ELSEVIER

Signal Processing 80 (2000) 525–534

SIGNAL PROCESSING

www.elsevier.nl/locate/sigpro

Detecting the number of signals for an undamped exponential model using cross-validation approach

Debasis Kundu*, Amit Mitra[1]

Department of Mathematics, Post-office, I.I.T. Kanpur, Pin 208016, India

Received 26 November 1998; received in revised form 8 October 1999

Abstract

Detecting the number of signals and estimating the parameters of the signals are important problems in statistical signal processing. Quite a number of papers appeared in the last 20 years in estimating the parameters of an exponential signal quite efficiently but not that much of attention has been paid in estimating the number of signals of an exponential signal model. Recently, it is observed that different information theoretic criteria can be used to estimate the number of signals in this situation. But it is also observed that the choice of the penalty function is very important particularly for small sample sizes. In this paper we suggest to use the cross-validation technique on estimating the number of signals and give its practical implementation procedures. Numerical experiments reveal that the new procedure performs quite comparable to the best performed information theoretic criteria at least for small sample sizes and it has certain desirable properties also.

Zusammenfassung

Die Detektion der Signalanzahl und Schätzung der Parameter der Signale sind wichtige Probleme in der statistischen Signalverarbeitung. Eine recht große Anzahl von Veröffentlichungen ist in den letzten 20 Jahre zur effizienten Parameterschätzung von Exponentialsignalen erschienen. Weniger Aufmerksamkeit wurde der Schätzung der Anzahl Signale im exponentialen Signalmodell gewidmet. Unlängst wurde festgestellt, daß underschiedliche informationstheoretische Kriteria zur Schätzung der Anzahl Signale in diesen Anwendungen benutzt werden können. Darüberhinaus wurde aber ebenso festgestellt, daß die Auswahl der Straffunktion von entscheidender Bedeutung ist, und zwar besonders dann, wenn nur wenige Abtastwerte vorliegen. In dieser Arbeit schlagen wir die Benutzung der Kreuzvalidierungstechnik *cross validation technique* zur Schätzung der Signalanzahl vor und geben hier die zugehörigen praktischen Implementierungsprozeduren an. Anhand numerischer Experimente zeigen wir, daß die neue Prozedur insbesondere bei nur wenigen Abtastwerten vergleichbare Resultate wie die besten informationstheoretischen Kriteria liefert und weitere attraktive Eigenschaften besitzt.

Résumé

La détection du nombre de signaux et l'estimation des paramètres de ces signaux constituent des problèmes importants en traitement statistique des signaux. Un nombre élevé d'articles ont été publiés durant les vingt dernières années sur l'estimation efficiente des paramètres d'un signal mais beaucoup moins d'attention a été portée à l'estimation du nombre de signaux dans un modèle de signaux exponentiels. Il a été observé récemment que différents critères issus de la théorie

* Corresponding author. Tel.: 0512-250-158; fax: 0512-250-260.

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

[1] Presently at the Reserve Bank of India.

0165-1684/00/$ - see front matter
PII: S0165-1684(99)00149-8

de l'information peuvent être utilisés pour estimer le nombre de signaux dans le contexte mentionné. Mais il a été également observé que le choix de la fonction de coût est crucial, tout particulièrement pour de petites tailles de données. Nous suggérons dans cet article l'utilisation de la technique de validation croisée pour estimer le nombre de signaux et décrivons son implantation pratique. Des expériences numériques révèlent que cette procédure nouvelle a des performances tout à fait comparables à celles des meilleurs critères issus de la théorie de l'information, tout du moins our de petites tailles de données, et qu'elle présente également un certains nombre de propriétés souhaitables.

Keywords: Exponential signals; Consistent estimates; Information theoretic criteria; Penalty function

1. Introduction

Estimation of number, amplitude and frequency of complex sinusoids in a signal is very important. The model can be expressed in the following way. Let $\{y_1, \ldots, y_n\}$ be a sample of size n, where y_k can be written as

$$y_k = \sum_{j=1}^{M} A_j e^{i\omega_j k} + e_k \quad \text{for } k = 1, \ldots, n. \tag{1}$$

The amplitudes, $A_j's$, are unknown complex numbers and the frequencies, $\omega_j's$, are unknown radian frequencies, between 0 to 2π. The additive errors, $e_k's$, are complex valued Gaussian random variables and they are independent and identically distributed (i.i.d.) with zero means. The real and imaginary parts of $e_k's$ are assumed to be independent. The number of signals, M, is unknown. The problem is to estimate M, $A_1, \ldots, A_M$, and $\omega_1, \ldots, \omega_M$.

In the last 20 years several iterative and non-iterative procedures were developed to estimate the parameters of an exponential model very efficiently, but not that much of attention has been paid on estimating the number of signals. See, for example [20] for an extensive list of references up to that point and see [11,12] for some of the recent references. In this paper we mainly concentrate on estimating M. Tufts and Kumaresan [24] also proposed some graphical techniques to estimate M, which is very subjective in nature. Some of the other techniques for example [4,14] can be used, but they also depend on the subjective choice of the individual and therefore the practical implementation becomes difficult.

Rao [15] proposed the information theoretic criteria (ITC) following the approach of Zhao et al., [26] on estimating the number of signals for an undamped exponential model (1). He did not provide any numerical results regarding the performances of his procedure. It is observed [8] that Rao's suggestion may not be implemented very easily in practice. A practical implementation procedure was suggested by [8]. It is also observed that different ITC depend very much on the penalty function used. Some suggestions about the penalty function were given in [8] based on the extensive computer simulations. It is not very difficult to show (as it is correctly mentioned by Rao [15]) that the ITC proposed by him will give consistent estimates of M, in the case of an undamped exponential model (1). Bai et al. [3] also proposed a method known as the equivariance linear prediction (EVLP) method on estimating M for an undamped exponential model and proved the strong consistency of the EVLP when the errors are not necessarily Gaussian. But extensive numerical simulations [8] suggest that the EVLP does not work well for small sample sizes even when the errors are Gaussian.

In this paper we estimate M, through cross-validation (CV) approach which usually performs well for small and moderate sample sizes. Rao [15] first mentioned that the CV technique can be used on estimating M for model (1). He provided certain CV scheme for practical implementation. However, he did not provide any numerical results regarding the performances of his proposed CV procedure. It is observed that Rao's CV approach for model (1) is not very easy to implement in practice as it was suggested. We propose a new simple CV procedure for an undamped exponential model using missing value technique discussed in Section 3. We perform detailed numerical experiments to compare

different ITC and the proposed CV procedure using different models. It is observed that the performance of the CV approach is quite similar with that of the best performed ITC and in certain cases it works marginally better also.

The rest of the paper is organized as follows. In Section 2, we give a brief description of the different ITC and the estimation of the different parameters in presence of missing value is discussed in Section 3. CV approach and its implementation are discussed in Section 4. The results of the numerical experiments are presented in Section 5 and finally we draw conclusions from our work in Section 6.

2. Different information theoretic criteria

Let $\{y_1, \ldots, y_n\}$ be a sample of size n from the model (1). Let L be the parameter ranges over all possible number of signals, i.e. $L \in \{1, \ldots, K\}$, where K is some preassigned fixed number. We make the assumption that the number of signals can be at most equal to K. Then it follows that the joint density function of the observed data is given by

$$f(\boldsymbol{y}|\theta_L) = \frac{1}{(\sqrt{2\pi})^n \sigma^n} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{k=1}^{n} |y_k - \mu_k(\theta_L)|^2 \right\}, \quad (2)$$

where

$$\theta_L = (A_1, \ldots, A_L, \omega_1, \ldots, \omega_L) \quad (3)$$

and

$$\mu_k(\theta_L) = \sum_{j=1}^{L} A_j e^{i\omega_j k}. \quad (4)$$

The problem can now be formulated as follows: Given the family of models

$$\{f(\boldsymbol{y}|\theta_L); L = 1, \ldots, K\} \quad (5)$$

select the true one. Posed in this way the problem becomes a model selection problem and perfectly suited for using the different ITC such as Akaike information criterion (AIC) of Akaike [1,2] or minimum description length (MDL) criterion (best information criterion (BIC)) of Schwartz [19] or Risannen [17] or the efficient detection criterion (EDC) of Zhao et al. [26]. The AIC, MDL or the EDC criteria are known as the penalized likelihood method in the statistical model selection literature. Here a penalty function is subtracted from the log-likelihood function before it is maximized. This serves to penalize or discourage the addition of more and more parameters.

The AIC suggests to choose $\hat{M}$, an estimator of M, which minimizes the following expression:

$$\mathrm{AIC(L)} = -\log f(\boldsymbol{y} \mid \hat{\theta}_L) + d(\hat{\theta}_L) \quad (6)$$

for $L = 1, \ldots, K$, where $\hat{\theta}_L$ is the maximum likelihood estimator (MLE) of θ_L and $d(\hat{\theta}_L) = 3L + 1$, the total number of independent parameters when the model order is L. Akaike's basic idea was to choose the model that minimizes the mean of the Kullback–Liebler distance between the true density $f(\boldsymbol{y}|\theta_L)$ and the estimated density $f(\boldsymbol{y}|\hat{\theta}_L)$. Since the distance is unknown, he proposed to estimate it by the log-likelihood value at the point of the MLE. The second term in (6) was added to make the log-likelihood function at the MLE an unbiased estimator of the Kullback–Leibler distance.

MDL criterion was introduced by Risannen [17]. The basic idea is that the best model is the one that provides the shortest description of the data. It is observed [18] that for large samples this estimator leads to the selection of the model that minimizes

$$\mathrm{MDL(L)} = -\log f(\boldsymbol{y}|\hat{\theta}_L) + \tfrac{1}{2} d(\hat{\theta}_L) \log n \quad (7)$$

for $L = 1, \ldots, K$, where $f(\boldsymbol{y}|\hat{\theta}_L)$ and $d(\hat{\theta}_L) = 3L + 1$ are same as defined before.

Schwartz [19] suggested a model selection criterion based on the Bayesian arguments. Assuming a priori probabilities for every competing models, he proposed selecting the model that maximizes the posterior probability. It was 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 (7) asymptotically.

The efficient detection criterion (EDC) of Zhao et al. [26] consists of choosing an estimator $\hat{M}$ of M, which minimizes

$$\mathrm{EDC(L)} = -\log f(\boldsymbol{y}|\hat{\theta}_L) + C_n d(\hat{\theta}_L) \quad (8)$$

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

$$\lim_{n\to\infty} \frac{C_n}{n} = 0, \quad \lim_{n\to\infty} \frac{C_n}{\log\log n} = \infty \tag{9}$$

and $d(\hat{\theta}_L) = 3L + 1$ (see [15]). Observe that MDL criterion is a special case of EDC. For MDL, C_n takes the value $\frac{1}{2}\log n$ in (8). It has been mentioned in [15] (the proof is not very difficult) that for an undamped exponential model the estimators obtained from EDC or MDL are strongly consistent estimators of M, where as AIC estimator is not consistent. Although any C_n satisfying (9) gives strongly consistent estimator of M, but unfortunately it is observed that the small sample performances of the estimators depend very much on the choice of C_n (see [8]). Kundu [8] used a wide variety of penalty functions, namely, $C_n = N^{0.1}$, $C_n = N^{0.2}$, $C_n = N^{0.3}$, $C_n = N^{0.4}$, $C_n = \log n$, $C_n = (\log n)^{0.2}$, $C_n = (\log n)^{0.4}$, $C_n = (\log n)^{0.6}$, $C_n = (\log n)^{0.8}$, $C_n = (n\log n)^{0.1}$, $C_n = (n\log n)^{0.3}$, $C_n = (n\log n)^{0.5}$, $C_n = (n\log n)^{0.7}$, $C_n = (n\log n)^{0.9}$. All of them satisfy (9) and they diverge to infinity at different rates. Extensive numerical simulations indicate that $C_n = \log n$ can be used as a good choice of C_n for an undamped exponential model, although no theoretical justification can be given in favor of this. Experimentally, it is observed that the penalty function $\frac{1}{2}\log n$ is relatively milder compared to $\log n$ and therefore it over estimates M. On the other hand, the penalty function $\log n$ looks appropriate at least for Gaussian error.

3. Estimation of the unknown parameters in presence of missing value for fixed M

In this section we discuss about the estimation of the unknown parameters if one observation is missing and when the model order is known a priori. Since all the existing eigen decomposition methods use the fact that the data are equispaced, therefore it is not immediate how they can be used if one observation is missing. It is well known that the modified forward–backward linear prediction (MFBLP) method of Tufts and Kumaresan [24] works very well with short data length and moderate signal-to-noise ratio. Its practical implementation is also quite simple. In this section we observe that the MFBLP can be further modified and can be used even if one observation is missing and when the model order is known.

Note that for known M, there exists a vector $\boldsymbol{g} = (g_1, \ldots, g_J)$ such that in the noiseless data, the forward–backward prediction equations are as follows [24]:

$$\begin{bmatrix} y_J & \cdots & y_1 \\ \vdots & \vdots & \vdots \\ y_{n-1} & \cdots & y_{n-J} \\ \bar{y}_2 & \cdots & \bar{y}_{J+1} \\ \vdots & \vdots & \vdots \\ \bar{y}_{n-J+1} & \cdots & \bar{y}_n \end{bmatrix} \begin{bmatrix} g_1 \\ \vdots \\ g_J \end{bmatrix} = - \begin{bmatrix} y_{J+1} \\ \vdots \\ y_n \\ \bar{y}_1 \\ \vdots \\ \bar{y}_{n-J} \end{bmatrix}, \tag{10}$$

here $M < J < n - M$ and ' $-$ ' denotes the complex conjugate of a complex number. Now if the mth observation is missing, we can delete the corresponding rows from the left as well as from the right wherever y_m is appearing. For example if $J + 1 < m \leqslant n - J - 1$, then (10) can be written as

$$\begin{bmatrix} y_J & \cdots & y_1 \\ \vdots & \vdots & \vdots \\ y_{m-2} & \cdots & y_{m-J} \\ y_{m+J} & \cdots & y_{m+1} \\ \cdots & \cdots & \cdots \\ y_{n-1} & \cdots & y_{n-J} \\ \bar{y}_2 & \cdots & \bar{y}_{J+1} \\ \vdots & \vdots & \vdots \\ \bar{y}_{m-J} & \cdots & \bar{y}_{m-1} \\ \bar{y}_{m+2} & \cdots & \bar{y}_{m+J+1} \\ \vdots & \vdots & \vdots \\ \bar{y}_{n-J+1} & \cdots & \bar{y}_n \end{bmatrix} \begin{bmatrix} g_1 \\ \vdots \\ g_J \end{bmatrix} = - \begin{bmatrix} y_{J+1} \\ \vdots \\ y_{m-1} \\ y_{m+J+1} \\ \vdots \\ y_n \\ \bar{y}_1 \\ \vdots \\ \bar{y}_{m-J-1} \\ \bar{y}_{m+1} \\ \vdots \\ \bar{y}_{n-J} \end{bmatrix}. \tag{11}$$

For $m \leqslant J + 1$ or $m > n - J - 1$ it can be defined similarly. The system of equations can be written as

$$\boldsymbol{A}_m \boldsymbol{g} = -\boldsymbol{h}_m, \tag{12}$$

where the matrix $\boldsymbol{A}_m$ and the vector $\boldsymbol{h}_m$ depend on the value of the missing observation. Now the minimum norm solution of $\boldsymbol{g}$ is given by

$$\boldsymbol{g} = -(\boldsymbol{A}_m^{\mathrm{H}}\boldsymbol{A}_m)^{-}\boldsymbol{A}_m^{\mathrm{H}}\boldsymbol{h}_m, \tag{13}$$

here 'H' and '$-$' denote the complex conjugate transpose of a matrix and pseudo-inverse of a matrix, respectively, as given in [16]. Now if we use the usual linear prediction notations, $\boldsymbol{R} = \boldsymbol{A}_m^{\mathrm{H}}\boldsymbol{A}_m$ and $\boldsymbol{r} = -\boldsymbol{A}_m^{\mathrm{H}}\boldsymbol{h}$, then it can be easily shown similarly as Tufts and Kumaresan [24], that in the noiseless data

$$\boldsymbol{g} = -\sum_{i=1}^{M} \frac{\boldsymbol{u}_i^{\mathrm{H}}\boldsymbol{r}}{\gamma_i}\boldsymbol{u}_i, \tag{14}$$

where $\gamma_1 \geqslant \cdots \geqslant \gamma_M > \gamma_{M+1} = \cdots = \gamma_J = 0$ are the eigen values of $\boldsymbol{R}$ and $\boldsymbol{u}_i$, $i = 1, \ldots, J$ are the corresponding orthonormal eigen vectors of $\boldsymbol{R}$. Form the prediction error polynomial equations with the vector $\boldsymbol{g}$ as follows:

$$H(z) = 1 + g_1 z + \cdots + g_J z^J = 0. \tag{15}$$

Eq. (15) has J roots. It can be shown in the same way as Kumaresan [7] that in case of noiseless data, out of J roots of (15), M of them will be at $\mathrm{e}^{\mathrm{i}\omega_j}$ for $j = 1, \ldots, M$ and $J - M$ other roots will have magnitudes strictly less than one and will be distributed uniformly over the unit circle.

In case of noisy data, first estimate $\boldsymbol{g}$ from (14), form the polynomial equation (15) and obtain the J roots of the prediction polynomial equation. Once we obtain the J roots, the estimates of ω_j's can be obtained from those M roots whose magnitudes are closest to one. Note that once we estimate the non-linear ω_j's the linear A_j's can be estimated very easily by the linear regression technique as given below. Model (1) can be written as

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \mathrm{e}^{\mathrm{i}\omega_1} & \cdots & \mathrm{e}^{\mathrm{i}\omega_M} \\ \vdots & \vdots & \vdots \\ \mathrm{e}^{\mathrm{i}n\omega_1} & \cdots & \mathrm{e}^{\mathrm{i}n\omega_M} \end{bmatrix} \begin{bmatrix} A_1 \\ \vdots \\ A_M \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_M \end{bmatrix}. \tag{16}$$

Let us write (16) as

$$\boldsymbol{Y} = \boldsymbol{\Omega}\boldsymbol{A} + \boldsymbol{E}.$$

If the matrix $\boldsymbol{\Omega}$ is completely known then the least-squares estimators, $\hat{\boldsymbol{A}}$, of $\boldsymbol{A}$ can be written as (see [9])

$$\hat{\boldsymbol{A}} = (\boldsymbol{\Omega}^{\mathrm{H}}\boldsymbol{\Omega})^{-1}\boldsymbol{\Omega}^{\mathrm{H}}\boldsymbol{Y}.$$

4. Cross-validation approach

In this section we observe how we can use the missing value technique discussed in Section 2 for model order selection of model (1). A practical and satisfactory method in model selection for small samples is the cross-validation approach, originally proposed by Lachenbruch [13] and Stone [21–23]. Dawid [6] suggested that in certain circumstances CV might lead to inconsistency but considerable interest has been shown recently in the use of cross validatory procedure because of its satisfactory performances in small samples. In the exponential signal model (1), Rao [15] proposed the following CV technique:

(1) For any fixed L, leave out one of the observation, say y_m, and replace it by $Y_m(L)$. Then for any choice of L and $Y_m(L)$, compute

$$R(Y_m(L), L) = \min_{\boldsymbol{A},\omega} \left\{ \sum_{k=1, k\neq m}^{n} \left| y_k - \sum_{j=1}^{L} A_j \mathrm{e}^{\mathrm{i}\omega_j k} \right|^2 + \left| Y_m(L) - \sum_{j=1}^{L} A_j \mathrm{e}^{\mathrm{i}\omega_j m} \right|^2 \right\}, \tag{17}$$

where $\boldsymbol{A} = (A_1, \ldots, A_L)$ and $\omega = (\omega_1, \ldots, \omega_L)$.

(2) For any given L, using a 'suitable computer program', find $\hat{Y}_m(L)$ such that

$$R(\hat{Y}_m(L), L) = \min_{Y_m(L)} R(Y_m(L), L), \tag{18}$$

which provides $\hat{Y}_m(L)$ is an estimate of y_m for a given L. Then comparing $\hat{Y}_m(L)$ with the observed y_m, the cross-validatory error is obtained as

$$R_*(L) = \sum_{m=1}^{n} |y_m - \hat{Y}_m(L)|^2 \quad \text{for } L = 1, \ldots, K. \tag{19}$$

Finally $\hat{M}$ is chosen such that it minimizes $R_*(L)$.

First, we would like to make some comments regarding the practical implementation and the numerical difficulties encountered about the above-mentioned CV algorithm. Observe that for a fixed $Y_m(L)$, (17) is a non-linear problem and it is not possible to obtain an explicit expression of $R(Y_m(L), L)$. Furthermore, the least-squares minimization problem of (17) is well known for its numerical instability (see, for example, [9,10,5,25]). The least-squares estimators often depend on the initial values and some times the iterative procedure may not even converge. Therefore, the minimization of $R(Y_m(L), L)$ with respect to $Y_m(L)$ may not be as

simple as it was suggested. We suggest the following CV procedure which is as follows:

(1) For fixed L, leave out one of the observation, say y_m.
(2) Estimate $\boldsymbol{A}$ and ω from $y_1, \ldots, y_{m-1}, y_{m+1}, \ldots, y_n$, using the missing value technique discussed in the previous section.
(3) Estimate y_m, say $\hat{y}_m(L)$, by

$$\hat{y}_m(L) = \sum_{j=1}^{L} \hat{A}_j e^{i\hat{\omega}_j m}, \tag{20}$$

where $\hat{A}_j$ and $\hat{\omega}_j$ are the estimates obtained from Step 2.

(4) Obtain the cross validatory error as Rao [15] suggested by

$$R^*(L) = \sum_{m=1}^{n} |y_m - \hat{y}_m(L)|^2 \quad \text{for } L = 1, \ldots, K. \tag{21}$$

(5) Finally $\hat{M}$ is chosen such that it minimizes $R^*(L)$.

Observe that our method is quite easy to implement in practice. Since the method discussed in Section 3, is non-iterative in nature, therefore the estimators of $\boldsymbol{A}$ and ω can be obtained quite easily in Step 2. Note that Rao's $\hat{Y}_M(L)$ is the best least-squares predictor of y_m for a given L. He did not require to estimate separately A_j's and ω_j's on estimating y_m. Since obtaining Rao's $\hat{Y}_m(L)$ is numerically difficult, we are approximating it by $\hat{y}_m(L)$, which is much easier to obtain. Intuitively, it seems

$$E(R^*(L)) \geqslant E(R_*(L)) \quad \text{for } L = 1, \ldots, K. \tag{22}$$

Since (22) it is true for all L, on estimating M it should not make much difference.

5. Numerical experiments and discussions

In this section we present some results of the numerical experiments. All the computations are performed in HP-9000 using the IMSL random number generator. We consider the following three models:

$$\text{Model 1: } y_k = \exp(\pi/4)\exp(i2\pi(0.50)k) + \exp(\pi/4)\exp(i2\pi(0.52)k) + e_k,$$

$$\text{Model 2: } y_k = \exp(\pi/4)\exp(i2\pi(0.50)k) + \exp(\pi/4)\exp(i2\pi(0.60)k) + e_k,$$

$$\text{Model 3: } y_k = \exp(\pi/4)\exp(i2\pi(0.50)k) + \exp(\pi/4)\exp(i2\pi(0.60)k) + \exp(\pi/2)\exp(i2\pi(0.20)k) + \exp(\pi/2)\exp(i2\pi(0.25)k) + e_k.$$

Note that for Models 1 and 2, the amplitudes are taken to be equal. The difference of the radian frequencies is more in Model 2 than in Model 1. It is well known that if the difference is more then the estimators are more accurate. Between, Models 2 and 3, two components are exactly same in both the models. Since Model 3 has more parameters compared to Model 2, it is expected that the estimators will be less accurate for Model 3 than Model 2. No such comparisons can be made between Models 1 and 3. The real and the imaginary parts of e_k's are normally distributed with zero mean and finite variance $\sigma^2/2$ and they are independent also. The sample sizes are taken as $n = 25$, 40 and 55 and SNR = 5, 10 and 15 dB. It is assumed that the maximum number of signals is 6, i.e. $K = 6$. For each data set we estimate M by AIC, MDL and by EDC. For EDC we take $C_n = \log n$, as suggested in [8]. We estimate M also by the proposed CV method. We use different order (within admissible range) J of the prediction equations in (10) to estimate the model parameters in presence of missing value and in turn estimating M. If the prediction order is J ($\geqslant 6$) then the corresponding CV method is denoted by CV(J). We take different values of J for different sample sizes. For each sample size and for each SNR and for each model, we replicate the process 500 times. The results are presented in Tables 1–3. We present the percentage of correct estimates (PCE), percentage of over estimates (POE) and the percentage of under estimates (PUE). The entry in each table represents the PUE, the PCE and the POE for different methods over 500 replications.

Comparing the tables it is observed that for fixed n as SNR increases the performances of almost all the methods improve and for fixed SNR as n increases the performances also improve except for AIC. It is observed that since the difference between the two frequencies in Model 2 is more than in

Table 1
Model 1

	SNR = 15			SNR = 10			SNR = 5		
Methods	PUE	PCE	POE	PUE	PCE	POE	PUE	PCE	POE
Sample size 25									
CV(6)	0	75	25	0	51	49	0	16	84
CV(8)	0	96	4	0	90	10	3	85	12
CV(10)	0	94	6	13	78	9	0	53	47
AIC	0	55	45	0	56	44	0	56	44
MDL	0	87	13	0	87	13	0	80	20
EDC	0	94	6	0	90	10	0	80	20
Sample size 40									
CV(6)	0	77	23	0	71	29	0	56	44
CV(8)	0	93	7	0	90	10	0	88	12
CV(10)	0	94	6	0	94	6	0	92	8
CV(13)	0	97	3	0	95	5	0	94	6
CV(15)	0	95	5	0	93	7	0	91	9
AIC	0	48	52	0	46	54	0	43	57
MDL	0	91	9	0	88	12	0	89	11
EDC	0	97	3	0	93	7	0	94	6
Sample size 55									
CV(6)	0	85	15	0	85	15	0	83	17
CV(8)	0	90	10	0	90	10	0	88	12
CV(12)	0	96	4	0	96	4	0	95	5
CV(18)	0	98	2	0	98	2	0	97	3
CV(20)	0	97	3	0	96	4	0	96	4
AIC	0	42	58	0	40	60	0	38	62
MDL	0	87	13	0	85	15	0	85	15
EDC	0	97	3	0	97	3	0	93	7

Model 1, estimators of the number of signals are more accurate in Model 2 than in Model 1. Between Models 2 and 3, since the number of parameters are more in Model 3, it is observed that the estimators become less accurate for Model 3. Among the different methods, the performance of AIC is quite poor. The percentage of correct estimation does not exceed 60% for AIC even at high SNR. As sample size increases the PCE decrease for AIC for fixed SNR. The inconsistency of the AIC is quite prominent in this situation. MDL works quite satisfactory at high SNR, although the percentage of correct estimation decreases between 80 to 85% if the SNR is low. The performance of EDC is quite satisfactory with $C_n = \log n$. The consistency of MDL or EDC can be verified from the experiment. The performance of the CV method depends on the order of the prediction equations. As the prediction order J increases, the percentage of the correct estimate increases up to a certain point then it starts decreasing. It is observed in our experiment that if $J \approx n/3$, then the percentage of correct detection of CV(J) is maximum, although we cannot give any theoretical justification of this. It seems more work is needed in this direction. It is also observed that if $J \approx n/3$, then the CV(J) works much better than AIC or MDL and the performance of CV(J) is quite comparable with the best performed EDC and in certain cases CV(J) works marginally better than the best performed EDC.

It is well known that the usual cross-validation method does not give consistent estimates of the

Table 2
Model 2

Methods	SNR = 15			SNR = 10			SNR = 5		
	PUE	PCE	POE	PUE	PCE	POE	PUE	PCE	POE
Sample size 25									
CV(6)	0	57	43	0	55	45	0	52	48
CV(8)	0	100	0	0	100	0	0	100	0
CV(10)	0	97	3	0	95	5	0	93	7
AIC	0	56	44	0	53	47	0	52	48
MDL	0	90	10	0	88	12	0	86	14
EDC	0	99	1	0	99	1	0	99	1
Sample size 40									
CV(6)	0	70	30	0	69	31	0	65	35
CV(8)	0	91	9	0	91	9	0	90	10
CV(10)	0	93	7	0	92	8	0	93	7
CV(13)	0	98	2	0	97	3	0	97	3
CV(15)	0	95	5	0	95	5	0	94	6
AIC	0	46	54	0	44	56	0	43	57
MDL	0	87	13	0	87	13	0	86	14
EDC	0	98	2	0	98	2	0	97	3
Sample size 55									
CV(6)	0	86	14	0	84	16	0	83	17
CV(8)	0	94	6	0	91	9	0	89	11
CV(12)	0	95	5	0	93	7	0	95	5
CV(18)	0	100	0	0	100	0	0	100	0
CV(20)	0	98	2	0	97	3	0	96	4
AIC	0	37	63	0	37	63	0	35	65
MDL	0	88	12	0	88	12	0	86	14
EDC	0	100	0	0	99	0	0	99	0

model order estimation, but sometimes it might provide consistent estimates for the model order estimation. In this case it seems if we take $J \approx n/3$, it might provide consistent estimates of M.

Let us consider computational complexities of the different methods. From computational point of view AIC, MDL or EDC are much faster than the CV method if K is small. If the maximum model dimension K is large, then AIC, MDL or EDC become more complicated, because they need to solve non-linear equations in a $2K$ dimensions. The solutions may depend on the initial values and they may lead to a local minimum rather than a global minimum. On the other hand if the sample size is large then the CV computations become more time consuming, although implementation is quite simple. The proposed CV computation does not require any initial values.

6. Conclusions

In this paper we consider the estimation of the number of signals of the undamped exponential models, which is a very important problem in Statistical Signal Processing. We propose a new CV approach based on the missing value technique. It is observed that the proposed CV method works quite well and it performs better than the usual AIC and MDL. The performance is quite comparable to the best performed EDC. Another point we would like to point out that although, we assume that the errors

Table 3
Model 3

Methods	SNR = 15			SNR = 10			SNR = 5		
	PUE	PCE	POE	PUE	PCE	POE	PUE	PCE	POE
Sample size 25									
CV(6)	0	62	38	0	52	48	0	22	78
CV(8)	0	97	3	0	93	7	0	88	12
CV(10)	0	92	8	0	88	12	0	63	37
AIC	0	53	47	0	54	46	0	54	46
MDL	0	87	13	0	85	15	0	80	20
EDC	0	93	7	0	89	11	0	82	18
Sample size 40									
CV(6)	0	58	42	0	63	37	0	60	40
CV(8)	0	88	12	0	88	12	0	85	15
CV(10)	0	90	10	0	92	8	0	89	11
CV(13)	0	97	3	0	96	4	0	95	5
CV(15)	0	94	6	0	93	7	0	92	8
AIC	0	46	54	0	41	59	0	41	59
MDL	0	83	17	0	85	15	0	83	17
EDC	0	94	6	0	95	5	0	91	9
Sample size 55									
CV(6)	0	82	18	0	82	18	0	82	18
CV(8)	0	90	10	0	89	11	0	88	12
CV(12)	0	93	7	0	90	10	0	90	10
CV(18)	0	98	2	0	98	2	0	97	3
CV(20)	0	96	4	0	95	5	0	95	5
AIC	0	41	59	0	39	61	0	37	63
MDL	0	84	16	0	84	16	0	83	17
EDC	0	96	4	0	95	5	0	92	8

are complex Gaussian random variables, but it is not being used except implicitly at (21). Therefore, it seems the CV procedure should work even if the errors are not from a complex Gaussian random variable but from any light tail distributions. Where as for AIC, MDL or EDC the exact distributional assumptions of the error random variables are very important for their implementation. Comparing all the points it is suggested that the CV method can be used to estimate the number of signals for the model (1) in many situations.

Acknowledgements

Most of the work was done when the first author was visiting The Pennsylvania State University and The University of Texas at San Antonio. He would like to thank the respective departments for their helps. The authors would like to thank two referees for their helpful and very constructive suggestions.

References

[1] H. Akaike, Information theory and an extension of the maximum likelihood principle, Proceeding of the Second International Symposium of Information Theory, 1973, pp. 267–281.
[2] H. Akaike, A new look at statistical model identification, IEEE Trans. Automat. Control 19 (1974) 716–723.
[3] Z.D. Bai, X.R. Chen, P.R. Krishnaiah, L.C. Zhao, Asymptotic properties of EVLP estimator for superimposed exponential signals in noise, Technical Report, Center for Multivariate Analysis, Pittsburgh, USA, 1987, pp. 87–19.

[4] M.S. Bartlett, A note on the multiplying factors for various χ^2 approximations, J. Roy. Statist. Soc. Ser. B 16 (1954) 296–298.
[5] Y. Breslar, A. Macovski, Exact maximum likelihood parameter estimation of superimposed exponential signals in noise, IEEE Trans. Acoust. Speech. Signal Process. 34 (5) (1986) 1081–1089.
[6] A.P. Dawid, Contribution to discussion of a paper by M. Stone, J. Roy. Statist. Soc. Ser. B 36 (1974) 136–138.
[7] R. Kumaresan, Estimating the parameters of exponentially damped or undamped sinusoidal signals in noise, Ph.D. Thesis, University of Rhode Island, 1982.
[8] D. Kundu, Detecting the number of signals for undamped exponential models using information theoretic criterion, J. Statist. Comput. Simulation 44 (1992) 117–131.
[9] D. Kundu, Estimating the parameters of undamped exponential signals, Technometrics 35 (1993) 215–218.
[10] D. Kundu, A modified Prony algorithm for estimating damped or undamped exponential signals, Sankhya A 56 (3) (1994) 524–544.
[11] D. Kundu, A. Mitra, Consistent method of estimating superimposed exponential signals, Scand. J. Statist. 22 (1) (1995) 73–82.
[12] D. Kundu, A. Mitra, On asymptotic behavior of the least squares estimators and the confidence intervals of the superimposed exponential signals, Signal Process. 72 (2) (1999) 129–139.
[13] P.A. Lachenbruch, Discriminant Analysis, Haffner Press, New York, 1975.
[14] D.N. Lawley, Tests of significance of the latent roots of the covariance and correlation matrices, Biometrica 43 (1956) 128–136.
[15] C.R. Rao, Some recent results in signal detection, in: S.S. Gupta, J. Berger (Eds.), Statistical Decision Theory and Related Topics, Vol. IV, Springer, New York, 1988, pp. 319–332.
[16] C.R. Rao, Linear Statistical Inference and its Applications, 2nd Edition, Wiley, New York, 1973.
[17] J. Risannen, Modeling by shortest data description, Automatica 14 (1978) 465–471.
[18] J. Risannen, A universal prior for the integers and estimation by minimum description length, Ann. Statist. 11 (1983) 417–431.
[19] G. Schwartz, Estimating the dimension of a model, Ann. Statist. 6 (1978) 461–464.
[20] P. Stoica, List of references on spectral analysis, Signal Process. 31 (1993) 329–340.
[21] M. Stone, Cross Validatory choice and assessment of statistical prediction, J. Roy. Statist. Soc. Ser. B 36 (1974) 111–147.
[22] M. Stone, An asymptotic equivalence of choice of model by cross validation and Akaike's criterion, J. Roy. Statist. Soc. Ser. B 39 (1977a) 44–47.
[23] M. Stone, Asymptotic for and against cross validation, Biometrika 64 (1977b) 29–35.
[24] D.W. Tufts, R. Kumaresan, Estimation of frequencies of multiple sinusoids: making linear prediction perform like maximum likelihood, Proc. IEEE 70 (9) (1982) 975–989.
[25] J.M. Varah, On fitting exponentials by non-linear least squares, SIAM J. Sci. Statist. Comput. 6 (1985) 30–44.
[26] L.C. Zhao, P.R. Krishnaiah, Z.D. Bai, On detection of the number of signals in the presence of white noise, J. Multivariate Anal. 20 (1986) 1–25.