The International Journal of Biostatistics

Volume 7, *Issue* 1 2011 *Article* 7

Classification of Stationary Signals with Mixed Spectrum

Pedro Saavedra, *University of Las Palmas de Gran Canaria*
Angelo Santana-del-Pino, *Universidad de Las Palmas de Gran Canaria*
Carmen N. Hernández-Flores, *Universidad de Las Palmas de Gran Canaria*
Juan Artiles-Romero, *Universidad de Las Palmas de Gran Canaria*
Juan J. González-Henríquez, *Universidad de Las Palmas de Gran Canaria*

Recommended Citation:
Saavedra, Pedro; Santana-del-Pino, Angelo; Hernández-Flores, Carmen N.; Artiles-Romero, Juan; and González-Henríquez, Juan J. (2011) "Classification of Stationary Signals with Mixed Spectrum," *The International Journal of Biostatistics*: Vol. 7: Iss. 1, Article 7.
DOI: 10.2202/1557-4679.1288

Classification of Stationary Signals with Mixed Spectrum

Pedro Saavedra, Angelo Santana-del-Pino, Carmen N. Hernández-Flores, Juan Artiles-Romero, and Juan J. González-Henríquez

Abstract

This paper deals with the problem of discrimination between two sets of complex signals generated by stationary processes with both random effects and mixed spectral distributions. The presence of outlier signals and their influence on the classification process is also considered. As an initial input, a feature vector obtained from estimations of the spectral distribution is proposed and used with two different learning machines, namely a single artificial neural network and the LogitBoost classifier. Performance of both methods is evaluated on five simulation studies as well as on a set of actual data of electroencephalogram (EEG) records obtained from both normal subjects and others having experienced epileptic seizures. Of the different classification methods, Logitboost is shown to be more robust to the presence of outlier signals.

KEYWORDS: classification, stationary processes, mixed spectrum, LogitBoost

Author Notes: The authors acknowledge Dr. José M. Pacheco for the valuable comments and the critical review of the article.

1 The problem of classifying stationary signals

In recent years, the demand for non-invasive diagnostic procedures has stimulated an increasing interest in the development of methods for classifying biomedical signals. The devices currently available in hospitals and healthcare facilities can easily produce signals such as electrocardiograms (EKG), electroencephalogram (EEG), electromyogram (EMG) or Doppler. These signals must be analysed quickly and accurately in order for the physician to be able to correctly diagnose the patient. In many cases the methods used for the analysis of these signals assume they are stationary with absolutely continuous spectral distributions. When the assumption of stationarity is admissible, the analysis of these signals in the frequency domain can lead to better discrimination than in the time domain (Shumway and Unger, 1974, Alagon, 1989, Vilar and Pértega, 2004, Chinipardaz and Cox, 2004). These authors consider the problem of discriminating between two sets of signals, where each one is generated by a Gaussian linear process, and propose a classification method based on the Kullback-Leibler discrimination information rate. Vilar and Pértega (2004) deal with the problem by using nonparametric estimators for the spectral density functions. Alagon (1989) analyzes the validity of the evoked potentials of the EEG for the diagnosis of certain neurological diseases.

The assumption of absolutely continuous spectral distribution is often unrealistic when biomedical signals are considered. Biological systems usually have periodicities in their behaviour patterns (Ahdesmäki, Lähdesmäki, Pearson, Huttunen, and Yli-Harja, 2005) and many of these periodicities can be detected as atoms in the spectrum of the signal. For example, the records of the electroencephalogram of healthy subjects in conditions of rest normally have a spectrum with discrete components. Pardey, Roberts, and Tarassenko (1996), in a review of modelling techniques for EEG analysis, consider mixed spectral distributions. Bhansali (1970) uses the mixed spectrum to analyse the annual record of the number of the Canadian lynx trapped in the Mackenzie River district of North-West Canada for the period 1821-1934 (Canadian lynx data set).

Another usual assumption in the analysis of biomedical signals is that of considering that all the times series measured on subjects of the same population are generated by the same stationary process. In a more realistic approach, Diggle and Al-Wasel (1993, 1997) suggest that the time series corresponding to levels of LH hormone in blood samples from subjects of a given population can be represented by a random effects model. This model means that the underlying spectrum of the stochastic process representing the time variation in hormone concentration varies randomly between subjects. The model is based on the asymptotic representation of the periodogram of linear processes and involves a population parameter (the population spectrum), a random component specific for each subject, and a term

related to the residuals of each periodogram. Hernández-Flores, Artiles-Romero, and Saavedra-Santana (1999) used a more general model to estimate the population spectrum by means of the bootstrap. Saavedra, Hernández, and Artiles (2000), Saavedra, Hernández, Luengo, Artiles, and Santana (2008) developed a theory for analyzing sets of time series in the frequency domain. Luengo, Hernández, and Saavedra (2006) compared the patterns of time series generated by two populations.

Modern techniques such as classification and regression trees (Breiman, Friedman, Stone, and Olshen, 1984), artificial neural networks (Ripley, 1996) and, more recently, Boosting (Freund and Schapire, 1996, Friedman, Hastie, and Tibshirani, 2000, Hastie, Tibshirani, and Friedman, 2001, Bühlman, 2006) open up a new approach to the problem of classifying time series. For instance, boosting decision trees have been used for the classification of gene expression data in Ben-Dor, Bruhn, Friedman, Nachman, Schummer, and Yakhini (2000) and Dudoit, Fridlyand, and Speed (2002). Both studies compare the original AdaBoost algorithm that was proposed by Freund and Schapire (1996) to other classifiers, and both recognize that the results obtained are not very impressive. However, Dettling and Buhlmann (2003) dramatically improved the results by performing a selection of variables and using the LogitBoost algorithm instead of AdaBoost.

In this paper we deal with the problem of discriminating between two sets of stationary signals. Any method that aims to make correct classifications of biomedical signals must take into account the aforementioned variability between subjects and the possible presence of periodicities in the series. Thus, we propose the use of a general framework for the modelling of the signals, assuming that these have been generated by processes with both random effects and mixed spectral distribution, including the possible existence of outlier signals (signals generated by patterns different and unrelated to the target population, possibly corresponding to anomalous subjects). We use four simulation studies and actual EEG data to show how an adequate modelling of the signals, combined with a discriminant method capable of incorporating the information provided by the model, can significantly improve the rate of correct classifications.

The class of stationary processes we consider for modelling random effects is described in Section 2. A review of a method due to Kooperberg, Stone, and Truong (1995) for the estimation of mixed spectra is presented in Section 3. In Section 4, three classification methods are described based, respectively, on Kullback-Leibler information, neural networks and Logitboost. The last two methods use the same feature vector which is proposed in Section 4.2, formed by combining the singular and absolutely continuous components of the spectral distributions estimated from each series. Finally four simulation studies and actual records of EEG corresponding to healthy subjects in normal state and with epileptic episodes are used to compare the performance of the classification achieved by each method.

DOI: 10.2202/1557-4679.1288

2 Random effects model for the set of time series

We consider a population of objects A, such that on each $a \in A$ a stationary process $X_t(a)$ can be observed. In order for this process to be general enough for modelling biomedical signals with possible periodicities in its behavior patterns, while allowing for random variability between subjects with respect to a common population pattern, we assume that this process is of the form:

$$X_t(a) = \sum_{k=1}^{p} R_{a,k} \cos\left(t\lambda_{a,k} + \phi_{a,k}\right) + Y_t(a) \tag{1}$$

where $\lambda_{a,k}$ and $R_{a,k}$ are random variables such that $0 < \lambda_{a,k} \leq \pi$, $R_{a,k} > 0$ and conditionally to $a \in A$, $\phi_{a,k}$ are independent and uniformly distributed random variables on the interval $[-\pi, \pi]$. Moreover, $Y_t(a)$ is a second order stationary process having an absolutely continuous spectral distribution, $\{f_a(\omega) : |\omega| \leq \pi\}$ being the set of spectral density functions. Saavedra et al. (2008) show that $\{Y_t(\cdot) : t \in \mathbb{Z}\}$ can be represented as a linear process of random coefficients. Thus, each spectral distribution function $F_a(\omega) : |\omega| \leq \pi$ can be considered as a realization of a stochastic process on the space A. In addition, it can be expressed (Kooperberg et al., 1995) by:

$$F_a(\lambda) = \int_{-\pi}^{\lambda} f_a(\omega)\, d\omega + \sum\nolimits_{\omega \leq \lambda} d_a(\omega) \tag{2}$$

Here, $d_a(\omega)$ are the so called spectral lines which take on the form $d_a(\omega) = R_{a,k}^2 \Big/ 2$ if $\omega = \pm\lambda_{a,k}$ and $d_a(\omega) = 0$ otherwise. We refer to $\pm\lambda_{a,k}$, $1 \leq k \leq p$ as the *atoms* of the spectral distribution for the subject a.

3 Spectral estimation

Data for the analysis are obtained from a sample of objects $a_1, \ldots, a_n$ randomly selected from the population A. On each object a_i, a stationary process $X_t(a_i)$ of the form (1) is observed at the same times $t = 1, \ldots, T$. Therefore, the data in the time domain take on the form:

$$\{X_t(a_i) : \ i = 1, \ldots, n \ ; \ t = 1, \ldots, T\}$$

while in the frequency domain, we have

$$\left\{ I_{a_i}^{(T)}\left(\omega_j\right) : i = 1, \ldots, n; \, j = 1, \ldots, [T/2] \right\}$$

where $\omega_j = 2\pi j/T$ are the Fourier frequencies, and $I_{a_i}^{(T)}(\omega_j)$ the periodogram corresponding to the i-th object. For any object $a \in A$, the periodogram is defined by:

$$I_a^{(T)}(\omega) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} X_t(a) \exp(-\mathbf{i}\omega t) \right|^2 \quad : \quad -\pi \leq \omega \leq \pi \tag{3}$$

Under the assumptions that all atoms are located at Fourier frequencies, i.e. have the form $\lambda_{a,k} = \omega_j$ for some j, and all time series are Gaussian, it can be shown (Brillinger, 1981, Theorem 5.2.6) that:

$$I_a^{(T)}(\omega_j) = \left\{ f_a(\omega_j) + \frac{T}{2\pi} d_a(\omega_j) \right\} U_{a,j} \tag{4}$$

where, conditionally to each a, the $U_{a,j}$ are asymptotically independent and follow an approximately exponential distribution with unit mean if $j < T/2$, or the χ^2 distribution with one degree of freedom if $j = T/2$. The function $g_a(\omega) = f_a(\omega) + \frac{T}{2\pi} d_a(\omega)$ is called the *mean function*. We use cubic splines and indicator functions, according to Kooperberg et al. (1995), to estimate the spectral densities $f_a(\omega)$ and the line spectrum $d_a(\omega)$. See the appendix for details.

Regarding the conditions for the decomposition (4), from a practical point of view the assumption of the $\lambda_{a,k}$ being located at Fourier frequencies is not too restrictive in the context of biomedical signals. With these kind of signals it is usually possible to have records with a sampling frequency greater enough for possible atoms being always in or very close to a Fourier frequency. Also, biomedical signals are usually affected by noises associated to different sources (Pander, 2008): electromagnetic effects on the measuremente devices, movements, electrical activity in the near tissues ... In practice, filters must be used to suppress these noises, particularly those of impulsive nature. Once the series have been correctly filtered it is reasonable to assume gaussianity.

4 Classification methods

In this section we describe briefly three different classifiers for a partition $\{A_0, A_1\}$ of the population A under study based on a data set of the form:

$$\{ (X_t(a_i)\ ,\ G_i)\ :\ i = 1, \ldots, n\ ;\ t = 1, \ldots, T\ \} \tag{5}$$

where $a_1, \ldots, a_n$ is a random sample of objects of A, $X_t(a_i)$ is a time series consisting of the realization of a stationary process of the form (1) on the object a_i, and

DOI: 10.2202/1557-4679.1288

$G_i \in \{1,0\}$ is the class label ($G_i = 1$ or 0 depending on $a_i \in A_1$ or $a_i \in A_0$). We consider first a classifier based on the Kullback-Leibler discrimination information, which requires that the stationary processes generating the signals have an absolutely continuous spectral distribution. The second and third classifiers are based on artificial neural networks and LogitBoost, respectively. Both can be used in more general scenarios, as is the case of stationary processes with mixed spectral distribution. These classifiers need a feature vector as input to proceed with the classification procedure. The same vector, based on the estimated spectrum, is used for both classifiers and is described in 4.2.

4.1 Classification based on the Kullback-Leibler discrimination information

As a first scenario for the classification problem let us assume that all time series in the class $A_k : k = 0,1$ have been generated by the same stationary process with an absolutely continuous spectral distribution given by the spectral density function $f_k(\omega)$. Let $p_k(\mathbf{x})$ be the probability density function corresponding to a random signal $x = (X_1(a), \ldots, X_T(a))$ measured on an object at class A_k (note that $p_k(\mathbf{x})$ is independent of a). One classical measure of disparity between $p_1(\mathbf{x})$ and $p_0(\mathbf{x})$ is the the Kullback-Leibler (KL) discrimination information, defined by:

$$I_T(p_1; p_0) = T^{-1} E_{p_1}\left[log\left(\frac{p_1(x)}{p_0(x)}\right)\right]$$

According to Shumway and Unger (1974) and Kakizawa, Shumway, and Taniguchi (1998), under certain conditions $I_T(p_1; p_0)$ can be asymptotically approximated by:

$$D(f_1; f_0) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left[\frac{f_1(\lambda)}{f_0(\lambda)} - \log\left(\frac{f_1(\lambda)}{f_0(\lambda)}\right) - 1\right] d\lambda \tag{6}$$

In order to classify the signals $\{X_t(a_i) : i = 1, \ldots, n\}$, Kakizawa et al. (1998) use a measure of the form (6) in the following algorithm:

1. First, for every $a \in A$, an adequate spectral estimator $\hat{f}_a$ of the true spectral density f_a is obtained from the time series $(X_1(a), \ldots, X_T(a))$.
2. Then, the disparity between $(X_1(a), \ldots, X_T(a))$ and the class $A_k : k = 1,0$ is evaluated by computing $D(f_k; \hat{f}_a)$.
3. Finally, the object a is classified into A_1 if $D(f_0; \hat{f}_a) - D(f_1; \hat{f}_a) > 0$ or into A_0 otherwise.

In this algorithm f_k, $k = 0, 1$ are assumed to be known. If this is not the case, they must be estimated from training samples of objects whose membership to each group is known. According to Vilar and Pértega (2004), we can obtain $\hat{f}_a$ by smoothing the corresponding periodogram via local polynomial techniques. Likewise, $f_1(\lambda)$ and $f_0(\lambda)$ can be estimated using the same techniques on the averaged periodograms of the training signals.

4.2 The feature vector

Using the procedure of Kooperberg et al. (1995) cited in 3, we estimate the discrete and absolutely continuous components of the spectral distribution from the set of periodograms. In what follows, we represent by D_a the set $\{\hat{d}_a(\lambda_{a,j})\}_j$ of estimates of the spectrum lines, and by $\hat{f}_a(\lambda)$ the estimates of the $f_a(\lambda)$. In order to establish a feature vector for the methods of classification that we will consider below, we fix a set of frequencies $0 < \phi_1 < \ldots < \phi_K < \pi$ and for every object a_i we define a vector $\mathbf{V}_i$:

$$\begin{aligned} \mathbf{V}_i \;=\; \Bigg(& \#(D_{a_i}) \;,\; \textstyle\sum_j \hat{d}_{a_i,j} \;;\; \frac{1}{T}\sum_{j=1}^{T} \hat{f}_{a_i}(\omega_j) \;, \\ & \frac{\int_0^{\phi_1} \hat{f}_{a_i}(\omega)\, d\omega}{\int_0^{\pi} \hat{f}_{a_i}(\omega)\, d\omega} \;,\; \cdots \;,\; \frac{\int_{\phi_{K-1}}^{\phi_K} \hat{f}_{a_i}(\omega)\, d\omega}{\int_0^{\pi} \hat{f}_{a_i}(\omega)\, d\omega} \Bigg) \end{aligned} \tag{7}$$

where:

- $\#(D_{a_i})$ is the number of atoms in the estimated spectral distribution;
- $\sum_j \hat{d}_{a_i,j}$ represent the contribution of the atoms to the overall spectral power;
- $\frac{1}{T}\sum_{j=1}^{T} \hat{f}_{a_i}(\omega_j)$ is the mean value of the estimated spectral density function in the object a_i;
- $\int_{\phi_{k-1}}^{\phi_k} \hat{f}_{a_i}(\omega)\, d\omega$ is the contribution of the frequency band $[\phi_{k-1}, \phi_k]$ to the spectral power of the absolutely continuous component of the spectrum.

We now describe the classifiers based on neural networks and LogitBoost. In both cases, the data set available for the construction of the classifier have the form $(\mathbf{V}_i, G_i) \;:\; i = 1, \ldots, n$ where $\mathbf{V}_i$ is the defined feature and G_i the class label.

DOI: 10.2202/1557-4679.1288

4.3 Classification using artificial neural networks

Artificial neural networks (ANNs) may be defined as structures comprised of densely interconnected adaptive simple processing elements (neurons) that are capable of performing massively parallel computations for data processing and knowledge representation (Schalkoff, 1997). ANNs can be trained to recognize patterns and the nonlinear models developed during training allow neural networks to generalize their conclusions and to apply them to patterns not previously encountered. Figure 1 shows a single hidden layer feed-forward neural network, which will be used in this paper. It consists of: *(i)* an input layer with neurons representing the feature $\mathbf{V}_i = (V_{i,1}, \ldots, V_{i,p})$ defined in (7), *(ii)* an output layer with neurons representing the dependent variables and *(iii)* one hidden layer containing neurons to help capture the nonlinearity in the data. Each neuron in the hidden layer sums its input signals after multiplying them by the strengths of the respective connection weights $\alpha_{m,j}$ and computes its output $Z_{i,m}$ as a function of the sum:

$$Z_{i,m} = \sigma\left(\alpha_{m,0} + \sum_j \alpha_{m,j} V_{i,j}\right) \; : \; m = 1, \ldots, M$$

where $\sigma(x)$ is some activation function that is necessary to transform the weighted sum of all signals impinging onto a neuron. As activation function we have used the logistic $\ell(z) = exp(z)/(1+exp(z))$. Neural networks also need a measure of fit between what the network predicts for each training pattern and the target value, or observed value, for that pattern. We have considered the entropy (deviance) as measure of fit.

4.4 LogitBoost classifier with decision trees

The boosting procedures introduced by Freund and Schapire (1996) are a powerful classification technique, especially in high dimensional spaces (Bühlman, 2006). Their aim is to produce an accurate combined classifier from a previous sequence of weak classifiers. In each boosting iteration $m = 1, \ldots, M$, objects incorrectly classified at the previous step have their weights increased, whereas weights are decreased for those correctly classified. Thus, the m-th classifier $h^{(m)}$ built in step m is forced to focus more on objects whose current classifications had been difficult to obtain at previous iterations. The resulting classifier has the following form:

$$C^M(V_i) = \text{sign}\left(\sum_{m=1}^{M} \alpha_m \cdot h^{(m)}(V_i)\right)$$

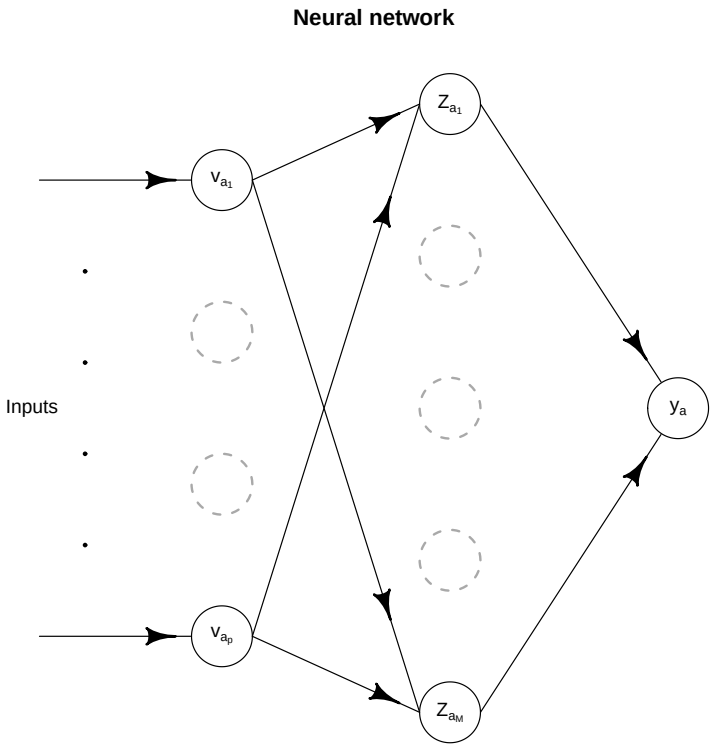

Figure 1: Schematic of a single hidden layer, feed-forward neural network

As weak classifiers $h^{(m)}$ we use decision trees with two terminal nodes (Breiman et al., 1984). A description of the LogitBoost algorithm is provided below.

Step 1: Initialization. Start with an initial committee function $H^{(0)}(V_i) \equiv 0$ and initial probabilities $p^{(0)}(V_i) = P(Y_i = 1 | V_i) = 1/2$ for all $i = 1, \ldots, n$.

Step 2: Iterations LogitBoost. For $m = 1, \ldots, M$ repeat:

A. Fitting the weak classifier:

I. Compute for $i = 1, \ldots, n$ the weights $w_i^{(m)}$ and the auxiliary variable $z_i^{(m)}$ by:

$$w_i^{(m)} = p^{(m-1)}(V_i) \cdot \left(1 - p^{(m-1)}(V_i)\right)$$

$$z_i^{(m)} = \frac{Y_i - p^{(m-1)}(V_i)}{w_i^{(m)}}.$$

DOI: 10.2202/1557-4679.1288

II. Fit the weak classifier using weighted least squares:

$$h^{(m)} = \arg\min_{h} \sum_{i=1}^{n} w_i^{(m)} \left(z_i^{(m)} - h\left(V_i\right) \right)^2$$

B. Updating

$$H^{(m)}\left(V_i\right) = H^{(m-1)}\left(V_i\right) + \frac{1}{2} h^{(m)}\left(V_i\right)$$

$$p^{(m)}\left(V_i\right) = \left(1 + \exp\left(-2 \cdot H^{(m)}\left(V_i\right)\right)\right)^{-1}$$

C. Output of the value assessed by the classifier:

$$C^{M}\left(\mathbf{V}_i\right) = sign\left(H^{(m)}(\mathbf{V}_i)\right)$$

5 Numerical study

We now proceed to evaluate the performance of the different classification methods presented in the previous section. This evaluation is carried out by using both simulated data and real data. For simulation we have considered five scenarios in order of increasing generality. Actual data were obtained from EEG records measured on healthy and epileptic subjects. In all cases, the dataset was split into a training data set and a validation set. Using this last one, misclassification rates were obtained, summarized in medians, interquartile ranges and maxima.

5.1 Simulations

The simulations were carried out in five different scenarios. In each one, 400 time series were generated, 200 in each group (cases and controls). All time series were generated by stationary processes of the form (1), with $Y_t\left(a_1, a_2\right) = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \varepsilon_t + a_1\,\varepsilon_{t-1} + a_2\,\varepsilon_{t-2}$, $\{\varepsilon_t \,:\, t \in \mathbb{Z}\}$ being a standard Gaussian white noise with variance 1, β_1, β_2 fixed coefficients, and $\mathbf{a} = (a_1, a_2)'$ the vector of random coefficients, which has a bivariate probability distribution $N_2\left(\mu, \mathbf{C}\right)$. In those cases in which a mixed spectra was considered, the number of atoms λ_{aj} and its values were fixed in each group, while the corresponding amplitudes R_{aj} had a multivariate normal distribution with the identity as covariance matrix.

Table 1 specifies the parameters of the singular and absolutely continuous components of the time series. In all cases, the number of observations was $T = 500$. From the 400 simulated series, 200 were randomly selected to build the training set. Using this set, the classifiers described in the previous section were constructed, namely the one based on the Kullback-Leibler divergence (KL), the artificial neural network, and the Boosting procedure. With the remaining 200 time series misclassification rates were obtained. The entire procedure (simulation and calculation of classification error rates) was iterated 100 times. The misclassification rates for each repetition are summarized as medians, interquartile ranges and maximum. For the feature defined in (7), used by the ANN and LogitBoost, we take $K = 3$ and $\phi_k = k/10 : k = 1,2,4$. The scenarios considered are briefly described below.

Scenario 1. The time series were generated by stationary processes of the form (1) without a singular component ($X_t = Y_t$). In addition, all ARMA process parameters were fixed (C = 0).

Scenario 2. The same pattern as in Scenario 1, except that now the coefficients of the process (part MA) were randomly selected with non-singular covariance matrix (see Table 1).

Scenario 3. This is the same scenario as above, but allowing in both classes the 5% of the observations to be outlier signals.

Scenario 4. A stationary process with mixed spectral distribution, the absolutely continuous part being the same as in scenario 2, was used to generate the time series. The parameters of the singular part are shown in Table 1.

Scenario 5. Same as in scenario 4, with 5% of outlier signals, as in scenario 3.

The simulation study was performed using the R software package, version 2.10 (R Development Core Team, 2010). The results are summarized in table 2.

DOI: 10.2202/1557-4679.1288

Table 1: Elements of the simulated signals

		Group	
Scenario	**Spectral Component**	A_1 **(Cases)**	A_0 **(Controls)**
1	Absolutely continuous	$\beta_1 = 0.8\,;\beta_2 = -0.5$ $\mu = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ $\mathbf{C} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$	$\beta_1 = 0.5\,;\beta_2 = -0.5$ $\mu = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$ $\mathbf{C} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$
2	Absolutely continuous	$\beta_1 = 0.8\,;\beta_2 = -0.5$ $\mu = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ $C = \begin{pmatrix} 1/6 & 1/6 \\ 1/6 & 1/6 \end{pmatrix}$	$\beta_1 = 0.5\,;\beta_2 = -0.5$ $\mu = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$ $\mathbf{C} = \begin{pmatrix} 1/6 & 1/6 \\ 1/6 & 1/6 \end{pmatrix}$
3	Absolutely continuous	As in scenario 2	As in scenario 2
	Proportion of outlier signals	5%	5%
4	Singular	$R \cong N_3\left((8;9;12)\,,\mathbf{I}_3\right)$ $\lambda = (0.38;0.18;0.025)$	$R \cong N_4\left((5;8;7;9)\,,\mathbf{I}_4\right)$ $\lambda = (0.38;0.12;0.6;0.025)$
	Absolutely continuous	As in scenario 2	As in scenario 2
5	Singular	As in scenario 4	As in scenario 4
	Absolutely continuous	As in scenario 2	As in scenario 2
	Proportion of outlier signals	5%	5%

Table 2: Evaluation of the classifiers according the five scenarios

Scenario	**Kullback-Leibler**		**Neural Network**		**LogitBoost**	
	Median (IQR)	Max	Median (IQR)	Max	Median (IQR)	Max
1	0	0	0.5 (0 ; 1.0)	3.0	0.5 (0 ; 1)	2.5
2	10.0 (8.0 ; 12.5)	15.5	3.5 (2.5 ; 4.5)	10.0	2.5 (1.5 ; 3.5)	7.0
3	11.7 (10.5 ; 14.0)	19.5	7.0 (5.5 ; 8.0)	14.5	4.7 (4.0 ; 6.0)	9.5
4	10.5 (9.0 ; 11.5)	16.0	4.0 (3.0 ; 5.0)	7.0	2.5 (2.0 ; 3.5)	6.5
5	12.0 (10.5 ; 13.5)	17.0	6.5 (5.0 ; 8.0)	13.5	4.5 (4.0 ; 5.5)	9.0

5.2 Diagnosis of epilepsy

Epilepsy is one of the most common neurological disorders. An electroencephalogram (EEG) signal is used to detect epilepsy because this signal reflects the electrical activity of the brain which has been related to this condition. Epilepsy is characterized by recurrent peaks in the EEG signal. In this section we consider two sets of EEG data (Andrzejak, Lehnertz, Mormann, Rieke, David, and Elger, 2001) corresponding to normal and epileptic subjects. Each set contains 100 single channel EEG segments of 23.6 seconds in duration, with $T = 4,096$ times of observation. The segments were selected and cut from multichannel records that were collected after a visual inspection of artifacts such as muscle activity or eye movement. Figure 2 shows fragments of 6 seconds from signals from an epileptic condition (a) and from a normal condition (b). The corresponding estimates of the spectral distribution functions using the method described in Section 3 are also shown (c and d). It can be seen that the spectral distribution for the normal record contains atoms with positive mass (signals with more regularities). In order to compare the three classifiers, we used the procedure described next:

1. The set of 200 EEG records was randomly split into two subsets: 60% of the signals (120) were used to build classifiers using each of the three methods. Classification error rates were obtained from the remaining 40% of the signals (80).
2. Step 1 was iterated 100 times.
3. The obtained misclassification rates were summarized as medians, interquartile ranges and maximums.

The results are shown in Table 3.

Table 3: EEG signals: misclassification rates.

Kullback-Leibler		ANN		LogitBoost	
Median (IQR)	Max	Median (IQR)	Max	Median (IQR)	Max
25.0 (20.0 ; 30.0)	38.7	6.2 (5.0 ; 7.5)	11.2	3.7 (2.5 ; 5.0)	8.7

DOI: 10.2202/1557-4679.1288

6 Discussion

The papers on classifiers based on Kullback-Leibler divergence (KL) assume that the time series are generated within each group by the same Gaussian linear process. Such an assumption has been considered in the three first scenarios of the simulation study. In the first scenario, as can be seen in Table 2, the 200 time series were correctly classified by the method used by Vilar and Pértega (2004). This classifier appears superior to those based on ANN and LogitBoost, possibly because it uses more effectively the conditions of linearity, Gaussianity and homogeneity. The introduction of intra-class heterogeneity in the second scenario significantly increases the misclassification rates of the classifier based on KL (the median of errors increases from 0% to 10.0%) while those based on the ANN and LogitBoost moderately increase, being similar those of the latter two. The contamination by a 5% of outliers under the third scenario tends to increase the misclassification rates of all procedures, but without doubt the most robust is the LogitBoost. The greater robustness of procedures based on decision trees versus neural networks has already been pointed out by Hastie et al. (2001, p. 313). In these three scenarios there are no discrete components, so ANN and LogitBoost do no use more information that KL does.

In practice it is often unrealistic to assume that actual biomedical signals are generated by processes with absolutely continuous spectral distribution. This is what occurs with the EEG signals described in 5.2, in most of which atoms are found (fig 2). Therefore, in scenarios 4 and 5 we have considered signals generated by processes with mixed spectra. It can be seen that the misclassification rates corresponding to the ANN and LogitBoost are similar and significantly lower than those obtained for the KL procedure. This is just what one would expect since KL can not use the information on the mixed spectrum. The differences in the observed error rates show the improvement achieved when the classification method adequately models the spectrum of the signal. In the fourth scenario, the highest rates of error are 7.0% for ANN and 6.5% for LogitBoost. The introduction of outliers on stage 5 increased the maximum error to 13.5% for the ANN and 9% for the LogitBoost. This result seems to confirm the greater robustness of LogitBoost compared to neural networks methods. In the analysis of the EEG, the high error rates for the KL method are undoubtedly attributable to the discrepancy between the series characteristics and the assumptions of this classification model. The lower classification error of LogitBoost with respect to ANN (maximum error rate of 11.2% for ANN compared to 8.7% for LogitBoost) could be attributed to neural networks being less robust to the presence of outliers.

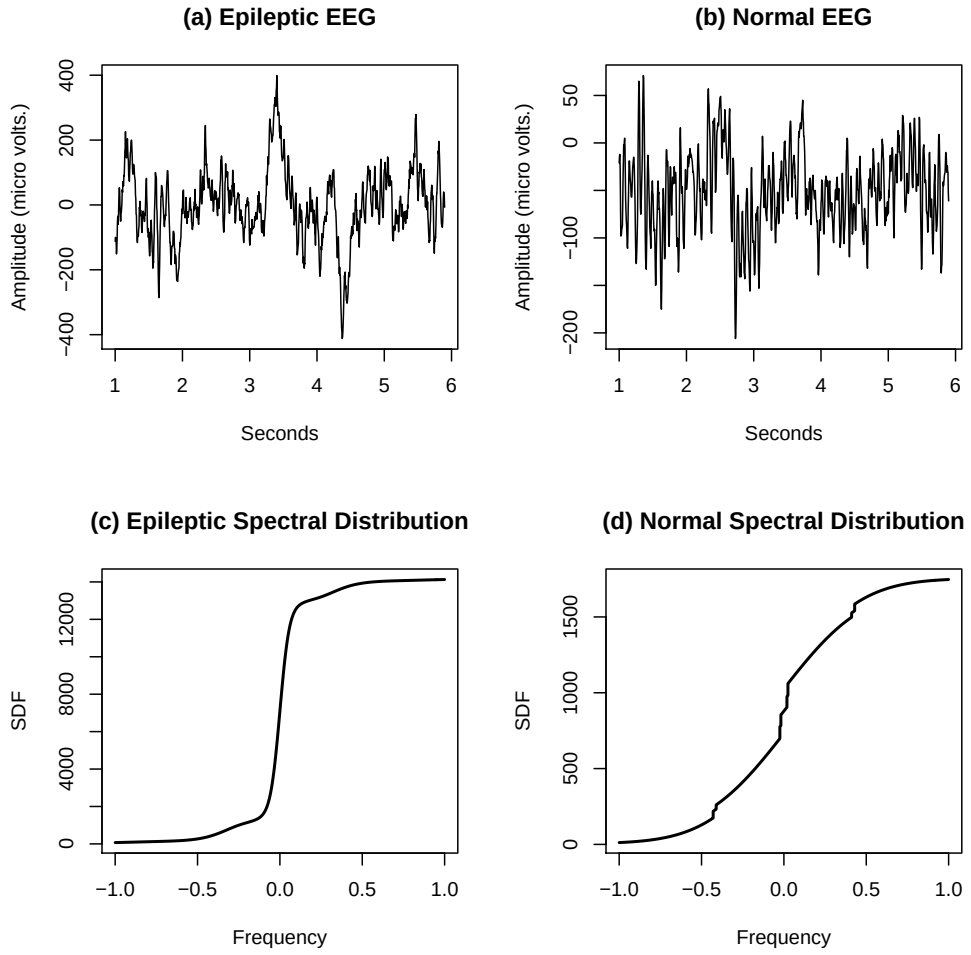

Figure 2: EEG signals corresponding to an epileptic subject (a) and to a normal subject (b) and their corresponding spectral distribution functions (c) and (d).

In practice, it is usually unknown if the true spectrum of the signals is mixed or absolutely continuous. In such conditions it is advisable that the classification method try to extract all the available information in the signals. For these reason, we recommend assuming that the signals are generated by mixed spectra, summarizing the spectral estimation in a feature as given in (7) and constructing a classifier based on the logitboost procedure.

DOI: 10.2202/1557-4679.1288

Appendix: Details of the spectral estimation

The representation (4) for each periodogram $I_a(\omega_j)$ determines a likelihood function l_a, which depends on the spectral density function $f_a(\lambda)$ and spectrum line $d_a(\lambda)$. According to Kooperberg et al. (1995), we parametrically model these functions by means of cubic splines as follows:

1. For the spectral density, set $\log f_a(\lambda) = \sum_{k=1}^{K_C} \beta_k B_k(\lambda)$, where $\{B_1, \ldots, B_{K_C}\}$ is a base for the space of twice continuously differentiable functions $s(\lambda)$ on $[0, \pi]$ such that the restriction of $s(\lambda)$ to each of the intervals $[\tau_{k-1}, \tau_k]$ of a certain partition $0 \leq \tau_1 < \ldots < \tau_{K_c} \leq \pi$ is a cubic polynomial, and moreover where the first and third derivatives of $s(\lambda)$ are equal to zero at 0 and π.
2. For the spectrum line, we consider $B_{k+K_C}(\lambda) = \delta_{\nu_k}(\lambda)$ for $1 \leq k \leq K_d$, where K_d is a nonnegative integer, $\delta_\nu(\cdot)$ is the indicator function and $\nu_1 < \ldots < \nu_{K_d}$ is a sequence of Fourier frequencies. The mean function is then modelled as $\log g_a(\lambda) = \sum_{k=1}^{K_C+K_d} \beta_k B_k(\lambda)$

It is easy to deduce that $d_a(\lambda) = \frac{2\pi}{T} f_a(\lambda) \left[\exp\left(\sum_{k=K_C+1}^{K_C+K_d} \beta_k B_k(\lambda)\right) - 1\right]$. Therefore, for any object $a \in A$, the log-likelihood (omitting constants) takes on the form:

$$l_a(\beta) = \sum_j \left\{ \frac{\delta_\pi(\omega_j)}{2} - 1 \right\} \left[\sum_{k=1}^{K_C+K_d} \beta_k B_k(\omega_j) + I_a(\omega_j) \exp\left\{ \sum_{k=1}^{K_C+K_d} \beta_k B_k(\omega_j) \right\} \right]$$

The maximum likelihood estimate $\hat{\beta}_a$ is given as usual by $l_a\left(\hat{\beta}_a\right) = \max_\beta l_a(\beta)$. From the maximum likelihood estimates $\hat{f}_a(\lambda)$ and $\hat{d}_a(\lambda)$, we obtained $\hat{F}_a(\lambda)$ using (2). The procedure for the selection of nodes is documented in Kooperberg et al. (1995).

References

Ahdesmäki, M., H. Lähdesmäki, R. Pearson, H. Huttunen, and O. Yli-Harja (2005): "Robust detection of periodic time series measured from biological systems." *BMC Bioinformatics*, 6, 6:117.

Alagon, J. (1989): "Spectral discrimination for two groups of time series." *Journal of Time Series*, 10, 203–214.

Andrzejak, R. G., K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger (2001): "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state." *Physical review. E, Statistical, nonlinear, and soft matter physics*, 64.

Ben-Dor, A., L. Bruhn, N. Friedman, I. Nachman, M. Schummer, and Z. Yakhini (2000): "Tissue classification with gene expression profiles." *Journal of Computational Biology*, 7, 559–583.

Bhansali, R. J. (1970): "A mixed spectrum analysis of the lynx data." *Journal of the Royal Statistical Society*, 199–209.

Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen (1984): *Classification and Regression Trees*, Chapman & Hall/CRC.

Brillinger, D. R. (1981): *Time Series: Data Analysis and Theory (Classics in Applied Mathematics, 36)*, Holden Day, San Francisco.

Bühlman, P. (2006): "Boosting for high-dimensional linear models." *Annals of Statistics*, 559–583.

Chinipardaz, R. and T. Cox (2004): "Nonparametric discrimination of time series." *Metrika*, 13–20.

Dettling, M. and P. Buhlmann (2003): "Boosting for tumor classification with gene expression data," *Bioinformatics*, 19, 1061–1069.

Diggle, P. and I. Al-Wasel (1993): "On periodogram-based spectral estimation for replicated time series," in *Subba Rao, ed., Developments in Time Series Analysis.*, Chapman and Hall, 341–354.

Diggle, P. and I. Al-Wasel (1997): "Spectral analysis of replicated biomedical time series," *Applied Statistics*, 31–71.

Dudoit, S., J. Fridlyand, and T. P. Speed (2002): "Comparison of discrimination methods for the classification of tumors using gene expression data," *Journal of the American Statistical Association*, 97, 77–87.

Freund, Y. and R. Schapire (1996): "Experiments with a new boosting algorithm," in *Proceedings of the Thirteenth International Conference on Machine Learning (ICML)*, 148–156.

Friedman, J., T. Hastie, and R. Tibshirani (2000): "Additive logistic regression: a statistical view of boosting," *Annals of Statistics*, 28, 337–407.

Hastie, T., R. Tibshirani, and J. H. Friedman (2001): *The Elements of Statistical Learning*, New York: Springer.

Hernández-Flores, C., J. Artiles-Romero, and P. Saavedra-Santana (1999): "Estimation of the population spectrum with replicated time series." *Computational Statistics and Data Analysis*, 271–280.

DOI: 10.2202/1557-4679.1288

Kakizawa, Y., R. Shumway, and M. Taniguchi (1998): "Discrimination and clustering for multivariate time series." *Journal of the American Statistical Association*, 328–340.

Kooperberg, C., C. Stone, and Y. Truong (1995): "Logspline estimation of a possibly mixed spectral distribution." *Journal of Time Series Analysis*, 259–388.

Luengo, I., C. Hernández, and P. Saavedra (2006): "Test to compare two population logspectra," *Computational Statistics*, 91–101.

Pander, T. (2008): *Information Technologies in Biomedicine, Advances in Soft Computing*, Springer-Verlag, volume 47, chapter An Application of Robust Kernel-Based Filtering of Biomedical Signals, 259–266.

Pardey, J., S. Roberts, and L. Tarassenko (1996): "A review of parametric modelling techniques for eeg analysis." *Med Eng Phys*, 18, 2–11.

R Development Core Team (2010): *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, URL `http://www.R-project.org`, ISBN 3-900051-07-0.

Ripley, B. D. (1996): *Pattern Recognition and Neural Networks*, Cambridge University Press.

Saavedra, P., C. Hernández, and J. Artiles (2000): "Spectral analysis with replicated time series." *Communications in Statistics: theory and methods*, 2343–2362.

Saavedra, P., C. Hernández, I. Luengo, J. Artiles, and A. Santana (2008): "Estimation of the population spectrum for linear processes with random coefficients." *Computational Statistics*, 79–98.

Schalkoff, R. (1997): *Artificial neural networks*, McGraw-Hill, New York.

Shumway, R. and A. Unger (1974): "Linear discriminant functions for stationary time series," *Journal of the American Statistical Association*, 948–956.

Vilar, J. and S. Pértega (2004): "Linear discriminant functions for stationary time series.: local linear fitting approach," *Journal of Nonparametric Statistics*, 443–462.