An Empirical Comparison of Arc-Cosine Distance, Generalised Fisher Ratio and Normalised Entropy Criteria for Model Selection

S. Zheng and C. G. Molina
Anglia Polytechnic University
Computer Science Department
East Road, Cambridge CB1 1PT, England
Email: szheng/cmolina@csd.anglia.ac.uk

Abstract

Three model selection criteria, the arc-cosine distance, the generalised Fisher ratio and the normalised entropy, are applied to several data sets sampled from different mixture models. Their performance is investigated and their ability to measure the mutual information between the components in a mixture model is compared. Experimental results show that the Arc-Cosine distance criterion outperforms the other two criteria.

1 Introduction

Choosing the right number of components in a mixture model is a difficult problem and is encountered in many applications where a priori knowledge about the distribution of the observed data is not available. When the objective is to partition a given data set into differentiate groups, it is implicitly assumed that each group can be approximately sampled from one of the mixture components.

Various self-adaptive algorithms have been proposed during the last few years. Most of these techniques are based on *pruning* [9] the components of which have minor importance, or *growing* [7] the model by adding new components when novelty is detected. Others are based on criteria such as *Akaike Information Criterion* (AIC)[1], *Bayesian Information Criterion* (BIC) [11], *Information Complexity Criterion* (ICC) [2]. Recently, Richardson *et al.* [10] developed a new technique which adapts the number of components and their parameters jointly by using reversible jump Markov Chain Monte Carlo (MCMC) methods.

0-7803-5060-X/98/$10.00 © 1998 IEEE

In this article, three model selection criteria, the *arc-cosine distance* [12], the *generalised Fisher ratio* (GFR) [13] and the *normalised entropy* [4] are compared by applying them to data sets sampled from different synthetic mixture models. Their performances are investigated and their abilities to measure the mutual information between the components in a mixture model are compared. In [4] the experimental results have shown improved performance using the normalised entropy criterion compared to AIC, BIC and ICC, so consequently these latter methods are not included in our comparison.

2 Three Criteria for Mixture Model Selection

A mixture model is defined as a linear combination of Gaussian distributions $\phi_i(\mathbf{x})$ in the form

$$y(\mathbf{x}) = \sum_{i=1}^{M} \lambda_i \phi_i(\mathbf{x}), \tag{1}$$

where M is the number of Gaussian distributions in the model, λ_i are the *mixing weights*, which are normalised,

$$\sum_{i=1}^{M} \lambda_i = 1 \quad and \quad \lambda_i \geq 0 \quad (i = 1, ..., M), \tag{2}$$

and $\phi_i(\mathbf{x})$ is the *pdf* of ith normalised Gaussian distribution $\mathbf{x} \sim N(\boldsymbol{\mu}_i, C_i)$, that is,

$$\phi_i(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|C_i|^{1/2}} \exp\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T C_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\}. \tag{3}$$

2.1 Arc-Cosine distance

The *arc-cosine* distance [12] is used to measure the distance between two Gaussian distributions ϕ_1 and ϕ_2 and is defined as

$$\Omega = \arccos\left(\frac{< \phi_1, \phi_2 >}{\| \phi_1 \|_2 \cdot \| \phi_2 \|_2}\right) \tag{4}$$

where $< \cdot, \cdot >$ is the inner product given by

$$< \phi_1, \phi_2 >= \int \phi_1(\mathbf{x}) \cdot \phi_2(\mathbf{x}) d\mathbf{x} \tag{5}$$

and $\| \cdot \|_2$ is the L_p-norm ($p = 2$), defined as

$$\| \phi \|_p = (\int [\phi(\mathbf{x})]^p d\mathbf{x})^{1/p} \tag{6}$$

The arc-cosine distance reduces to a *Mahalanobis distance* when the two Gaussian distributions have the same covariance matrices. In the more general case, if we want to measure the arc-cosine distance between a mixture distribution Φ_M and a Gaussian distribution ϕ_j, the inner product reduces to a weighted sum of inner products between the Gaussians

$$< \Phi_M, \phi_j >= \sum_{i=1}^{M} \lambda_i < \phi_i(\mathbf{x}), \phi_j(\mathbf{x}) > \tag{7}$$

where

$$< \phi_i(\mathbf{x}), \phi_j(\mathbf{x}) >= \frac{\| \phi_k \|_1}{\| \phi_i \|_1 \cdot \| \phi_j \|_1} \cdot \mathbf{A} \tag{8}$$

with

$$\mathbf{A} = \exp\left(-\frac{1}{2}(\boldsymbol{\mu}_i^T C_i^{-1} \boldsymbol{\mu}_i + \boldsymbol{\mu}_j^T C_j^{-1} \boldsymbol{\mu}_j - \boldsymbol{\mu}_k^T C_k^{-1} \boldsymbol{\mu}_k)\right) \tag{9}$$

and

$$C_k = (C_i^{-1} + C_j^{-1})^{-1} \tag{10}$$

$$\boldsymbol{\mu}_k = (\boldsymbol{\mu}_i C_i^{-1} + \boldsymbol{\mu}_j C_j^{-1}) \cdot C_k \tag{11}$$

In order to find the number of components in a given mixture model, an overestimated initial model is generated and its parameters are updated on the data set using the EM algorithm [5]. Then a merging technique [12] is applied to reduce the number of components in the overestimated model.

When a Gaussian mixture $\Phi_m(\mathbf{x})$ with m Gaussian components is to be merged into one, the *zero*, *first* and *second* order moments of the new merged Gaussian $\phi_t(\mathbf{x})$ are calculated as

$$\lambda_t = \sum_{k=1}^{m} \lambda_k \tag{12}$$

$$\boldsymbol{\mu}_t = \int \Phi_m(\mathbf{x})\, \mathbf{x}\, d\mathbf{x} = \sum_{k=1}^{m} \lambda_k \int \phi_k(\mathbf{x})\, \mathbf{x}\, d\mathbf{x} = \sum_{k=1}^{m} \lambda_k \boldsymbol{\mu}_k \tag{13}$$

$$C_t = \int \Phi_m(\mathbf{x}) \cdot (\mathbf{x} - \boldsymbol{\mu}_t)(\mathbf{x} - \boldsymbol{\mu}_t)^T d\mathbf{x} = \sum_{k=1}^{m} \lambda_m \int \phi_m(\mathbf{x})(\mathbf{x} - \boldsymbol{\mu}_t)(\mathbf{x} - \boldsymbol{\mu}_t)^T d\mathbf{x}$$

$$= \sum_{k=1}^{m} \lambda_m (C_k + (\boldsymbol{\mu}_k - \boldsymbol{\mu}_t)(\boldsymbol{\mu}_k - \boldsymbol{\mu}_t)^T) \tag{14}$$

If the arc-cosine distance between a mixture of two Gaussians

$$\Phi_2(\mathbf{x}) = \lambda_i \phi_i(\mathbf{x}) + \lambda_j \phi_j(\mathbf{x}) \tag{15}$$

and their new merged component $\phi_{new}(\mathbf{x})$ is smaller than a threshold, then the merging is accepted and this reduces the number of Gaussians by one in the generated mixture model. Thus, calculating the number of components is equivalent to find the best threshold. A probabilistic function $P_i(N_i|\Delta_i)$ is defined as follows

$$P_i(N_i|\Delta_i) = \frac{length(\Delta_i)}{\pi/2} \tag{16}$$

which gives the highest probability to the best distance threshold. Here Δ_i is a continuous range of thresholds (in the distance range $\Omega \in [0, \pi/2]$) under which the number of components is equal to N_i. After the probabilities $P_i(N_i|\Delta_i)$ for each interval are calculated, the underlying number of components in the original mixture distribution can be estimated

$$N_{components} = max_{N_i}(P_i(N_i|\Delta_i)). \tag{17}$$

2.2 The Generalised Fisher ratio

Fisher ratio [6] is a common method for feature extraction in pattern recognition and defined as:

$$f = trace(C_1 + C_2)^{-1} det[(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T] \tag{18}$$

where C_1, C_2, $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are the class covariance matrices and means for the feature. The term $C_1 + C_2$ is a measure of the spread of the data around the mean of each class and is referred to as an *intra-class scatter matrix*. The second term $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T$ is a measure of the separation between the classes and is referred to as an *inter-class scatter matrix*. It is easy to see that this measure is maximum when the inter-class separation is maximised and the intra-class spread is minimised.

In [8], the *generalised Fisher ratio* (GFR) has been defined for a multi-class problem as:

$$F = \frac{1}{m(m-1)} \frac{\sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \lambda_j f_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \lambda_j}, \quad i \neq j \tag{19}$$

where m is the number of classes, λ_i and λ_j are the mixing proportions for classes i, j , and f_{ij} is the Fisher ratio between classes i, j as defined in equation (18).

In the case of modelling the probability of each class with a single Gaussian, the whole probability distribution can be expressed as a mixture model.

Assuming that there are as many classes as components, the optimal model $y(\mathbf{x})$ that matches the right number of components in the underlying mixture distribution is the one that maximises the GFR. In order to find the maximum GFR that corresponds to the optimal number of Gaussians, the *generalised Fisher ratio model selection algorithm* [13] starts with an overestimated mixture model with more Gaussians than the number of expected components in the underlying density distribution; then reduces the number of Gaussians using the same merging technique [12]. In the proposed algorithm, only the two nearest Gaussians are merged every time according to their Euclidean distance. This reduces the number of Gaussians by one in the mixture model. If the GFR of the new model is greater than that of the original model, then the new one is accepted, otherwise the new model is rejected and the algorithm tries merging other pairs of Gaussians from the original mixture model. This process is repeated until no more merging can be achieved.

2.3 Normalised Entropy

The *normalised entropy criterion* (NEC) is proposed by Celeux, G. *et al.* [4] for assessing the number of components in a mixture model. It is defined as

$$NEC(m) = \frac{E(m)}{L(m) - L(1)} \tag{20}$$

where $L(m) = C(m) + E(m)$ with

$$\begin{aligned} C(m) &= \sum_{k=1}^{m}\sum_{i=1}^{n} t_{ik} ln[\lambda_k \phi_k(\mathbf{x}_i)] \\ E(m) &= -\sum_{k=1}^{m}\sum_{i=1}^{n} t_{ik} ln(t_{ik}) \geq 0 \end{aligned} \tag{21}$$

$C(m)$ is the *classification maximum likelihood* [3] and

$$t_{ik} = \frac{\lambda_k \phi_k(\mathbf{x}_i)}{\sum_{k=1}^{m} \lambda_k \phi_k(\mathbf{x}_i)}. \tag{22}$$

The entropy term $E(m)$ measures the overlap of the mixture components. Experiments reported in [4] show improved results using the NEC method compared to the performance of other criteria such as AIC, BIC, ICC.

3 Numerical Experiments and Results

In this section, the arc-cosine distance, the generalised Fisher ratio and the normalised entropy criteria are applied to the synthetic problems. We compare their abilities to match the number of components in several synthetic

	Experiments		Arc-cos	GFR	NEC
	Experiment 1.1				
	$\mu_1 = 0.2, \mu_2 = 0.5, \mu_3 = 0.8$	M = 2	0	0	5
	$\sigma_1 = 0.06, \sigma_2 = 0.05, \sigma_3 = 0.07$	M = 3	95	90	85
	$\lambda_1 = 0.3, \lambda_2 = 0.4, \lambda_3 = 0.3$	M = 4	5	10	10
	Experiment 1.2				
100	$\mu_1 = 0.15, \mu_2 = 0.35, \mu_3 = 0.6,$	M = 3	0	5	15
	$\mu_4 = 0.8, \sigma_1 = 0.04, \sigma_2 = 0.05,$	M = 4	95	95	75
	$\sigma_3 = 0.045, \sigma_4 = 0.03, \lambda_1 = 0.25,$	M = 5	5	0	5
	$\lambda_2 = 0.2, \lambda_3 = 0.25, \lambda_4 = 0.3$	M = 6	0	0	5
	Experiment 1.1				
	$\mu_1 = 0.2, \mu_2 = 0.5, \mu_3 = 0.8$				
	$\sigma_1 = 0.06, \sigma_2 = 0.05, \sigma_3 = 0.07$	M = 3	100	95	90
	$\lambda_1 = 0.3, \lambda_2 = 0.4, \lambda_3 = 0.3$	M = 4	0	5	10
	Experiment 1.2				
	$\mu_1 = 0.15, \mu_2 = 0.35, \mu_3 = 0.6,$				
200	$\mu_4 = 0.8, \sigma_1 = 0.04, \sigma_2 = 0.05,$	M = 3	0	0	5
	$\sigma_3 = 0.045, \sigma_4 = 0.03, \lambda_1 = 0.25,$	M = 4	100	100	90
	$\lambda_2 = 0.2, \lambda_3 = 0.25, \lambda_4 = 0.3$	M = 5	0	0	5

Table 1: Percent frequencies of choosing M components for the univariate distributions.

underlying mixture models. 20 data sets are sampled from each mixture model. Three different types of Gaussian mixtures are considered: (1) univariate Gaussian mixtures, (2) bivariate Gaussian mixtures with covariance matrices which are scalar multiples of the identity matrix, and (3) bivariate Gaussian mixtures with diagonal covariance matrices. For each type, two distributions are simulated and for each simulated distribution, two data sets are considered with different sizes.

We start with an overestimated mixture model with more Gaussians than the components expected in the underlying model. The parameters of the generated mixtures are updated by running the EM algorithm for 20 epochs. Using the criteria described in section 2, the number of the components in the underlying mixture model are estimated by reducing the Gaussians in the generated model.

All the results for univariate and bivariate simulated distributions are listed in tables 1, 2, 3 and 4. Not surprisingly, all the criteria perform better for larger data sets. When running the algorithms with these criteria, the NEC method has a complexity of $O(m \cdot n)$ for each calculation of the criterion value for every mixture model; it is very time consuming when the size of data set n is large. The arc-cosine distance and the GRF method have a complexity of $O(m^2)$. Since the size of the data set n is much bigger than the number of components m, the Arc-cosine distance and the GFR method

Experiments (Data set size = 200)		**Arc-cos**	**GFR**	**NEC**
Experiment 2.1				
$\mu_1 = [0.8, 0.5]$, $\mu_2 = [0.5, 0.2]$,				
$\mu_3 = [0.2, 0.5]$, $\mu_4 = [0.5, 0.8]$,				
$\sigma_1 = 0.045$, $\sigma_2 = 0.047$,				
$\sigma_3 = 0.043$, $\sigma_4 = 0.032$	M = 3	0	0	5
$\lambda_i = 0.25, i = 1, ..., 4.$	M = 4	100	100	95
Experiment 2.2				
$\mu_1 = [0.76, 0.65]$, $\mu_2 = [0.76, 0.35]$,				
$\mu_3 = [0.50, 0.20]$, $\mu_4 = [0.24, 0.35]$,				
$\mu_5 = [0.24, 0.65]$, $\mu_6 = [0.50, 0.80]$,				
$\sigma_1 = 0.050$, $\sigma_2 = 0.046$, $\sigma_3 = 0.040$,	M = 4	0	5	10
$\sigma_4 = 0.036$, $\sigma_5 = 0.030$, $\sigma_6 = 0.022$,	M = 5	0	15	15
$\lambda_i = 1/6, i = 1, ..., 6.$	M = 6	100	80	75
Experiment 3.1				
$\mu_1 = [0.2, 0.2]$, $\mu_2 = [0.5, 0.8]$,				
$\mu_3 = [0.8, 0.3]$,				
$C_1 = [0.005, 0; 0, 0.006]$,				
$C_2 = [0.010, 0; 0, 0.009]$,				
$C_3 = [0.006, 0; 0, 0.008]$,				
$\lambda_1 = 0.3$, $\lambda_2 = 0.4$, $\lambda_3 = 0.3$.	M = 3	100	100	100
Experiment 3.2				
$\mu_1 = [0.78, 0.60]$, $\mu_2 = [0.68, 0.26]$,				
$\mu_3 = [0.32, 0.26]$, $\mu_4 = [0.21, 0.60]$,				
$\mu_5 = [0.50, 0.80]$,				
$C_1 = [0.0050, 0; 0, 0.0060]$,				
$C_2 = [0.0075, 0; 0, 0.0080]$,				
$C_3 = [0.0062, 0; 0, 0.0075]$,				
$C_4 = [0.0045, 0; 0, 0.0050]$,				
$C_5 = [0.0050, 0; 0, 0.0055]$,	M = 3	5	5	20
$\lambda_1 = 0.2$, $\lambda_2 = 0.18$, $\lambda_3 = 0.22$,	M = 4	15	20	20
$\lambda_4 = 0.15$, $\lambda_5 = 0.25$.	M = 5	80	75	60

Table 2: Percent frequencies of choosing M components for the bivariate distributions with data set size of 200.

Experiments (Data set size = 300)		**Arc-cos**	**GFR**	**NEC**
Experiment 2.1				
$\mu_1 = [0.8, 0.5]$, $\mu_2 = [0.5, 0.2]$,				
$\mu_3 = [0.2, 0.5]$, $\mu_4 = [0.5, 0.8]$,				
$\sigma_1 = 0.045$, $\sigma_2 = 0.047$,				
$\sigma_3 = 0.043$, $\sigma_4 = 0.032$				
$\lambda_i = 0.25, i = 1, ..., 4.$	M = 4	100	100	95
Experiment 2.2				
$\mu_1 = [0.76, 0.65]$, $\mu_2 = [0.76, 0.35]$,				
$\mu_3 = [0.50, 0.20]$, $\mu_4 = [0.24, 0.35]$,				
$\mu_5 = [0.24, 0.65]$, $\mu_6 = [0.50, 0.80]$,				
$\sigma_1 = 0.050$, $\sigma_2 = 0.046$, $\sigma_3 = 0.040$,	M = 4	0	0	5
$\sigma_4 = 0.036$, $\sigma_5 = 0.030$, $\sigma_6 = 0.022$,	M = 5	0	5	10
$\lambda_i = 1/6, i = 1, ..., 6.$	M = 6	100	95	85
Experiment 3.1				
$\mu_1 = [0.2, 0.2]$, $\mu_2 = [0.5, 0.8]$,				
$\mu_3 = [0.8, 0.3]$,				
$C_1 = [0.005, 0; 0, 0.006]$,				
$C_2 = [0.010, 0; 0, 0.009]$,				
$C_3 = [0.006, 0; 0, 0.008]$,				
$\lambda_1 = 0.3$, $\lambda_2 = 0.4$, $\lambda_3 = 0.3$.	M = 3	100	100	100
Experiment 3.2				
$\mu_1 = [0.78, 0.60]$, $\mu_2 = [0.68, 0.26]$,				
$\mu_3 = [0.32, 0.26]$, $\mu_4 = [0.21, 0.60]$,				
$\mu_5 = [0.50, 0.80]$,				
$C_1 = [0.0050, 0; 0, 0.0060]$,				
$C_2 = [0.0075, 0; 0, 0.0080]$,				
$C_3 = [0.0062, 0; 0, 0.0075]$,				
$C_4 = [0.0045, 0; 0, 0.0050]$,				
$C_5 = [0.0050, 0; 0, 0.0055]$,	M = 3	0	0	10
$\lambda_1 = 0.2$, $\lambda_2 = 0.18$, $\lambda_3 = 0.22$,	M = 4	5	5	10
$\lambda_4 = 0.15$, $\lambda_5 = 0.25$.	M = 5	95	95	80

Table 3: Percent frequencies of choosing M components for the bivariate distributions with data set size of 300.

Experiments	Arc-cos	GFR	NEC
Experiment 1.1	$\mu(M) = 3.0250$ $\sigma(M) = 0.1581$	$\mu(M) = 3.0750$ $\sigma(M) = 0.2067$	$\mu(M) = 3.0750$ $\sigma(M) = 0.3499$
Experiment 1.2	$\mu(M) = 4.0250$ $\sigma(M) = 0.1581$	$\mu(M) = 3.9750$ $\sigma(M) = 0.1581$	$\mu(M) = 4$ $\sigma(M) = 0.5064$
Experiment 2.1	$\mu(M) = 4$ $\sigma(M) = 0$	$\mu(M) = 4$ $\sigma(M) = 0$	$\mu(M) = 3.9750$ $\sigma(M) = 0.1581$
Experiment 2.2	$\mu(M) = 6$ $\sigma(M) = 0$	$\mu(M) = 5.85$ $\sigma(M) = 0.4267$	$\mu(M) = 5.7250$ $\sigma(M) = 0.5986$
Experiment 3.1	$\mu(M) = 3$ $\sigma(M) = 0$	$\mu(M) = 3$ $\sigma(M) = 0$	$\mu(M) = 3$ $\sigma(M) = 0$
Experiment 3.2	$\mu(M) = 4.85$ $\sigma(M) = 0.4267$	$\mu(M) = 4.825$ $\sigma(M) = 0.4465$	$\mu(M) = 4.55$ $\sigma(M) = 0.7494$

Table 4: Means μ and standard deviations σ of the number of components M calculated by the three different methods.

run considerably faster than the NEC method.

From tables 1, 2 and 3, when the components in the underlying mixture models are well separated, the three criteria work well and the differences between them are not significant. However, for the second example in the experiment 3, all the methods tend to underestimate the number of components in the underlying mixture model when the size of data set is small and when there is some overlapping between the components. Especially, the NEC method has a tendency to merge some of the components even when the size of the data set is increased. The GFR method has the intermediate position between the NEC method and the arc-cosine distance, and the latter has produced best results in most cases with more accurate means and lower variances for the number of components chosen M, as shown in table 4.

4 Conclusion

This paper presents a study of the performance of three model selection criteria, the Arc-cosine distance, the generalised Fisher ratio and the normalised entropy, by applying them to data sets sampled from different synthetic mixture models. These three criteria measure the mutual information or the overlap between the components in a mixture model. The numerical experiments show that the NEC has a tendency to underestimate the number of components when there is some overlap between the components. The GFR method produces better results compared to the NEC method in our experiments and the arc-cosine distance has the best performance.

References

[1] H. Akaike. A new look at the statistical identification model. *IEEE Transactions on Automatic Control*, 19:pp716–723, 1974.

[2] H. Bozdoğan. Choosing the number of component clusters in the mixture model using a new information complexity criterion of the inverse-Fisher information matrix. In O. Opitz and B. Lausen, editors, *Information and Classification*, pages 40–54, Heidelberg, 1993. Springer – Verlag.

[3] P.G. Bryant. Large-sample results for optimisation based clustering methods. *Journal of Classification*, 8:pp31–44, 1991.

[4] G. Celeux and G. Soromenho. An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13:pp195–212, 1996.

[5] N.M. Dempster, A.P. Laird and D.B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society B*, 39:pp1–38, 1977.

[6] R. Duda and P. Hart. *Pattern Classification and Scene Analysis.* Wilsey, New York, 1973.

[7] S.E. Fahlman and C. Lebiere. The cascade-correlation learning architecture. In D.S. Touretzky, editor, *Neural Information Processing Systerms 2*, pages 524–532. Morgan-Kaufmann, 1990.

[8] S. Krishnan and P.V.S. Rao. Feature selection for pattern classification with Gaussian mixture models: A new objective criterion. *Pattern Recognition Letters*, 17:pp803–809, 1996.

[9] R. Reed. Pruning algorithms - a survey. *IEEE Transactions on neural networks*, 4(5):pp740–747, Sept. 1993.

[10] S. Richardson and P.J. Green. On Bayesian analysis of mixtures with an unknown number of components. *Journal of Royal Statistical Society B*, 59(4):pp731–792, 1997.

[11] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:pp461–464, 1978.

[12] C. M. Stow, A. C. T. Kennington, G. Molina, and W. J. Fitzgerald. Experimental Issues of Functional Merging on Probability Density Estimation. In *Artificial Neural Networks*, pages 123–128, London, 1997. Institution of Electrical Engineers.

[13] S. Zheng and C.G. Molina. Measuring tree-ring parameters using the generalised Fisher ratio. *Accepted by IX European Signal Processing Conference September 8 - 11, Island of Rhodes, Greece*, 1998.