ELSEVIER Computational Statistics & Data Analysis 36 (2001) 245–256

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Estimating the number of signals of the damped exponential models

Debasis Kundu [*,1], Amit Mitra [2]

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

Received 1 November 1998; accepted 1 May 1999

Abstract

In this paper we consider the problem of estimating the number of signals of the damped exponential models. We use different information theoretic criteria to detect the number of signals and compare their small sample performances by Monte Carlo simulations study. © 2001 Elsevier Science B.V. All rights reserved.

MSC: 62J99

Keywords: Damped Exponential Signals; Information Theoretic Criteria; Monte Carlo Simulations

1. Introduction

Detecting the number of signals and estimating the parameters of the damped exponential signals are important problems in signal processing. We formulate the problems as follows:

Let $y_1, y_2, \ldots, y_n$ be a sample of size n, where y_t is given by

$$y_t = \sum_{k=1}^{M} \alpha_k \exp(-s_k t + i2\pi f_k t) + \varepsilon_t. \tag{1.1}$$

Here α_k's are unknown complex numbers called amplitude of the kth signal, f_k's are distinct real numbers lying between 0 and 1, s_k's are the damping factors and

* Corresponding author.

E-mail address: kundu@iitk.ac.in (D. Kundu).

[1] Part of this work was done when the author was visiting the Department of Statistics, The Pennsylvania State University, State College, PA 16802-2111, USA.

[2] Presently at the Reserve Bank of India, Mumbai, India.

PII: S 0167-9473(00)00036-0

are positive real numbers, $i = \sqrt{-1}$. $\{\varepsilon_t\}$ is a sequence of independent identically distributed random variables with mean zero and finite variance for both the real and the imaginary part. The real and imaginary parts of $\{\varepsilon_t\}$ are assumed to be independent and normally distributed. M, the number of signals is also assumed to be unknown. Given the sample of size n, the problem is to estimate the unknown parameters α_k, s_k, f_k for $k = 1, \dots M$ and M also.

The estimation of the parameters of a damped exponential model (1.1) is an old problem (Kay, 1987) and the readers are referred to Stoica (1993) for an extensive list of references. A lot of methods for estimating the frequencies have been proposed by researchers over the last 20 years. Among the notables, are the methods of Errikson et al. (1994), Kay (1984), Kundu and Mitra (1995), Stoica and Nehorai (1989), Stoica et al. (1989), Tufts and Kumaresan (1982) and Yan and Bressler (1993). All these methods of estimation assume that the number of signals, M is known. The aim of this paper is to estimate the number of signals M, which is usually unknown, under the assumption that the number of signals can be at the most K, which is known in advance.

Wax and Kailath (1985) developed information theoretic criteria for detecting the number of signals received by a sensor array. Fuchs (1988) developed a criterion, based on the perturbation analysis of the data auto correlation matrix, for detecting the number of sinusoids. More recently, Reddy and Biradar (1993), following the information theoretic approach to model selection, developed a criterion for detecting the number of damped/undamped exponentials. The detection performance of these criteria were compared with that of Fuchs (1988) and their results showed that the Minimum Description Length (MDL) criterion as developed by them performs nearly same as that of Fuchs (1988). A more general information theoretic criterion in model selection has been proposed by Zhao et al. (1986a,b) called the Efficient Detection Criterion (EDC). Rao (1988) suggested the use of EDC to estimate the number of signals for damped or undamped case but he did not perform any numerical experiments. It is known (Bai et al., 1987) that the EDC give consistent estimates for estimating the number of signals in undamped exponential signals, although the same result is not applicable for damped exponential model. Kundu (1992) gave a detailed comparison of the different information theoretic criteria for estimating the number of undamped signals, but nowhere, at least not known to the authors, the comparison of the different information theoretic criteria exist for damped exponential model.

Note that for the damped exponential model the data sequence is pure noise as the sample size goes to infinity. Therefore, one can not obtain any asymptotically consistent estimate of the number of signals. However, when the damping factor is not that first, it is hoped that some good detection criterion can surely be obtained by suitable algorithms, which should be able to estimate the number of signals reasonably well. That is the main aim of this paper.

For the undamped exponential models all the information theoretic criterion can be written in form (2.8), where C_n represents penalty function. It has to satisfy the conditions given in (2.7). Note that, the penalty function C_n goes to infinity for the undamped model to give consistent estimate of the number of signals. For the damped model if C_n goes to infinity, then for large sample size any criterion will underestimate the number of signals. In fact, the penalty function should go to zero as n tends to infinity. We modify C_n for the damped model and propose the modified information theoretic criteria where the penalty function depends on the amplitude as well as the damping factor. If there is no damping factor it coincides with the information theoretic criteria for the undamped model. We obtain the probability of the wrong detection. The probability of wrong detection depends on the unknown parameters. We propose to use bootstrap techniques to estimate the probability of wrong detection for a particular penalty function. Once we estimate the probability of wrong detection, we choose that penalty function for which the wrong detection is minimum. Some simulations are performed to see the effectiveness of the proposed criterion.

The organization of the rest of the paper is as follows. In Section 2 we introduce different information theoretic criteria and propose the modified efficient detection criteria for the damped model. The practical implementation procedures are provided in Section 3. In Section 4 we present the numerical experiments and finally we draw conclusions in Section 5.

2. Different information theoretic criteria

In this section we discuss the different information theoretic criteria for estimating the number of signals of the damped exponential signal models. We introduce the Akaike Information Criteria (AIC), Minimum Description Length (MDL) criteria and Efficient Detection Criteria (EDC).

Let $y_1, \ldots, y_n$ be a sample of size n from the model (1.1). Let P, be the parameter that ranges over all possible number of signals, i.e., $P \in \{1, \ldots, K\}$. Then the joint density function of the data set can be written as

$$f(y|\theta_P) = \frac{1}{\pi^n \sigma^{2n}} \exp\left(-\frac{1}{2} \sum_{t=1}^{n} |y_t - \mu_t(\theta_P)|^2 \right), \tag{2.1}$$

where

$$\theta_P = (\alpha_1, \ldots, \alpha_P, s_1, \ldots, s_P, f_1, \ldots, f_p)$$

and

$$\mu_t(\theta_P) = \sum_{k=1}^{P} \alpha_k \exp(-s_k t + i2\pi f_k t).$$

We now formulate the problem as follows: Given a set of n observations and a family of models $\{f(y|\theta_P);\ P = 1, \ldots, K\}$ which is a parameterized family of probability densities $f(y|\theta_P)$, our problem is to select the true one.

Posed in this way, this problem is perfectly suited for using different information theoretic criteria such as AIC, MDL or the EDC. The AIC, MDL and EDC criteria are known as penalized likelihood method in the general statistical literature. Here a penalty function is subtracted from the log-likelihood before it is maximized.

This serves to penalize or discourage the addition of more and more parameters. In this set up the best model would be one for which the penalized likelihood is maximum. For the general problem on this topic, one can refer to Akaike (1973, 1974, 1978), Hannan and Quinn (1979), Rissanen (1978), Schwartz (1983) and Zhao et al. (1986a,b).

Akaike (1973, 1974) proposed the Akaike Information Criterion (AIC). The AIC suggests choosing $\hat{M}$, an estimator of M, which minimizes the following expression.

$$\mathrm{AIC}(P) = -\log f(y|\hat{\theta}_P) + \mathrm{d}(\theta_P); \tag{2.2}$$

for $P = 1,\ldots,K$, where $\hat{\theta}_P$ is the maximum likelihood estimator (MLE) of θ_P and $\mathrm{d}(\theta_P)$ is the number of independent parameters of the parameter vector θ_P.

Akaike's basic idea was to choose the model that minimizes the mean of the Kullback–Leibler distance between the true density $f(y|\theta_P)$ and the estimated density $f(y|\hat{\theta}_P)$. Since the distance is unknown, he proposed to estimate it by the log-likelihood of the MLE. The second term in (2.2) was added to make the log-likelihood at the MLE an unbiased estimator of the Kullback–Leibler distance.

In the exponential signals model, with the assumption of the Gaussian error the AIC takes the following form:

$$\mathrm{AIC}(P) = -n \log R_P - 8P \tag{2.3}$$

(see Rao, 1988), where R_P, denotes the minimum value of

$$\sum_{t=1}^{n} |y_t - \mu_t(\theta_P)|^2 \tag{2.4}$$

and the minimization is performed with respect to $\alpha_1,\ldots,\alpha_P$, $s_1,\ldots,$ $s_P, f_1,\ldots,f_P$.

Minimum Description Length (MDL) criterion was introduced by Rissanen (1978). The basic idea is that the best model is the one that provides the shortest description of the data. It has been shown (Rissanen, 1983) that for large samples this criterion leads to the selection of the model that minimizes

$$\mathrm{MDL}(P) = -\log f(y|\hat{\theta}_P) + \tfrac{1}{2}\mathrm{d}(\theta_P)\log n \tag{2.5}$$

for $P = 1,\ldots,K$, where $f(y|\hat{\theta}_P)$ and $\mathrm{d}(\theta_P)$ are as defined before.

Schwartz (1978) suggested a model selection approach based on Bayesian arguments. Assuming a priori probabilities for every competing model, he proposed selecting the model that maximizes the posterior probability. It has been shown that for a model belonging to an exponential family, the maximization of the posterior probability leads to the minimization of the criterion given by (2.5) asymptotically.

The Efficient Detection Criterion (EDC) method of Zhao et al. (1986a,b) consists of choosing as an estimator of M, the number $\hat{M}$, which minimizes

$$\mathrm{EDC}(P) = -\log f(y|\hat{\theta}_P) + C_n \mathrm{d}(\theta_P) \tag{2.6}$$

for $P = 1,\ldots,K$, where C_n's are such that

$$\lim_{n\to\infty} \frac{C_n}{n} = 0, \quad \text{and} \quad \lim_{n\to\infty} \frac{C_n}{\log\log n} = \infty. \tag{2.7}$$

In the exponential signal model, with the assumption of Gaussian error, the EDC takes the following form (see Rao, 1988):

$$\mathrm{EDC}(P) = -n \log R_P - C_n(8P). \tag{2.8}$$

Observe that MDL criterion is a special case of the EDC. For MDL, C_n takes the value $\frac{1}{2}$ log n in (2.8). The estimators of M, obtained from (2.8) are strongly consistent for the undamped exponential model. For a detailed proof of the consistency for the undamped model see Bai et al. (1987). For the damped model however the consistency results do not hold, therefore, it is important to observe the behavior of the different information theoretic criteria in this situation at least for small samples.

Now, we try to analyze what kind of problem we might encounter if we directly use (2.8) for estimating the number of signals for damped exponential model. Note that, (2.7) implies C_n tends to infinity as n tends to infinity. Suppose, M is the correct order model, then C_n should be such that

$$\mathrm{EDC}(M) < \mathrm{EDC}(P); \quad \text{for } P = 1,\ldots,K, P \neq M. \tag{2.9}$$

Now (2.9) implies

$$n \log R_M + C_n(8M) < n \log R_P + C_n(8P); \quad \text{for } P = 1,\ldots,K, P \neq M. \tag{2.10}$$

Since $R_1 > R_2 > \cdots > R_K$ almost surely, (2.10) implies that C_n must satisfy

$$n \log \left(\frac{R_M}{R_{M+1}}\right) < 8C_n < n \log \left(\frac{R_{M-1}}{R_M}\right). \tag{2.11}$$

For undamped model

$$\lim_{n\to\infty} \frac{R_M}{R_{M+1}} = 1 \quad \text{and} \quad \lim_{n\to\infty} \frac{R_{M-1}}{R_M} > 1 \tag{2.12}$$

for damped model

$$\lim_{n\to\infty} \frac{R_M}{R_{M+1}} = 1 \quad \text{and} \quad \lim_{n\to\infty} \frac{R_{M-1}}{R_M} = 1. \tag{2.13}$$

Because of the damped factor, note that for large n,

$$\frac{R_{M-1}}{R_M} = 1 + \mathrm{O}(\mathrm{e}^{-\delta n}), \tag{2.14}$$

where $\delta = \max\{s_1,\ldots,s_M\} > 0$. If we divide by $\log\log n$ in (2.11) and take the limit, we obtain

$$\infty = 8 \lim_{n\to\infty} \frac{C_n}{\log\log n} < \lim_{n\to\infty} n \log(1 + \mathrm{O}(\mathrm{e}^{-\delta n})) = 0. \tag{2.15}$$

Therefore, if C_n tends to infinity, for large n, (2.11) may not satisfy. On the other hand, it looks more reasonable that the penalty function should be more if the amplitudes are more (suggested by a referee). Based on the above observations, we propose the following modified EDC (MEDC) for the damped model

$$\mathrm{MEDC}(P) = -n \log R_P - AC_n \mathrm{e}^{-\delta n}(8P) \tag{2.16}$$

here $A = \max\{A_1,\ldots,A_M\}$ and $\delta = \max\{\delta_1,\ldots,\delta_M\}$. If the damping factor is zero, then MEDC coincides with the usual EDC. Note that we need to know A and δ to implement MEDC in practice. We will describe that in the next section.

3. Practical implementation

Consider the following data matrix:

$$\mathbf{A} = \begin{bmatrix} y_1 & \cdots & y_L \\ \vdots & \ddots & \vdots \\ y_{n-L+1} & \cdots & y_n \end{bmatrix}.$$

Here L is any integer such that $K < L < N - K$. Let us denote the matrix $\mathbf{T} = (1/n)\mathbf{A}^*\mathbf{A}$, where '$*$' denotes the conjugate transpose of a matrix or of a vector. We obtain the spectral decomposition of the matrix $\mathbf{T}$ as follows:

$$\mathbf{T} = \sum_{i=1}^{L+1} \hat{\sigma}_i^2 \hat{\mathbf{U}}_i \hat{\mathbf{U}}_i^*$$

here $\sigma_1^2 > \cdots > \sigma_{L+1}^2$ are the ordered eigenvalues of $\mathbf{T}$ and $\hat{\mathbf{U}}_i$'s are the normalized eigenvalues corresponding to $\hat{\sigma}_i^2$.

Assuming that the true order of the model is K (the maximum one), we estimate, first, the K damping factors and the K amplitudes say $\delta_1 > \cdots > \delta_K$ and $A_1, \dots, A_K$, respectively, by using the NSD method of Kundu and Mitra (1995) from $\mathbf{T}$. We use $\delta = \delta_1$ and $A = \max\{A_1, \dots, A_M\}$. Note that the values of δ and A depend on L, we provide some suggestions to choose L in the next section.

For a given choice of C_n and from the estimated A and δ, we can compute MEDC(P) for different values of $P = 1, \dots, K$ and choose $\hat{M}$ an estimate of M such that MEDC($\hat{M}$) is minimum.

Note that we have a wide choice of C_n, but we would like to choose that C_n so that $P(\hat{M} \neq M)$ is minimum. First let us compute $P(\hat{M} \neq M)$.

$$\begin{aligned} P(\hat{M} \neq M) &= P(\hat{M} < M) + P(\hat{M} > M) \\ &= \sum_{q=0}^{M-1} P(\hat{M} = q) + \sum_{q=M+1}^{K} P(\hat{M} = q) \\ &= \sum_{q=0}^{M-1} P(MEDC(q) - MEDC(M) < 0) \\ &\quad + \sum_{q=M+1}^{K} P(MEDC(q) - MEDC(M) < 0) \\ &= \sum_{q=0}^{M-1} P(n \log R_q - n \log R_M > AC_n \mathrm{e}^{-\delta n} 8(M - q)) \\ &\quad + \sum_{q=M+1}^{K} P(n \log R_M - n \log R_q < AC_n \mathrm{e}^{-\delta n} 8(q - M)). \end{aligned} \qquad (3.1)$$

Unfortunately, $P(\hat{M} \neq M)$ depends on the unknown model parameters. Without knowing the original parameters we can not calculate the theoretical probabilities. We would like to estimate these probabilities with the help of the given sample and using the bootstrap technique. The idea is as follows. From any particular realization of the model, we compute the matrix $\mathbf{T}$ and obtain the corresponding eigenvalues and eigenvectors. We estimate the error variance σ^2 by averaging the last $L-K$ eigenvalues of $\mathbf{T}$, say $\hat{\sigma}^2$. Now suppose, using the penalty function C_n, we estimate the order of the model as $M(C_n)$. We generate n complex Gaussian random variables with mean zero and variance $\hat{\sigma}^2$, say $\varepsilon_1, \ldots, \varepsilon_n$. We obtain the new bootstrap sample as

$$y_t^B = y_t + \varepsilon_t \quad \text{for } t = 1, \ldots, n.$$

Assuming $M(C_n)$ is the correct order model, we check for $q < M(C_n)$, whether

$$n\log(R_q) - n\log(R_{M(C_n)}) > AC_n \mathrm{e}^{-\delta n} 8(M(C_n) - q),$$

or, for $q > M(C_n)$, check whether

$$n\log(R_{M(C_n)}) - n\log(R_q) < AC_n \mathrm{e}^{-\delta n} 8(q - M(C_n)).$$

Repeating the process, say B times, we can estimate (3.1). Finally, we choose that C_n for which the estimated $P(\hat{M} \neq M)$ is minimum.

Some justifications regarding this kind of bootstrap estimates of (3.1) can be given. Note that the realization of y_t^B can be thought of coming from a model (1.1) with $V(\varepsilon_t) \approx 2\sigma^2$. Note that for the damped exponential model R_q/R_M for $q = 1, \ldots, K$ is independent of σ^2. Therefore, (3.1) remains invariant if we change the error variance from σ^2 to $2\sigma^2$.

4. Numerical experiments

In this section we present the Monte Carlo simulations done for small samples to compare the different information theoretic criteria. All these computations have been done on $HP-9000$, machine at the Indian Institute of Technology, Kanpur.

We consider four different models with different parameters and different standard deviations of the error random variables. The four models are given as follows:

$$\text{Model 1}: \; y_t = \mathrm{e}^{\pi/4}\mathrm{e}^{(-0.01t + i2\pi(0.52)t)} + \mathrm{e}^{\pi/2}\mathrm{e}^{(-0.02t + i2\pi(0.42)t)} + \varepsilon_t,$$

$$\text{Model 2}: \; y_t = \mathrm{e}^{\pi/4}\mathrm{e}^{(-0.01t + i2\pi(0.52)t)} + \mathrm{e}^{\pi/2}\mathrm{e}^{(-0.02t + i2\pi(0.50)t)} + \varepsilon_t,$$

$$\text{Model 3}: \; y_t = \mathrm{e}^{\pi/4}\mathrm{e}^{(-0.01t + i2\pi(0.52)t)} + \mathrm{e}^{\pi/2}\mathrm{e}^{(-0.03t + i2\pi(0.42)t)} + \varepsilon_t,$$

$$\text{Model 4}: \; y_t = 1 + \mathrm{e}^{\pi/4}\mathrm{e}^{(-0.01t + i2\pi(0.52)t)} + \mathrm{e}^{\pi/2}\mathrm{e}^{(-0.02t + i2\pi(0.50)t)} + \varepsilon_t.$$

The data are generated using the different standard deviation, viz $\sigma = 0.01$, 0.1, 0.5 and 1.0 and with different sample sizes $n = 25$, 50, 75 and 100. The random deviates are generated with the help of the IMSL random deviate generator. For each of the four models one hundred replications of the data set for different n and σ are generated. Observe that in Models 1 and 2, the amplitudes and the damping factors

Table 1

SS	ITC	$\sigma = 0.01$			$\sigma = 0.1$			$\sigma = 0.5$			$\sigma = 1.0$		
		PUE	PCE	POE	PUE	PCE	POE	PUE	PCE	POE	PUE	PCE	POE
	MEDC	0.0	0.99	0.01	0.0	0.99	0.01	0.03	0.97	0.0	0.24	0.52	0.19
25	AIC	0.0	0.42	0.58	0.0	0.40	0.60	0.00	0.40	0.60	0.00	0.38	0.62
	MDL	0.0	0.86	0.14	0.0	0.86	0.14	0.00	0.83	0.17	0.27	0.52	0.21
	MEDC	0.0	1.0	0.00	0.0	1.0	0.00	0.00	0.99	0.01	0.10	0.75	0.15
50	AIC	0.0	0.33	0.67	0.0	0.35	0.65	0.00	0.33	0.67	0.00	0.36	0.64
	MDL	0.0	0.82	0.18	0.0	0.83	0.17	0.00	0.82	0.18	0.10	0.72	0.18
	MEDC	0.0	0.98	0.02	0.0	0.97	0.03	0.04	0.96	0.0	0.27	0.49	0.24
75	AIC	0.0	0.15	0.85	0.0	0.15	0.85	0.00	0.13	0.87	0.00	0.14	0.86
	MDL	0.0	0.77	0.23	0.0	0.77	0.23	0.00	0.75	0.25	0.04	0.76	0.20
	MEDC	0.0	1.0	0.00	0.0	1.0	0.00	0.00	0.98	0.02	0.03	0.74	0.23
100	AIC	0.0	0.16	0.84	0.0	0.16	0.84	0.00	0.12	0.88	0.00	0.14	0.86
	MDL	0.0	0.75	0.25	0.0	0.77	0.23	0.00	0.72	0.28	0.03	0.70	0.27

are kept fixed, whereas the difference of the radian frequencies is more in Model 1 than in Model 2. Between Models 1 and 3, the amplitudes and the radian frequencies are kept fixed, whereas the difference between the damping factor is more in Model 3 than in Model 1. Model 4 is a higher-order model than Models 1,2 or 3. As far as the estimation of frequencies are concerned, it is known (Kundu and Mitra, 1995) that it is difficult to estimate the parameters in Model 2 than in Model 1 and similarly in Model 1 than in Model 3. No such comparison can be made between Models 2 and 3. Between Models 4 and 2 it is expected that Model 2 will be easier than Model 4 as the number of parameters are more in Model 4 than that of Model 2. It is expected that in estimating the number of signals also, the same pattern will exist.

We compare the usual AIC and usual MDL with the proposed MEDC. Note that for AIC and MDL, $C_n = 1$ and $C_n = 1/2\log n$, respectively, in (2.8). For MEDC, we take a varied choice of C_n satisfying (2.7) (except when $C_n = 1$) but diverging to infinity at different rates from very slow to very fast. The different choices of C_n considered are as follows: $C_n = 1$, $C_n = n^{0.1}$, $C_n = n^{0.5}$, $C_n = n^{0.9}$, $C_n = \log n$, $C_n = (\log n)^{0.1}$, $C_n = (\log n)^{0.5}$, $C_n = (\log n)^{0.9}$, $C_n = (n\log n)^{0.1}$, $C_n = (n\log n)^{0.5}$, $C_n = (n\log n)^{0.9}$ and $C_n = \frac{1}{2}\log n$. It is assumed that for all the four models the maximum number of signals is 6. First, we assume that the model order is $K = 6$. Now, using the modified noise space decomposition method with $L \approx \min\{\frac{3}{5}N, 20\}$ we obtain the estimates of A and δ. Using that A and δ, we compute MEDC(P) from (2.16) for different values of $P = 1,\ldots,6$ for a particular choice of C_n. We obtain an estimate of σ^2 by averaging the last $(L-K)$ eigenvalues of $\mathbf{A}^{\mathbf{H}}\mathbf{A}$ and also obtain an estimate of $P(\hat{M} \neq M)$ as the method suggested in the previous section. We take $B = 100$, in our calculations. The results are reported in Tables 1–4. We report the percentage of under estimate (PUE), percentage of correct estimate (PCE) and the percentage of over estimate (POE) for AIC, MDL and MEDC over five hundred replications.

Table 2

SS	ITC	$\sigma = 0.01$			$\sigma = 0.1$			$\sigma = 0.5$			$\sigma = 1.0$		
		PUE	PCE	POE	PUE	PCE	POE	PUE	PCE	POE	PUE	PCE	POE
	MEDC	0.04	0.96	0.00	0.04	0.90	0.06	0.24	0.76	0.00	0.61	0.39	0.00
25	AIC	0.00	0.41	0.59	0.0	0.42	0.58	0.66	0.17	0.17	0.60	0.20	0.20
	MDL	0.00	0.84	0.16	0.01	0.85	0.14	0.76	0.24	0.00	0.73	0.27	0.00
	MEDC	0.00	1.0	0.00	0.0	1.0	0.00	0.15	0.85	0.01	0.51	0.49	0.00
50	AIC	0.00	0.28	0.72	0.0	0.28	0.72	0.40	0.13	0.47	0.50	0.19	0.31
	MDL	0.00	0.84	0.16	0.0	0.85	0.15	0.73	0.27	0.00	0.72	0.28	0.18
	MEDC	0.00	1.0	0.00	0.02	0.98	0.00	0.17	0.83	0.00	0.49	0.51	0.00
75	AIC	0.00	0.13	0.87	0.00	0.13	0.87	0.21	0.07	0.72	0.33	0.18	0.49
	MDL	0.00	0.78	0.22	0.0	0.79	0.21	0.75	0.25	0.00	0.77	0.23	0.00
	MEDC	0.00	1.0	0.00	0.02	0.98	0.00	0.17	0.83	0.00	0.49	0.51	0.00
100	AIC	0.00	0.15	0.85	0.0	0.15	0.85	0.87	0.07	0.06	0.85	0.13	0.02
	MDL	0.00	0.79	0.21	0.0	0.78	0.22	0.80	0.20	0.00	0.78	0.22	0.00

Table 3

SS	ITC	$\sigma = 0.01$			$\sigma = 0.1$			$\sigma = 0.5$			$\sigma = 1.0$		
		PUE	PCE	POE	PUE	PCE	POE	PUE	PCE	POE	PUE	PCE	POE
	MEDC	0.00	0.98	0.02	0.00	0.98	0.02	0.04	0.96	0.00	0.27	0.49	0.24
25	AIC	0.00	0.40	0.60	0.00	0.40	0.60	0.00	0.38	0.62	0.00	0.41	0.59
	MDL	0.00	0.86	0.14	0.00	0.86	0.14	0.00	0.85	0.15	0.22	0.48	0.30
	MEDC	0.00	1.0	0.00	0.00	1.0	0.00	0.00	0.99	0.01	0.09	0.70	0.21
50	AIC	0.00	0.35	0.65	0.00	0.33	0.67	0.00	0.33	0.67	0.00	0.38	0.62
	MDL	0.00	0.84	0.16	0.00	0.83	0.17	0.00	0.84	0.16	0.12	0.75	0.13
	MEDC	0.00	1.0	0.00	0.00	1.0	0.00	0.00	0.96	0.04	0.04	0.74	0.22
75	AIC	0.00	0.15	0.85	0.00	0.15	0.85	0.00	0.15	0.85	0.00	0.18	0.82
	MDL	0.00	0.79	0.21	0.0	0.79	0.21	0.00	0.77	0.23	0.04	0.75	0.21
	MEDC	0.00	1.0	0.00	0.00	1.0	0.00	0.00	0.98	0.02	0.03	0.76	0.21
100	AIC	0.00	0.16	0.84	0.0	0.14	0.86	0.00	0.16	0.84	0.00	0.16	0.84
	MDL	0.00	0.77	0.23	0.0	0.77	0.23	0.00	0.70	0.30	0.04	0.69	0.27

5. Conclusions

In this paper, we consider the problem of estimating the number of damped exponential signals. We use different information theoretic criteria for estimating the number of signals.

We consider the AIC, MEDC and the MDL criteria for the detection problem. It is well known that the AIC criteria does not provide the consistent estimates in general model selection problem. This fact is well reflected in the results of the simulations

Table 4

SS	ITC	$\sigma = 0.01$			$\sigma = 0.1$			$\sigma = 0.5$			$\sigma = 1.0$		
		PUE	PCE	POE	PUE	PCE	POE	PUE	PCE	POE	PUE	PCE	POE
	MEDC	0.00	1.0	0.00	0.10	0.78	0.12	0.44	0.56	0.00	0.71	0.29	0.00
25	AIC	0.00	0.38	0.62	0.09	0.31	0.60	0.77	0.11	0.12	0.84	0.10	0.06
	MDL	0.00	0.67	0.39	0.49	0.29	0.22	0.85	0.15	0.00	0.82	0.18	0.00
	MEDC	0.00	1.0	0.00	0.00	1.0	0.00	0.19	0.81	0.00	0.60	0.40	0.00
50	AIC	0.00	0.33	0.67	0.04	0.29	0.67	0.73	0.10	0.17	0.88	0.08	0.04
	MDL	0.00	0.78	0.22	0.49	0.51	0.00	0.58	0.18	0.24	0.81	0.17	0.02
	MEDC	0.00	1.0	0.00	0.02	0.98	0.00	0.19	0.81	0.00	0.63	0.37	0.00
75	AIC	0.00	0.13	0.87	0.00	0.14	0.86	0.37	0.11	0.52	0.45	0.14	0.41
	MDL	0.00	0.65	0.35	0.00	0.68	0.32	0.78	0.22	0.00	0.79	0.21	0.00
	MEDC	0.00	1.0	0.00	0.05	0.95	0.00	0.19	0.81	0.00	0.65	0.35	0.00
100	AIC	0.00	0.10	0.90	0.00	0.09	0.91	0.81	0.07	0.12	0.93	0.07	0.00
	MDL	0.00	0.61	0.39	0.00	0.64	0.36	0.83	0.17	0.00	0.83	0.17	0.00

given in Tables 1–4. Comparing the Tables 1–4 it is observed that, although the MEDC and MDL criteria give consistent estimates for undamped signals model, the same cannot be said for the damped models. It is well known (Wu, 1981; Kundu, 1994) that although it is possible to estimate consistently the parameters of the undamped exponential model, it is not possible to obtain the consistent estimates of the parameters of the damped model. It may not be surprising if we look at the damped model carefully. From the model it is clear (if the damping factor is negative) that as the sample size n increases the signal component vanishes to zero and we are left with the error components only. Therefore, even if we increase the sample sizes, we may not extract any more information about the signal parameters from the sample. In fact, the inconsistency is clearly indicated in the simulation results. It is observed that the number of correct selections by different methods do not increase for a fixed σ as n increases. In fact for AIC and MDL in many cases they even decrease. For fixed n, as σ decreases, it is observed that for MDL and MEDC, the performances improve. This indicates the consistency of the MDL and MEDC methods as σ decreases to zero for fixed n. It is also observed that for a fixed n as σ increases the methods have a tendency to over estimate for models 1 and 3, whereas, they have a tendency to under estimate for models 2 and 4. Comparing the tables, it is observed that in most of the cases, for fixed n and σ, the number of correct detection is more in Table 1 than in Table 2, which is not very surprising as the difference of the radian frequencies ($|f_1 - f_2|$) is more in model 1 than in model 2. This fact was also seen for the undamped signals by Kundu (1992). Interestingly, the number of correct detection in Table 3 is more or less same as that of Table 1; although, the difference in the damping factor ($|s_1 - s_2|$) is more in model 3 than in model 1. The difference in performance is more marked if the two models differ significantly with respect to the frequencies. The performance of most of the methods is much better for model 1 than that of model 2 if σ is > 0.01. Between Models

2 and 4, the behavior is quite similar in nature for almost all the cases considered, although the number of correct detection is more in Model 2 compared to Model 4.

Now, comparing the three methods it is quite clear that AIC does not work well for this particular model. Our simulations show that for AIC the probability of correct detection never exceeds 0.45 also the inconsistency of the AIC is very prominent. MDL criterion works reasonably well if the error variance is not very high. If the error variance is high and the difference of the radian frequencies is small (Model 2 and Model 4) then the performance of MDL is also very poor. MEDC works very well if the error variance is low. It can detect almost 90–100% for all the models considered if $\sigma \leq 0.1$. If the error variance is high and the radian frequencies are close to each other the performance drops significantly. It may not be very surprising, since if the radian frequencies are close to each other and the error variance is high it is very difficult to estimate the unknown parameters by any method. Although, eventually as n tends to infinity MEDC also will give inconsistent estimates but at least for finite sample it works reasonably well and better than the existing known methods. Therefore, even though MEDC are quite involved computationally compared to AIC or MDL, it can be used to estimate the number of components for the damped exponential model.

Acknowledgements

The authors would like to thank two referees for their valuable suggestions and to Professor Dr. Peter Naeve for his encouragements.

References

Akaike, H., 1973. Information theory and an extension of the ML principle. Proceedings of the second International Symposium on Information Theory Supp. to Problems of Information Theory and Control. Tokyo, pp. 267–281.

Akaike, H., 1974. A new look at statistical model identification. IEEE Trans. Automat. Control 19, 716–723.

Akaike, H., 1978. A Bayesian analysis of the minimum AIC procedure. Ann. Inst. Stat. Math. 30, 9–14.

Bai et al, 1987. Asymptotic properties of the EVLP estimation for superimposed exponential signals in noise. Technical Report No. 87-19 CMA, University of Pittsburgh.

Errikson, A., Stoica, P., Soderstorm, T., 1994. Markov-based eigen analysis method for frequency estimation. IEEE Trans. Signals Process. 42 (3), 586–594.

Fuchs, J.J., 1988. Estimating the number of sinusoids in additive white noise. IEEE Trans. Acoust. Speech Signal Proc. ASSP-36 36, 1846–1853.

Hannan, E.J., Quinn, B.G., 1979. The determination of the order of an autoregression. J. Roy. Statist. Soc. B Vol. 41, 190–195.

Kay, S.M., 1984. Accurate frequency estimation at low Signal to Noise ratio. IEEE Trans. Acoust. Speech Signal Process. ASSP-32, 540–547.

Kay, S.M., 1987. Modern Spectral Estimation; Theory and Applications. Prentice-Hall, Englewood Cliffs, NJ.

Kundu, D., 1992. Estimating the number of signals using the information theoretic criterion. J. Statist. Comput. Simulat. 44, 117–131.

Kundu, D., 1994. A modified Prony algorithm for damped or undamped exponential signals. Sankhya Ser A 56 (3), 524–544.

Kundu, D., Mitra, A., 1995. Estimating the parameters of exponentially damped or undamped sinusoids in noise: A non iterative approach. Signal Processing 46 (3), 363–368.

Rao, C.R., 1988. Some results in signal detection. In: Gupta, S.S., Berger, J.O. (Eds.), Statistical Decision Theory and Related Topics IV, Vol. 2. Springer, Berlin, pp. 319–332.

Reddy, V.U., Biradar, L.S., 1993. SVD-based information theoretic criteria for detection of the number of damped/undamped sinusoids and their performance analysis. IEEE Trans. Signal Process. ASSP-41, 2872–2881.

Rissanen, J., 1978. Modeling of shortest data description. Automatica 14, 465–471.

Rissanen, J., 1983. A universal prior for the integers and the estimation by MDL. Ann. Statist. 11, 417–431.

Schwartz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6, 461–464.

Stoica, P., 1993. List of references on spectral analysis. Signal Processing 31, 329–340.

Stoica, P., Nehorai, A., 1989. MUSIC, maximum likelihood and Cramer-Rao bound. IEEE Trans. Acoust. Speech Signal Process. 37, 720–741.

Stoica, P., Moses, R.L., Friedlander, B., Soderstorm, T., 1989. Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements. IEEE Trans. Acoust Speech Signal Process. 37, 378–392.

Tufts, D.W., Kumaresan, R., 1982. Estimating the parameters of exponentially damped and pole zero modeling in noise. IEEE Trans. Acoust. Speech Signal Process. 30, 833–840.

Wax, M., Kailath, T., 1985. Detection of signals by information theoretic criteria. IEEE Trans. Acoust Speech Signal Process. 30, 387–392.

Wu, C.F.J., 1981. Asymptotic theory of nonlinear least squares. Ann. Statist. 9, 501–513.

Yan, S.F., Bressler, Y., 1993. Maximum likelihood parameter estimation of superimposed signals by dynamic programming. IEEE Trans. Acoust Speech Signal Process. 41 (2), 804–820.

Zhao, L.C., Krishnaiah, P.R., Bai, Z.D., 1986a. On detection of the number of signals in the presence of white noise. J. Multivariate Anal. 20, 1–25.

Zhao, L.C., Krishnaiah, P.R., Bai, Z.D., 1986b. On detection of the number of signals when the noise covariance is arbitrary. J. Multivariate Anal. 20, 26–49.