ELSEVIER

Signal Processing 56 (1997) 103–110

SIGNAL PROCESSING

Estimating the number of sinusoids in additive white noise[1]

Debasis Kundu*

Department of Mathematics, I.I.T. Kanpur, Pin 208016, India

Received 4 December 1995; revised 10 April 1996 and 29 August 1996

Abstract

In this article we propose a simple method for estimating the number of sinusoids in additive noise by using the penalty function approach. We establish the strong consistency of the proposed method under the assumption of independent and identically distributed random errors. We also provide an upper bound of the probability of wrong detection under the same assumptions. © 1997 Elsevier Science B.V.

Zusammenfassung

Wir schlagen hier ein Verfahren zur Schätzung der Anzahl von Sinusanteilen in additivem Rauschen vor und verwenden dazu den Straffunktions-Ansatz. Wir weisen die strenge Konsistenz der vorgeschlagenen Methode nach unter der Annahme unabhängiger, gleichartig verteilter Zufallsfehler. Wir liefern unter derselben Annahme auch eine Obergrenze für die Wahrscheinlichkeit von Fehldetektionen. © 1997 Elsevier Science B.V.

Résumé

Nous proposons dans cet article une méthode simple, basée sur l'approche par fonction de pénalité, pour l'estimation du nombre de sinusoïdes dans du bruit additif. Nous établissons la forte consistance de la méthode proposée sous l'hypothèse d'erreurs indépendantes et identiquement distribuées. Nous fournissons également une borne supérieure pour la probabilité de fausse détection sous la même hypothèse. © 1997 Elsevier Science B.V.

Keywords: Sinusoidal model; Strong consistency; Information theoretic criteria; Extended order modeling

1. Introduction

The problem of estimating the parameters of sinusoids has a long history dating from the 16th century. In this paper we consider the followiρng model:

$$Y_t = \sum_{j=1}^{p} \rho_j \cos(\lambda_j t + \phi_j) + \varepsilon_t \tag{1.1}$$

under the following conditions:

(1) $\{\phi_j\}$ and $\{\rho_j\}$ are unknown real numbers and, without loss of generality, we can assume that $\rho_1 \geqslant \rho_2 \geqslant \cdots > \rho_p$.

(2) λ_j's are unknown real numbers lying between 0 to 2π.

*Corresponding address: tel.: 250 158; fax: 0512 250 260; e-mail: kundu@iitk.ernet.in.

[1] Part of this work has been supported by the Department of Science and Technology (Grant No. SR/OY/M-06/93) Government of India.

PII S0165-1684(96)00153-3

(3) Let $\{\varepsilon_t\}$ be i.i.d. sequence of random variables with mean zero and finite variance $\sigma^2 > 0$.
(4) We assume that the maximum number of terms can be at most J, which is known in advance.

We are interested in estimating p, the number of sinusoidal components, present in the model (1.1).

This is an old problem and different methods are available in the literature to estimate the frequencies of the sinusoids very efficiently (see [8]), but not much attention has been paid to estimating the number of sinusoids when it is unknown. Some work on estimating the number of sinusoidal components of the model (1.1) can be found in [2–4, 6, 7, 9]. Some of the most elegant methods of estimating the number of signals and their performance analysis of the direction of arrival model can be found in [10, 11].

In this article, we propose a simple method to estimate the number of components, i.e. p, of the sinusoidal model under the assumptions of independent and identically distributed (i.i.d.) additive error random variables, following the approach of [10]. We also give an upper bound of the probability of wrong detection for large n as of [10]. The rest of the paper is organized as follows. In Section 2 we develop the criterion. Section 3 gives the consistency results. An upper bound on the probability of wrong detection is obtained in Section 4, some numerical experiments are performed in Section 5, finally, we draw the conclusions in Section 6.

2. Estimation procedure

Consider the following data matrix:

$$\underset{n-L+1\times L}{A_L} = \begin{pmatrix} Y_1 & \cdots & Y_L \\ \vdots & \ddots & \vdots \\ Y_{n-L-1} & \cdots & y_n \end{pmatrix} \tag{2.1}$$

for some fixed $L > 2J$. Compute the $L\times L$ matrix $R_L = (1/n)A_L^{\mathrm{T}}A_L$. Let the eigenvalues of R_L be $\hat\sigma_1^2 > \hat\sigma_2^2 \cdots > \hat\sigma_L^2$. Observe that although $\hat\sigma_i^2$, $i = 1, \ldots, L$, depends on L and n, we do not make it explicit for brevity. Now consider

$$\phi(i, C_n) = \hat\sigma_{2i+1}^2 + iC_n, \quad i = 1, \ldots, J. \tag{2.2}$$

Here $C_n > 0$, satisfies the following two conditions:

$$\text{(a)} \lim_{n\to\infty} C_n = 0, \quad \text{(b)} \lim_{n\to\infty}\left(\frac{\log\log n}{n}\right)^{1/2}\Big/ C_n = 0. \tag{2.3}$$

Choose that value of i, $1 \leqslant i \leqslant J$, for which $\phi(i, C_n)$ is minimum and that will be an estimator, $\hat p_n$ of p.

3. Consistency of the estimation procedure

In this section we prove the consistency of the estimation procedure proposed in Section 2 following the approach of [10, 11]. Observe that the model (1.1) can be written in the following form:

$$\begin{aligned} Y_t &= \sum_{j=1}^{p} \tilde\rho_j \exp(\mathrm{i}\lambda_j t) + \exp(-\mathrm{i}\lambda_j t)) + \varepsilon_t \\ &\quad (\text{where } \tilde\rho_j = (1/2)\rho_j \exp(\mathrm{i}\phi_j)) \\ &= \sum_{j=1}^{2p} \alpha_j \exp(\mathrm{i}\omega_j t) + \varepsilon_t, \end{aligned} \tag{3.1}$$

where $\alpha_j = \alpha_{p+j} = \tilde\rho_j$, $\omega_j = \lambda_j$ and $\omega_{j+p} = -\lambda_j$ for $j = 1, \ldots, p$. Let us denote the (a, b)th element of $\boldsymbol{R}_L$ by r_{ab}; therefore,

$$\begin{aligned} r_{ab} &= \frac{1}{n}\sum_{s=0}^{n-L}\left(\sum_{j=1}^{2p} \bar\alpha_j \exp(-\mathrm{i}(a+s)\omega_j) + \varepsilon_{a+s}\right) \\ &\qquad \times \left(\sum_{j=1}^{2p} \alpha_j \exp(\mathrm{i}(b+s)\omega_j) + \varepsilon_{b+s}\right) \\ &\quad (\text{here ` $-$ ' denotes the conjugate of a complex number}) \\ &= \frac{1}{n}\sum_{s=0}^{n-L}\left(\left(\sum_{j=1}^{2p} \bar\alpha_j \exp(-\mathrm{i}(a+s)\omega_j)\right)\right. \\ &\qquad \times \left(\sum_{j=1}^{2p} \alpha_j \exp(\mathrm{i}(b+s)\omega_j)\right) \\ &\qquad + \left(\sum_{j=1}^{2p} \bar\alpha_j \exp(-\mathrm{i}(a+s)\omega_j)\varepsilon_{b+s}\right) \\ &\qquad \left. + \left(\sum_{j=1}^{2p} \alpha_j \exp(\mathrm{i}(b+s)\omega_j)\varepsilon_{a+s}\right) + \varepsilon_{a+s}\varepsilon_{b+s}\right). \end{aligned} \tag{3.2}$$

By the law of iterated logarithm for the M-dependent sequences, we can say that

$$r^a_{ab} \overset{\text{a.s.}}{=} (\Omega^{\mathrm{H}} \boldsymbol{D} \Omega + \sigma^2 I_L)_{ab} + \mathrm{O}(\log\log n/n)^{1/2}, \quad (3.3)$$

where $\boldsymbol{D} = \mathrm{diag}\{|\alpha_1|^2, \ldots, |\alpha_{2p}|^2\}$ is a $2p \times 2p$ matrix and

$$\underset{2p \times L}{\Omega} = \begin{pmatrix} \exp(\mathrm{i}\lambda_1) & \cdots & \exp(\mathrm{i}L\lambda_1) \\ \vdots & \ddots & \vdots \\ \exp(\mathrm{i}\lambda_p) & \cdots & \exp(\mathrm{i}L\lambda_p) \\ \exp(-\mathrm{i}\lambda_1) & \cdots & \exp(-\mathrm{i}L\lambda_1) \\ \vdots & \ddots & \vdots \\ \exp(\mathrm{i}\lambda_p) & \cdots & \exp(-\mathrm{i}L\lambda_p) \end{pmatrix}, \quad (3.4)$$

where 'H' denotes the conjugate transpose of a matrix or a vector, $(\Omega^{\mathrm{H}} \boldsymbol{D} \Omega + \sigma^2 I_L)_{ab}$ denotes the (a, b)th element of $(\Omega^{\mathrm{H}} \boldsymbol{D} \Omega + \sigma^2 I_L)$ and I_L is the identity matrix of order $L \times L$. It can be seen that $\phi(i, C_n)$ defined in (2.2) is a 0-regular function as defined in [10]. Therefore, following an approach similar to that of [10], it can be shown that the estimation procedure as suggested in Section 2, is a consistent estimation procedure.

4. Upper bound on the probability of wrong detection

In this section we give an upper bound on the probability of wrong detection under the assumptions (1)–(4). Since $\phi(i, C_n)$ is a 0-regular function, following the approach similar to that of [10], we obtain for a sufficiently large n,

$$\begin{aligned} \Pr\{\hat{p}_n \neq p\} &\leqslant J \Pr\left\{\sum_{i=1}^{2J} |\hat{\sigma}_i^2 - \sigma_i^2|/2J > C_n\right\} \\ &\leqslant J \Pr\left\{\sum_{i=1}^{2p} (\hat{\sigma}_i^2 - \sigma_i^2)^2/2J > C_n^2\right\} \\ &\quad \text{(follows from [1])} \\ &\leqslant J \Pr\left\{\sum_{i=1}^{L} (\hat{\sigma}_i^2 - \sigma_i^2)^2/2J > C_n^2\right\} \\ &\leqslant J \Pr\left\{\sum_{i=1}^{L} \sum_{j=1}^{L} |r_{ij} - \sigma_{ij}|^2/2J > C_n^2\right\} \\ &\quad \text{(by [10, Lemma 4.2])}, \end{aligned} \quad (4.1)$$

where σ_{ij} is the (i, j)th element of $\Sigma = (\Omega^{\mathrm{H}} \boldsymbol{D} \Omega + \sigma^2 I_L)$ and r_{ij} is the (i, j)th element of $\boldsymbol{R}_L$. Now we want to compute

$$\Pr\left\{\sum_{i=1}^{L} |r_{ij} - \sigma_{ij}|^2 > 2J\, C_n^2\right\}.$$

Let us denote the asymptotic dispersion matrix of $\sqrt{n}\,\mathrm{vec}(\boldsymbol{R}_L - \Sigma)$ (here vec of an $L \times L$ matrix is an $L^2 \times 1$ vector obtained by writing the L columns one below another) by Γ. Therefore, we can say that

$$\sqrt{n}\,\mathrm{vec}(R_L - \Sigma) \xrightarrow{\mathrm{d}} \mathrm{N}_{L^2}(0, \Gamma),$$

here '$\xrightarrow{\mathrm{d}}$' means convergence in distribution and $\mathrm{N}_{L^2}(0, \Gamma)$ denotes L^2 variate multivariate normal distribution with mean $\mathbf{0}$ and dispersion matrix Γ. So

$$n\,\mathrm{vec}(\boldsymbol{R}_L - \Sigma)^{\mathrm{T}}\,\mathrm{vec}(\boldsymbol{R}_L - \Sigma) \xrightarrow{\mathrm{d}} x^{\mathrm{T}} \Gamma x,$$

here x is an L^2 variate multivariate normal distribution, with mean $\mathbf{0}$ and dispersion matrix I. Therefore,

$$x^{\mathrm{T}} \Gamma x \overset{\mathrm{d}}{=} [\tau_1 V_1 + \tau_2 V_2 + \cdots + \tau_{L^2} V_{L^2}],$$

where $\tau_1, \ldots, \tau_{L^2}$ are the eigenvalues of Γ and $V_1, \ldots, V_{L^2}$ are i.i.d. random variables each with chi-square one degree of freedom. Observe that the expression for Γ for any particular model can be easily obtained using the central limit theorem.

5. Numerical experiments and discussions

We perform a Monte Carlo simulation study to compare different penalty functions. All these computations are performed on PC-486, using the random deviate generator suggested in [5].

We consider the following model:

$$Y_t = 1.0\cos(1.40\pi t) + 1.0\cos(1.50\pi t + 0.1) + \varepsilon_t.$$

Here ε_t's are i.i.d. random variables with mean zero and variance σ^2. We take three different error distributions, namely (a) normal, (b) uniform and (c) double exponential. We take $\sigma^2 = 1.0$ and

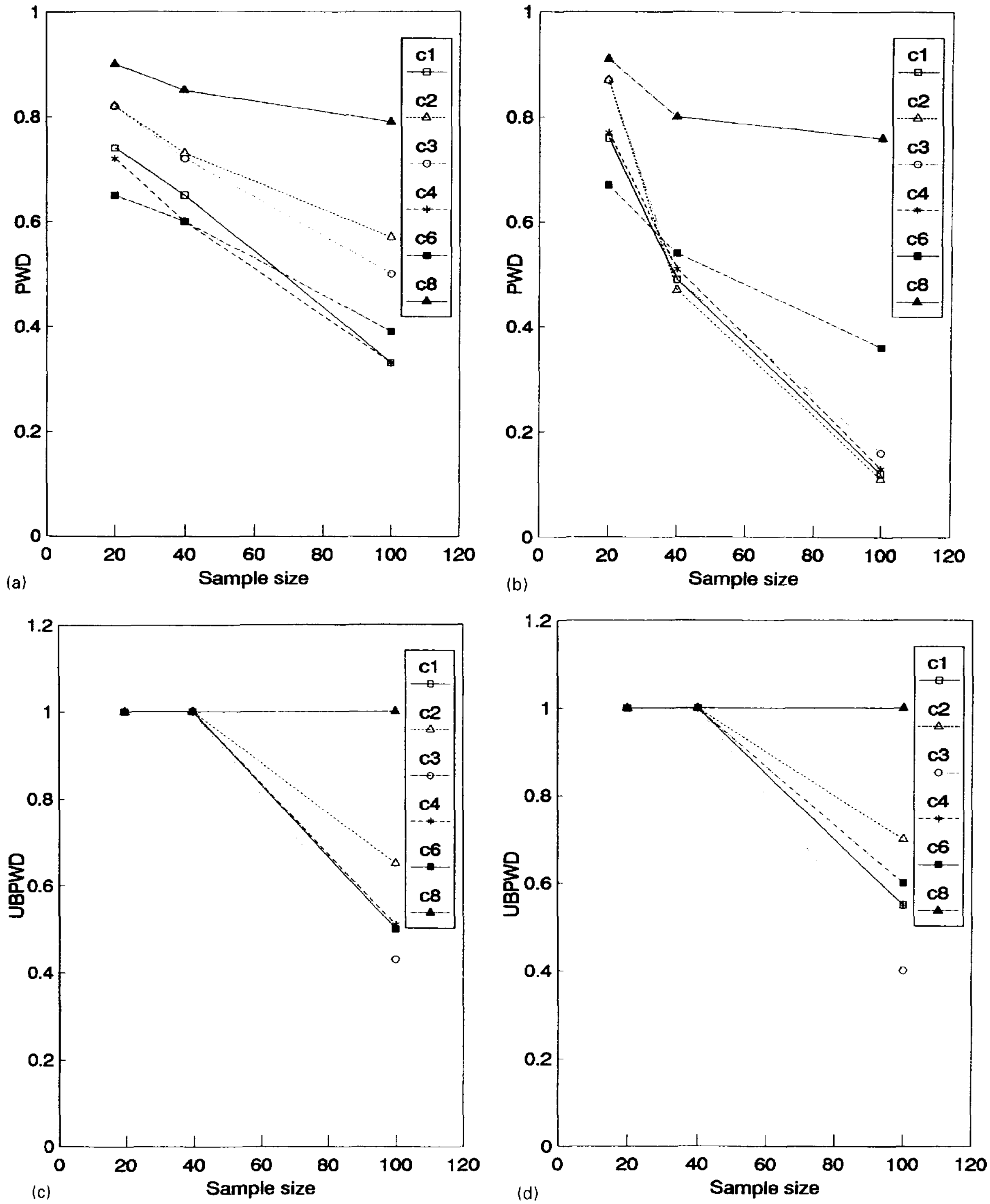

Fig. 1. Probability of wrong detection for different penalty function in case of Gaussian error, when (a) $L = 11$ and (b) $L = 15$. Upper bound of the probability of wrong detection for different penalty function in case of Gaussian error, when (c) $L = 11$ and (d) $L = 15$.

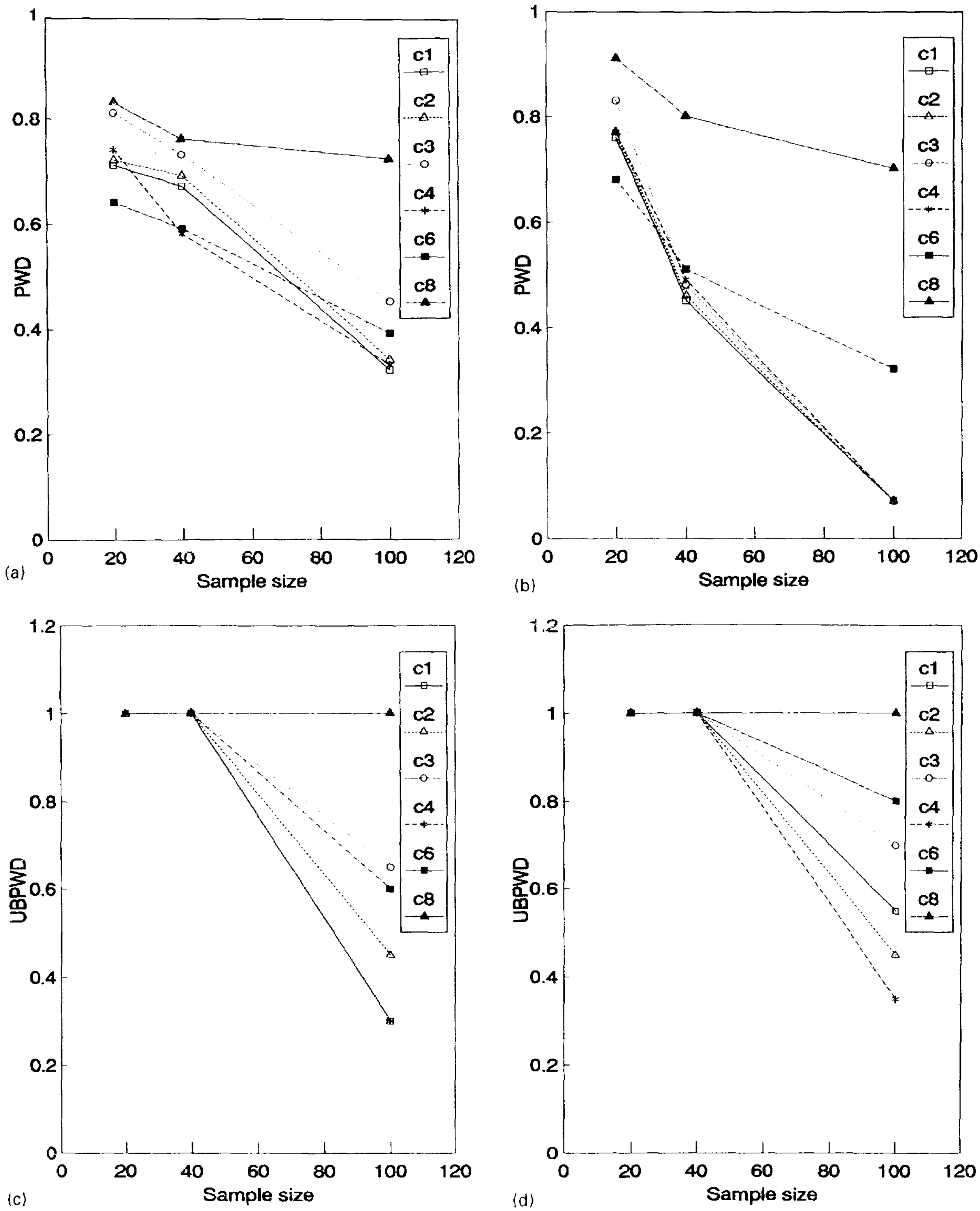

Fig. 2. Probability of wrong detection for different penalty function in case of uniform error, when (a) $L = 11$ and (b) $L = 15$. Upper bound of the probability of wrong detection for different penalty function in case of uniform error, when (c) $L = 11$ and (d) $L = 15$.

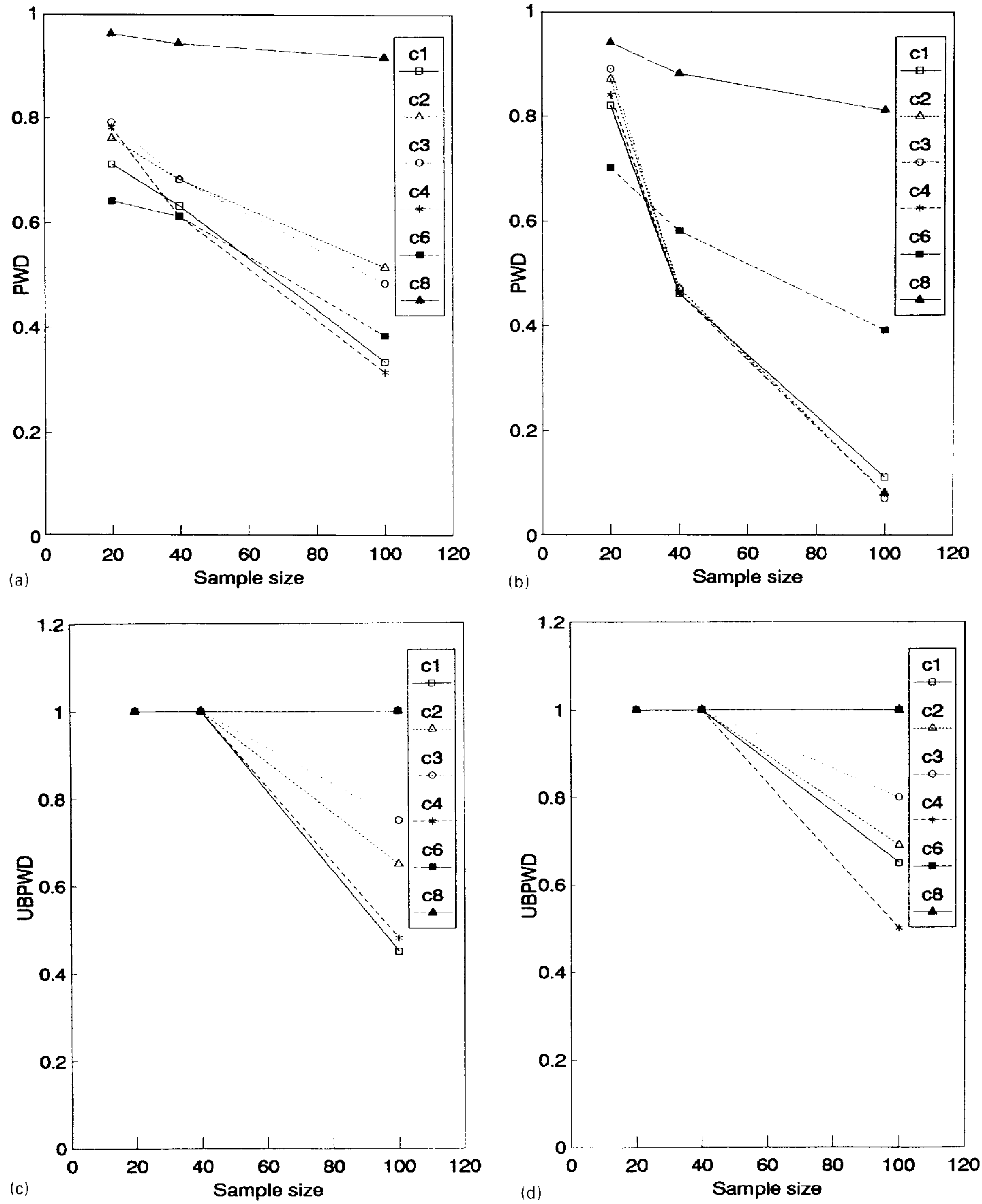

Fig. 3. Probability of wrong detection for different penalty function in case of double exponential error, when (a) $L = 11$ and (b) $L = 15$. Upper bound of the probability of wrong detection for different penalty function in case of exponential error, when (c) $L = 11$ and (d) $L = 15$.

consider $n = 20$ (small sample), $n = 40$ (moderate sample), $n = 100$ (large sample) and $L = 11$ and 15. One hundred replications of the data set for different n for all the distributions are generated. We take a wide range of C_n, satisfying (2.3) and converging to zero at different rates. We denote them as $C_1, \ldots, C_9$, where $C_1 = (1/n)^{0.1}$, $C_2 = (1/n)^{0.05}$, $C_3 = (1/\log n)^{0.4}$, $C_4 = (1/\log n)^{0.5}$, $C_5 = (1/\log n)^{0.6}$, $C_6 = (1/n \log n)^{0.2}$, $C_7 = (1/n \log n)^{0.3}$, $C_8 = (1/n \log n)^{0.4}$, $C_9 = (1/n \log n)^{0.5}$. It is assumed that the maximum number of terms is 5 or equivalently $J = 5$. For each data set and for each penalty function, we estimate p, the number of terms present in the model, by the method described in Section 2 and repeat it over one hundred replications. We also obtain the upper bound of the probability of wrong detection as obtained in Section 4. We report the probability of wrong detection (PWD) and the upper bound of the probability of wrong detection (UBPWD) in Figs. 1–3. Figs. 1(a) and (b) represent the PWD and Figs. 1(c) and (d) represent the UBPWD for $L = 11$ and $L = 15$, respectively, when the error distribution is Gaussian. Similarly, Figs. 2 and 3 represent the results for uniform and double exponential errors. We report the results for $C_1, C_2, C_3, C_4, C_6, C_8$, because C_5, C_7 behave very similarly with C_6 and C_9 behaves very similarly with C_8.

Comparing the results, it is observed that for fixed σ, the probability of correct detection increase as n increases. It indicates that all the methods give consistent estimates of p. It is observed that for all the three distributions considered, the PWD is quite close to each other although the UBPWD differ significantly from each other. The UBPWD for the normal error is relatively better than the other two distributions although even for the normal error the bound is not very tight and sometimes UBPWD is lower than the PWD. It seems that further work is needed to obtain a better bound. Regarding the effect of L on the detection performance of the different methods, it is observed that as L increases from $2J + 1$, the PWD decreases and then starts increasing. It is observed that for $L \cong (n/3)$, the PWD reaches its minimum, although the UBPWD increases as L increases from $2J + 1$. Therefore, achieving the best performances of both PWD and UBPWD is not possible at the same time.

Finally, it comes to the choice of proper C_n, which is one of the important problems in practice. Calculating C_n at $n = 20$, 40 and 100 allows the various choices to be ranked in order of increasing severity as follows:

$$C_9, C_8, C_7, C_6, C_5, C_1, C_4, C_3, C_2.$$

Out of these $\{C_9, C_8\}$, $\{C_7, C_6, C_5\}$ and $\{C_1, C_4\}$ behave very similarly. The simulations show that the least severe penalties (especially C_9) are too weak, causing p to be generally overestimated. The most severe penalties (namely C_2), on the other hand, are too severe and p is then generally underestimated. An optimal choice will lie somewhere between these extremes. The simulations show that $\{C_1, C_4\}$ behave satisfactorily at least for moderate and large sample sizes and we recommend $C_4 = (1/\log n)^{1/2}$ for its simplicity, although no theoretical justification can be given.

6. Conclusions

In this paper we provide a simple new method of estimation of the number of terms in a sum of sinusoidal model and also obtain the upper bound of the probability of wrong detection following the approach of [10, 11]. We do not need any distributional assumption on the error terms like [7]. Moreover, computationally, it involves only the calculation of eigenvalues of an $L \times L$ matrix, where as in most of the existing methods it involves the minimization of a certain function which involves some search procedure. It is observed that if we use the penalty $(1/\log n)^{0.5}$, then it works well for different error distributions. Observe that the criterion (2.2) is not scale invariant, but it can be easily made scale invariant by dividing it by some estimate of σ^2, for example σ^2_{2J+1}. We have observed that it does not make much difference in the detection purposes. We have also given an upper bound on the estimation of wrong detection, which may be useful in choosing the proper C_n theoretically. It seems that more work is needed in this direction.

Acknowledgements

Part of the work was done when I was visiting the ECE Department of The Indian Institute of Science, Bangalore. I would like to thank Professor V.U. Reddy for his suggestions and help. I would also like to thank the two referees for their comments to improve the earlier version of this paper.

References

[1] K.L. Chung, *A Course in Probability Theory*, Academic Press, New York, 2nd Edition, 1974.

[2] J.J. Fuchs, "Estimating the number of sinusoids in additive white noise", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-36, 1988, pp. 1846–1853.

[3] E.J. Hannan, "Determining the number of jumps in a spectrum", in: T. Subba Rao, ed., *Developments in Time Series Analysis*, Chapman and Hall, London, 1993, pp. 127–138.

[4] L. Kavalieries and E.J. Hannan, "Determining the number of terms in trigonometric regression", *J. Time Ser. Anal.*, Vol. 15, 1994, pp. 613–625.

[5] W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, *Numerical Recipes*, Cambridge University Press, Cambridge, 1986.

[6] B.G. Quinn, "Estimating the number of terms in a sinusoidal regression", *J. Time Ser. Anal.*, Vol. 10, No. 1, 1989, pp. 71–75.

[7] V.U. Reddy and L.S. Biradar, "SVD-based information theoretic criteria for detection of the number of damped/undamped sinusoids and their performance analysis", *IEEE Trans. Signal Process.*, Vol. ASSP-41, No. 9, 1993, pp. 2872–2881.

[8] P. Stoica, "List of references on spectral line analysis", *Signal Processing*, Vol. 31, 1993, pp. 329–340.

[9] X. Wang, "An AIC type estimator for the number of sinusoids", *J. Time Ser. Anal.*, Vol. 14, 1993, pp. 431–440.

[10] Y.Q. Yin and P.R. Krishnaiah, "On some nonparametric methods for detection of the number of signals", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-35, No. 11, 1987, pp. 1533–1538.

[11] L.C. Zhao, P.R. Krishnaiah and Z.D. Bai, "On detection of the number of signals in the presence of white noise", *J. of Multivariate Anal.*, Vol. 20, 1986, pp. 26–49.