TIME SERIES SPECTRAL DISCRIMINANTS BASED ON KULLBACK INFORMATION MEASURES (WITH AN APPLICATION TO SEISMOLOGY)

Kristin Hedberg, University of Kansas Medical Center; Joseph E. Cavanaugh, University of Missouri-Columbia
Kristin Hedberg, Dept. of Preventive Medicine, KUMC, 3901 Rainbow Blvd., Kansas City, KS 66160

Key Words: Directed divergence, discriminant analysis, Kullback-Leibler discrepancy, simulation

I. Introduction

Discriminant analysis is a well-known and frequently used technique in multivariate statistics, which involves classifying an observed vector of unknown origin into one of several defined populations. Discriminant techniques are also valuable in time series analysis, since many real problems occur in which an observed series must be classified as belonging to one of a collection of known sources. Time series discrimination has been applied to several areas, including medicine, engineering and geophysics. There are a number of different discriminant measures with which we may distinguish between two or more groups, based on both time and frequency domain methodologies.

Gersch (1981) classifies the anesthesia levels of pre-surgery patients, to determine whether or not patients are sufficiently anesthetized for surgery, by analyzing their electroencephalogram (EEG) readings. He uses a nearest neighbor rule, based on the Kullback-Leibler information measure, which compares the test series with each series of known origin. This method classifies the test series into the group containing the most similar known series.

Another medical application is presented in Alagón (1989), in which the EEG is used to determine whether or not a patient has Multiple Sclerosis. He bases his conclusions on two methods: an F-test of the average spectrum of the first population to the average spectrum of the second population, and a "windowed discriminant function," based on the quadratic discriminant function and the periodogram.

One of the most frequently published applications for time series discriminants involves classifying a seismogram reading as arising from either an earthquake or an underground explosion. Shumway and Unger (1974) derive spectral forms of the Kullback-Leibler discrimination information, to distinguish between earthquakes and underground explosions. Shumway (1982) discusses both time and frequency domain methods for discriminant analysis. He suggests that spectral analysis is more meaningful than time domain methods. An analysis of power, involving an F-statistic based on the ratio of between-group power to within-group power, is used to discriminate between earthquakes and explosions.

Cavanaugh, McQuarrie and Shumway (1993) base discriminants upon the entire seismogram using several different methods including logarithms of P and S phase amplitudes and mean-squared errors, third order autoregressive model coefficients, the Kullback-Leibler discrimination information, and the α-entropy. The last three are determined to be the best methods for their purposes. The Kullback-Leibler discrimination information, along with the Chernoff information measure, is employed, again to classify earthquakes and explosions, in Kakizawa, Shumway and Taniguchi (1998). The Chernoff measure is found to have the highest rate of correct classification among the discriminants used.

In this paper we will focus on discriminants based on Kullback's discrimination information, modified for spectral analysis. Section II presents relevant background material pertaining to time series analysis, followed by the development of our discriminants in Section III. Section IV features a discussion of the procedure and results from a simulation, and Section V presents an application of our discriminants in the setting of seismology.

II. Time Series Background

Consider a set of random variables measured over time, x_t, $t = 0, 1, \dots, T\text{-}1$. Assume the series has mean value function μ_t, and autocovariance function

$$C(s,t) = E[(x_s - \mu_s)(x_t - \mu_t)],$$

which is analogous to the usual covariance function. If x_t has a constant mean value function and the autocovariance is a function of the difference, or lag, between s and t, then x_t is called a weakly stationary process. In this case, we may write

$$C(s,t) = R(s-t) = R(m) = E[(x_{t+m} - \mu)(x_t - \mu)],$$

where $m = s - t$.

Spectral analysis consists of the study of salient frequencies in a time series, where the frequency v is the rate of oscillation for a recurring cyclic pattern in the series. To find the predominant frequencies in the series, we will consider the discrete frequencies $v_k = k/T$, $k = 0, 1, \dots, T/2$; i.e., values of v ranging from 0 to ½. The period, $1/v_k$, is the time required to complete one cycle. Since time series have cyclic tendencies, we would like to match our weakly stationary series with a known series, so we may find which frequencies contribute most to the series. We will assume our series x_t has a mean of zero. Matching x_t with sine or cosine waves of varying frequencies gives us the sine transform,

$$X_S(k) = T^{-1/2} \sum_{t=0}^{T-1} x_t \sin(2\pi v_k t) ,$$

and the cosine transform,

$$X_C(k) = T^{-1/2} \sum_{t=0}^{T-1} x_t \cos(2\pi v_k t) ,$$

where k is the index for the frequency, v_k. We then may combine the cosine and sine transforms to construct the periodogram,

$$P_x(v_k) = [X_C^2(k) + X_S^2(k)] .$$

If the series is effectively matched with either the sine wave or the cosine wave at frequency v_k, then $P_x(v_k)$ will be large. We may consider the periodogram to be a measure of the variability in the series; if $P_x(v_k)$ is large for v_k, then much of the variability in the series is due to periodic behavior at that particular frequency.

The periodogram may be written in terms of the discrete Fourier transform (DFT) of x_t, defined as

$$\begin{aligned} X(k) &= T^{-1/2} \sum_{t=0}^{T-1} x_t \exp\{-2\pi i v_k t\} \\ &= T^{-1/2} \sum_{t=0}^{T-1} x_t [\cos(2\pi v_k t) - i \sin(2\pi v_k t)] \\ &= X_C(k) - iX_S(k) . \end{aligned}$$

Note that we have

$$|X(k)|^2 = X_C^2(k) + X_S^2(k) = P_x(v_k) .$$

The power spectrum can be considered the time series analogue of analysis of variance. It is a nonnegative function whose graph displays the variance plotted as a function of the frequency; i.e., the plot is a visual descriptor of how much of the variability corresponds to various frequencies in the series. A prominent peak at frequency v indicates that the series is dominated by periodic behavior at that particular frequency. The power spectrum is given by

$$f_x(v) = \sum_{m=-\infty}^{\infty} R_x(m) \exp\{-2\pi i v m\} ,$$

where $f_x(v)$ satisfies

$$R_x(m) = \int_{-1/2}^{1/2} f_x(v) \exp\{2\pi i v m\} dv .$$

Since $f_x(v)$ can only be computed exactly when the true autocovariance function is known, the power spectrum must be estimated. The periodogram is an approximately unbiased estimator for $f_x(v)$, but $P_x(v_k)$ is highly variable and therefore, not very informative. To attenuate this variability, we consider averaging $P_x(v_k)$ over a frequency band of width L (where L is odd), with $(L-1)/2$ discretized frequency values on each side of $P_x(v_k)$. This local averaging produces the smoothed spectral estimator,

$$\begin{aligned} \hat{f}_x(v_k) &= \frac{1}{L} \sum_{l=-(L-1)/2}^{(L-1)/2} P_x(v_{k+l}) \\ &= \frac{1}{L} \sum_{l=-(L-1)/2}^{(L-1)/2} \left[X_C^2(k+l) + X_S^2(k+l) \right]. \end{aligned}$$

This smoothed spectral estimator generates the plot of variance versus frequency, which we will call the spectrum in our application.

III. Development of Discriminants

Suppose we have a time series, x_t, t = 0, 1, . . . , $T-1$, which we know belongs to one of two known populations, say earthquakes or explosions. If the two populations produce similar time series and we have no way of identifying the source, we need a method to match the series to one of the two populations, and thus, to determine from which population the series in question came.

Assuming that the two populations are similar in all characteristics excepting their spectra, e.g., both populations have approximately equal means and variances, we would like to use a discriminant based on the spectra. The Kullback-Leibler discrepancy, or the non-normalized Kullback-Leibler information, is a popular method of discrimination between two densities, used in a wide range of statistical applications. The discrimination information, or directed divergence, for two normal processes that differ only in their spectra, is approximately

$$I(f_1, f_2) = \frac{1}{2} \sum_{k=0}^{T-1} \left\{ \frac{f_1(v_k)}{f_2(v_k)} - \log \frac{f_1(v_k)}{f_2(v_k)} - 1 \right\}$$

(Shumway & Unger, 1974). If we consider $f_1(v)$ to be the spectrum representing population 1, and $f_2(v)$ to be the spectrum representing population 2, then the discrimination information is simply a measure of discrepancy between population 1 and population 2, with respect to population 1. It follows that we might modify the discrimination information to assess the conformity of an individual series to the characteristics of a population.

We will use three different forms of Kullback's information measures in this paper: two forms of the directed divergence and a combination of those two forms, called the symmetric divergence. For the following, let $f_{TEST}(v)$ be the spectrum for the series we wish to classify, let $f_j(v)$ be the average spectrum for the comparison population, where j = 1 is the population known to be earthquakes and j = 2 is the population known to be explosions. The first directed divergence, the I-divergence between the test spectrum and spectrum j with respect to the test spectrum is given by

$$I_1(f_{TEST}, f_j) = \frac{1}{2}\sum_{k=0}^{T-1}\left\{\frac{f_{TEST}(v_k)}{f_j(v_k)} - \ln\frac{f_{TEST}(v_k)}{f_j(v_k)} - 1\right\}.$$

The I-divergence between the test spectrum and spectrum j with respect to spectrum j is given by

$$I_2(f_{TEST}, f_j) = \frac{1}{2}\sum_{k=0}^{T-1}\left\{\frac{f_j(v_k)}{f_{TEST}(v_k)} - \ln\frac{f_j(v_k)}{f_{TEST}(v_k)} - 1\right\}.$$

Finally, the symmetric divergence, or J-divergence, is given by

$$\begin{aligned} J(f_{TEST}, f_j) &= I_1(f_{TEST}, f_j) + I_2(f_{TEST}, f_j) \\ &= \frac{1}{2}\sum_{k=0}^{T-1}\left\{\frac{f_{TEST}(v_k)}{f_j(v_k)} + \frac{f_j(v_k)}{f_{TEST}(v_k)} - 2\right\}. \end{aligned}$$

The first I-divergence, I_1, will classify the test series as an earthquake if $I_1(f_{TEST}, f_2) - I_1(f_{TEST}, f_1) \geq 0$, and as an explosion otherwise. The I_2- and the J-divergence are used in the same manner.

IV. Simulations

To test the proposed spectral discriminants, we will consider a simple setting in which the time series are generated via the first-order autoregressive (AR) model, given by

$$x_t = \phi x_{t-1} + w_t,\ w_t \sim iid(0, \sigma_w^2).$$

We generate eight series of the form

$$x_t = 0.9 x_{t-1} + w_t,\ t = 0, 1, \ldots, 299,$$

which will serve as the baseline population (see Figures 1-2). An AR process with $\phi = 0.9$ is known to be dominated by low frequencies. We then generate sets of eight series from several other populations with different values of ϕ, to determine when the three discriminants work well, when they begin to make classification errors, and when they fail to provide adequate classification.

The first test of the discriminants will be a comparison of the baseline population with its opposite,

$$x_t = -0.9 x_{t-1} + w_t,$$

known to have prominent high frequencies. These two sets of series are so dramatically different that any observer could distinguish the series just by inspecting their spectra, or even the actual series (see Figures 3-4). We test the discriminants using a hold-out method, in which we consider one series to be a test series that we wish to classify. We form our average spectrum for each population as the average of the remaining spectra, not including the test series. We will cycle through all sixteen series, until each has acted as the test series. As would be expected, all three discriminants work perfectly, even with a bandwidth as small as L = 3.

We generate several more sets of series corresponding to different values of ϕ, gradually increasing ϕ by increments of 0.1, until errors are encountered in the discrimination. With a bandwidth of L = 5, all three measures demonstrate perfect discrimination until ϕ > 0.65; with L = 11, the procedures work as high as ϕ = 0.72 (see Figures 5-6). None of the three measures appear to be consistently better than the others, and the performances of the measures do not vary noticeably with different bandwidths.

V. Application

The data used in this application are seismogram readings from eight earthquakes and nine underground explosions from stations in Scandinavia (Cavanaugh, McQuarrie and Shumway, 1993). Both of these types of seismograms have two distinct phases, known as the P phase and the S phase, representing different kinds of movements in and below the earth's surface. The P phase is the first part of the seismogram reading, and is the section of the seismogram that we will consider in this paper. The P phase is caused by surface waves that travel about 1.7 times as fast as the perpendicular S waves, which produce the S phase.

Defense agencies are interested in identifying the source of a seismogram reading to determine if another country is doing underground nuclear testing. We will analyze our collection of seventeen readings and try to distinguish between them based on their prominent frequencies.

After an examination of our seventeen time series, we find that the magnitude of the oscillations vary considerably from one series to another. Since our discriminants depend on approximately constant variance, both within each series and between series, we will perform the actual analysis on a standardized version of the series. The spectrum for each series is calculated, and then the series are classified, following the same procedure as outlined in the simulations. For each of the three discrimination measures, $f_1(v)$ will represent the average spectrum for the earthquakes, and $f_2(v)$ will represent the average spectrum for the explosions. We will determine the test series to be an earthquake if the measure is nonnegative, and an explosion otherwise. We should note that, according to the average spectrum for each group (see Figure 7), earthquake P phases lean toward lower frequencies than do explosion P phases. This difference in average spectra should allow us to discriminate between the spectra; our discriminants will conclude that the test series belongs to the group whose average spectrum is the most like the test spectrum.

If we look at the average spectrum for each group, we see that the values of each spectrum are practically the same beyond a frequency of about 0.22. Thus, we may consider analyzing the spectra for only the lower frequencies, where the values are distinctly different for the two groups. We also will, to an extent, see better classification with larger bandwidths; we find that L = 37 (18 values on either side of any given frequency for the periodogram) gives the best results. With these modifications we see 76.5% correct classifications with the I_1-divergence, and 82.4% correct classifications with I_2- and J-divergence. The complete results from SAS are presented in Table 1.

Table 1. SAS Results for Application

Series	I_1	I_2	J
Earthquake #1	-4.96563	-16.1548	-21.1204
Earthquake #2	0.37463	0.070652	0.44528
Earthquake #3	5.33630	2.58330	7.91961
Earthquake #4	2.06496	1.62068	3.68564
Earthquake #5	13.3553	5.71742	19.0727
Earthquake #6	3.58771	2.92491	6.51262
Earthquake #7	1.46424	0.56786	2.03210
Earthquake #8	-2.93081	-4.08207	-7.01288
Explosion #1	-0.87904	-1.10895	-1.98799
Explosion #2	0.80675	-5.07160	-4.26485
Explosion #3	-3.64376	-6.83440	-10.4782
Explosion #4	-4.79244	-23.4998	-28.2922
Explosion #5	-6.22543	-46.2105	-52.4359
Explosion #6	-4.37301	-36.3892	-40.7622
Explosion #7	-3.33701	-5.00275	-8.33976
Explosion #8	18.3449	7.49093	25.8358
Explosion #9	-1.63082	-3.62294	-5.25376

We should examine the series that were misclassified, to determine if a logical explanation for the errors exists. All three discriminants misclassified Earthquake #1 and #8, and Explosion #8. The I_1-divergence misclassified Explosion #2, as well. The spectrum for Earthquake #1 (see Figure 8) indicates the presence of higher frequencies than that which are typical for earthquakes, so the discriminants classified that series as an explosion. The spectrum for Earthquake #8 (see Figure 9) almost exactly matches the average explosion spectrum, so it is natural that is would be classified as an explosion. The spectrum for Explosion #8 (see Figure 10) exhibits much lower frequencies than would be expected in an explosion, which explains why it is misclassified as an earthquake. Finally, Explosion #2 was misclassified by the I_1-divergence, but was correctly classified by the other two methods (see Figure 11); this may be an indication that the J-divergence should be the preferred discriminant. Since the J-divergence is the sum of the two I-divergences, it will agree with the I-divergence with the greater magnitude, and may then be slightly more sensitive.

VI. Conclusion

In this paper we have explored three forms of the Kullback-Leibler discrepancy in the context of both simulations and actual data. The three discriminants performed nearly equally well, with a possible advantage toward the use of the J-divergence. This should not be a surprise, since all three discriminants have very similar foundations. We did find that the discriminants performed better in the setting of simulations than with earthquakes and explosions, but the misclassifications in the application are easily explained and should be expected, due to random chance.

References

Alagón, J. (1989). Spectral discrimination for two groups of time series. *Journal of Time Series Analysis*, **10**, 203-214.

Cavanaugh, J.E., A.D.R. McQuarrie & R.H. Shumway (1993). Parametric and nonparametric discriminants for regional earthquakes and explosions. Scientific report PL-TR-93-2164, Phillips Laboratory, Directorate of Geophysics, Air Force Materiel Command, Hanscom AFB, MA 01731-3010.

Gersch, W. (1981). Nearest neighbor rule classification of stationary and nonstationary time series. In *Applied Time Series Analysis II*, ed. David Findley, 328-340. New York: Academic Press.

Kakizawa, Y., Shumway, R.H. & Taniguchi, M. (1998). Discrimination and Clustering for Multivariate Time Series. *Journal of the American Statistical Association*, **93**, 328-340.

Shumway, R.H. (1988). *Applied Statistical Time Series Analysis*. Englewood Cliffs: Prentice-Hall.

Shumway, R.H. (1982). Discriminant analysis for time series. In *Handbook of Statistics, Vol. 2, Pattern Recognition and Reduction of Dimensionality*, ed. P.R. Krishnaiah, 1-43. Amsterdam: North-Holland.

Shumway, R.H. & Unger A.N. (1974). Linear discriminant functions for stationary time series. *Journal of the American Statistical Association*, **69**, 948-956.