \hat{\sigma}_{m_o}^2 - \frac{G_k}{n} + o\left(\frac{\ln n}{n}\right) \text{ a.s. as } n \to \infty$$

where $G_k = \sum_{j=1}^{k} I_\varepsilon\left(\hat{\omega}_{m_o+j}\right)$ and $I_\varepsilon(\omega) = \frac{1}{n}\left|\sum_{t=1}^{n} \varepsilon_t e^{-it\omega}\right|^2$. $I_\varepsilon(\omega)$ corresponds to periodogram of underlying white noise process and $\hat{\omega}_{m_o+1}, \hat{\omega}_{m_o+2}, ..., \hat{\omega}_{m_o+k}$ are the k largest frequencies corresponding to $I_\varepsilon(\omega)$.   Thus, G_k is the sum of k largest elements of the periodogram of noise.

Lemma 3: Under assumptions A1-A3, r_m satisfies

$$r_m = \begin{cases} 0, & m = 1, \\ O(n), & 2 \leq m \leq \tilde{m}. \end{cases} \tag{7}$$

Lemma 4: Under the assumptions A1-A3, ρ_m satisfies

$$\rho_m = \begin{cases} O(n), & m \leq m_o, \\ O_p(1), & m = m_o+1, m_o+2, \ldots, \tilde{m}. \end{cases} \tag{8}$$

Remark 2: Using Lemma 3 and Lemma 4, the ratio $\dfrac{\ln(r_m+1)}{\ln(\rho_m+1)}$, therefore, satisfies

$$\frac{\ln(r_m+1)}{\ln(\rho_m+1)} = \begin{cases} O(1), & m \leq m_o, \\ O(\ln n), & m = m_o+1, m_o+2, \ldots, \tilde{m}. \end{cases} \tag{9}$$

Theorem: Under A1-A3, if m_o is true order ($m_o \leq \tilde{m}$) and $\hat{m}$ is the estimated model order using PAL rule then

$$P(\hat{m} \neq m_o) \to 0 \text{ as } n \to \infty.$$

Proof: Under assumption of the model, the PAL rule can be written as

$$PAL(m) = 2n \ln\left(\hat{\sigma}_m^2\right) + (3m+1)\ln(3\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)} + \gamma$$

where γ is a constant independent of *m*.

Case 1: Underestimation ($m \leq m_o$)

$$PAL(m) - PAL(m_o) = 2n \ln\left(\frac{\hat{\sigma}_m^2}{\hat{\sigma}_{m_o}^2}\right) + (3m+1)\ln(3\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)} - (3m_o+1)\ln(3\tilde{m}+1)\frac{\ln(r_{m_o}+1)}{\ln(\rho_{m_o}+1)}$$

Using Lemma 1, $\ln\left(\dfrac{\hat{\sigma}_m^2}{\hat{\sigma}_{m_o}^2}\right) \to 2\ln\left(1 + \sum_{j=m+1}^{m_o} \alpha_j^H \alpha_j \middle/ \sigma^2\right)$ a.s as $n \to \infty$, where r.h.s. is strictly positive and bounded.   Further, using Lemma 3 and 4 we get,

$$PAL(m) - PAL(m_o) = O(n) + (3m+1)\ln(3\tilde{m}+1)\frac{O(\ln n)}{O(\ln n)} - (3m_o+1)\ln(3\tilde{m}+1)\frac{O(\ln n)}{O(\ln n)} \text{ a.s.}$$

and hence $\frac{1}{n}\left(PAL(m)-PAL(m_o)\right)\to 2\ln\left(1+\sum_{j=m+1}^{m_o}\alpha_j^H\alpha_j\Big/\sigma^2\right)$ a.s. as $n\to\infty$ where r.h.s. is a

strictly positive and bounded quantity. Thus, $PAL(m)>PAL(m_o)$ $\forall m<m_o$ with probability 1. Further, note that

$$P\left(\hat{m}<m_o\right)=P\left(PAL(m)<PAL(m_o)\text{ for some }m<m_o\right)$$

$$=P\left(\bigcup_{m<m_o}PAL(m)<PAL(m_o)\right)$$

$$\leq\sum_{m=1}^{m_o-1}P\left(\frac{1}{n}\left(PAL(m)-PAL(m_o)\right)<0\right)\to 0\text{ as }n\to\infty.$$

Case 2: Overestimation ($m>m_o$)

We observe that

$$PAL(m)-PAL(m_o)=2n\ln\left(\frac{\hat{\sigma}_m^2}{\sigma_{m_o}^2}\right)+(3m+1)\ln(3\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)}-(3m_o+1)\ln(3\tilde{m}+1)\frac{\ln(r_{m_o}+1)}{\ln(\rho_{m_o}+1)}$$

We know from Remark 1 that $\forall m<m_o$ and $\forall$ n,

$$\frac{\ln(r_m+1)}{\ln(\rho_m+1)}\geq\frac{\ln(r_{m_o}+1)}{\ln(\rho_{m_o}+1)}$$

as the ratio is an increasing function of m. Therefore, we can write

$$\frac{1}{\ln n}\left(PAL(m)-PAL(m_o)\right)\geq\frac{1}{\ln n}\left(2n\ln\left(\frac{\hat{\sigma}_m^2}{\hat{\sigma}_{m_o}^2}\right)+3(m-m_o)\ln(3\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)}\right)\ \forall n.$$

Thus,

$$P\left(\frac{1}{\ln n}(PAL(m)-PAL(m_o))>0\right)=P\left(\frac{1}{\ln n}\left(2n\ln\left(\hat{\sigma}_m^2/\sigma_{m_o}^2\right)\right)+\frac{1}{\ln n}3(m-m_o)\ln(3\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)}>0\right)$$

It follows from the asymptotic theory of likelihood ratios that, $2n\ln\left(\hat{\sigma}_m^2/\sigma_{m_o}^2\right)\sim\chi_{3(m-m_o)}^2$ and hence

$$\frac{1}{\ln(n)}2n\ln\left(\hat{\sigma}_m^2/\sigma_{m_o}^2\right)=o_p(1).$$

Also, $\frac{1}{\ln n}3(m-m_o)\ln(3m+1)>0$ and the ratio $\frac{1}{\ln n}\frac{\ln(r_m+1)}{\ln(\rho_m+1)}=o(1)$. Hence

$$P\left(\frac{1}{\ln n}2n\ln\left(\hat{\sigma}_m^2/\sigma_{m_o}^2\right)+\frac{1}{\ln n}3(m-m_o)\ln(3\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)}>0\right)\to 1\text{ as }n\to\infty.$$

Thus, we have

$$P\left(\hat{m}>m_o\right)=P\left(PAL(m)<PAL(m_o)\text{ for some }m>m_o\right)$$

$$=P\left(\bigcup_{m>m_o}PAL(m)<PAL(m_o)\right)$$

$$\leq\sum_{m=m_o+1}^{\tilde{m}}P\left(\frac{1}{\ln n}\left(PAL(m)-PAL(m_o)\right)<0\right)\to 0\text{ as }n\to\infty.$$

This proves the result that $P(\hat{m}\neq m_o)\to 0$ as $n\to\infty$.

Numerical Examples

We consider the following 1-D cisoid model for simulation to compare performance of PAL method based estimator with other popular model order selection rules:

$$y_t = \sum_{k=1}^{m} \alpha_k e^{it\omega_k} + \varepsilon_t, \quad \alpha_1 = 3 + i2, \quad \alpha_2 = 2 + i1.66, \quad \alpha_3 = 1.75 + i, \omega_1 = 0.8\pi, \quad \omega_2 = 1.2\pi, \quad \omega_3 = 1.4\pi.$$

ε_t are i.i.d complex valued circularly symmetric gaussian error with zero mean and variance $\sigma^2 / 2$. We have considered the maximum model order to be 10, and sample size is varied from 5 to 200. We estimate the model order using different model order selection rules including PAL, BIC, BIC corrected (BICc) and AIC and report the probabilities of correct selection based on 200 simulation runs. Some representative plots from the simulation results are given in Figure 1 –Figure 4.
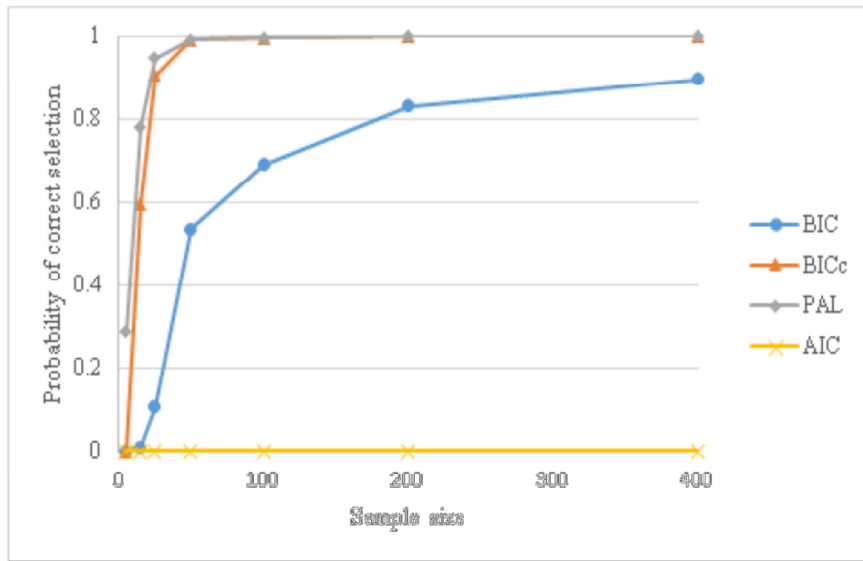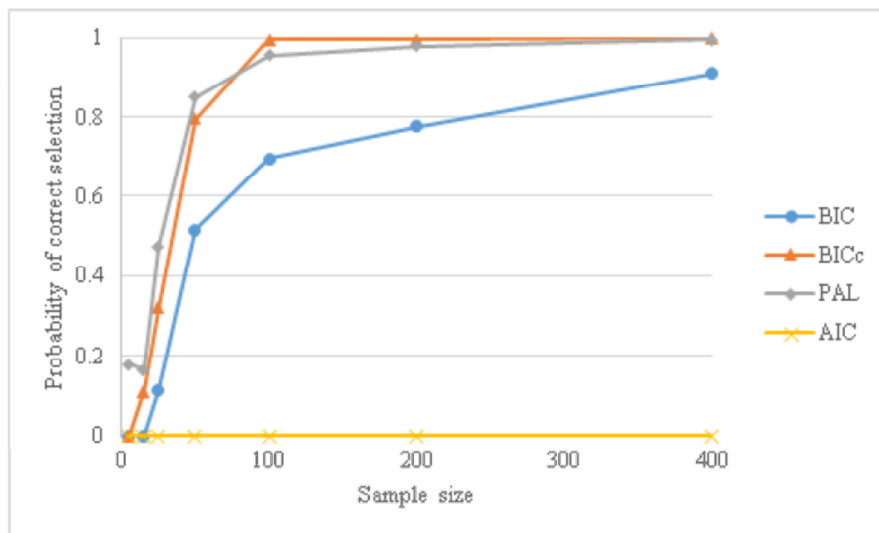

Figure 1. Results for $\sigma^2 = 3$.

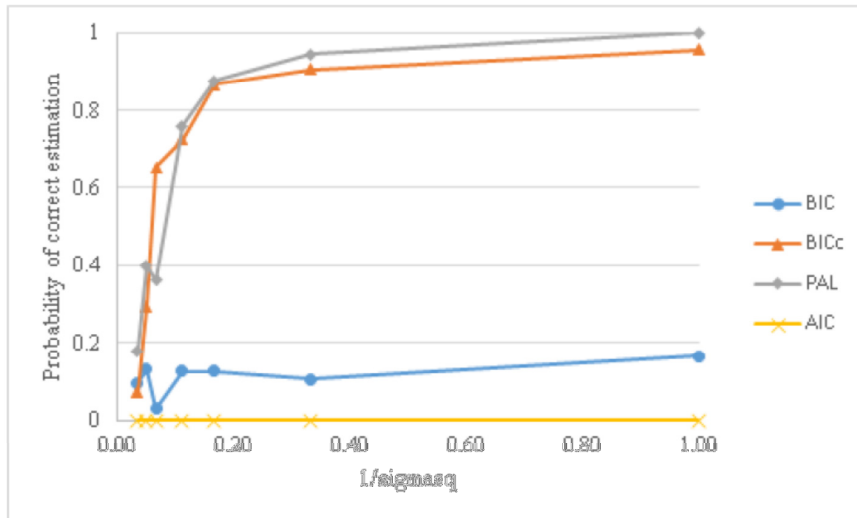Figure 2. Results for $\sigma^2 = 20$.
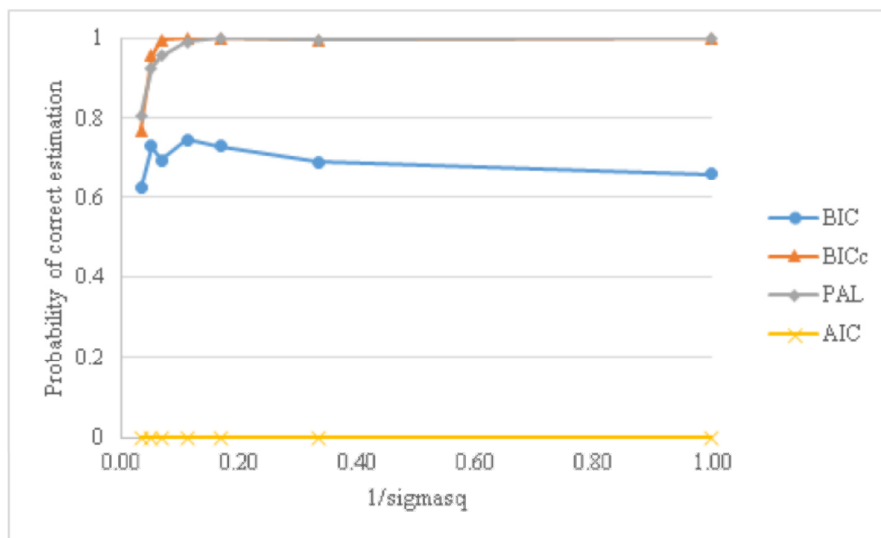
Figure 3. Results for *n=25*.

Figure 4. Results for *n=100*.

References

[1] P. Stoica and P. Babu, Model order estimation via penalizing adaptively the likelihood (PAL), *Signal Processing*, vol. 93(11), (2013), 2865-2871.

[2] P. Stoica and Y. Selen, Model-order selection: a review of information criterion rules, *IEEE Signal Processing Magazine*, vol.21(4), (2004), 36-47.

[3] K. Surana, S. Mitra, A. Mitra and P. Stoica, Estimating the order of sinusoidal models using the adaptively penalized likelihood approach: Large sample consistency properties, *Signal Processing*, vol.128, (2016), 204-211.

[4] M. Kliger and J. M. Francos, Strong consistency of a family of model order selection rules for estimating 2D sinusoids in noise, *Statistics & Probability Letters*, vol.78(17), (2008), 3075-3081.

[5] A. Prasad, D. Kundu, and A. Mitra, Sequential estimation of the sum of sinusoidal model parameters, *Journal of Statistical Planning and Inference*, vol.138(5), (2008), 1297-1313.