A Sequential Comparison of Two Population Spectra

Quintana, M. P.[1]   Saavedra, P.[2]

Abstract

Two populations of objects are considered in such way that on each one a random signal can be evaluated. This signal generally is a realization of a stationary process. In each population, the signal is generated by a same pattern. The aim of this paper consists of comparing both patterns. A sequential statistical test is proposed and its distribution of propability is investigated under the null and alternative hypothesis.

Introduction

Let us consider two populations of objects on which a certain signal can be evaluated and this one can be modeled by a stationary process. We will suppose that in each population a same pattern which generates such signals exists.

The aim of this paper is to compare two different patterns corresponding to two populations in order to determine if the considered signal has predictive value.

In practical applications, one of such populations can be constituted by subjects with a certain pathology (cases) and the other one, by people without that pathology (controls). The signal can consist of the realization of electroencephalogram in a certain cerebral zone. Then it analyzes its discriminant character. With that aim we propose a sequential statistical test based on the spectrum of the signals.

The Statistical Test

Let $\{X_{li}(t);\ l=1,2;\ i=1,\cdots,r_l;\ t=1,\cdots,N\}$ be a set of time series evaluated on random samples of r_l objects, chosen from population C_l with l = 1,2, in the same N times. Each of such time series can represent in practical applications the evaluation of a certain signal or a set of replicated measures over the subjects of the populations aforementioned. The periodogram of each time serie, for which the jth Fourier frequency, is defined by:

$$I_{li}(\omega_j)=\frac{1}{2\pi N}\left|\sum_{t=1}^{N}X_{li}(t)e^{-i\omega_j t}\right|^2 \tag{1}$$

Let us suppose that each periodogram satisfies the relation $I_{li}(\omega_j)=f_l(\omega_j)\cdot U_{lij}^N$, being $f_l(\omega)$ the spectral density corresponding to the lth population and $\{U_{lij}^N\}$, for each l = 1,2 and $i=1,\cdots,r_l$, are independent random variables in $j=1,\cdots,\nu=[N/2]$ with exponential distribution of parameter 1. Such a model is based on known results for the periodogram of Gaussian linear processes [4]. To test $H_0: f_1(\omega) \equiv f_2(\omega)$, we propose a *sequential test*, as follows:

Let $Y_{lij} = \mu_l(\omega_j) + \xi_{lij} - C$, being $Y_{lij} = \log I_{li}(\omega_j) - C$, $\mu_l(\omega) = \log f_l(\omega)$, C the Euler constant and $\xi_{lij} = \log U_{lij}$.

i. Let $J_2 = \frac{1}{v}\sum_{j=1}^{v}\left(\overline{Y}_{1.j} - \overline{Y}_{2.j}\right)^2$ be the statistical test, with $\overline{Y}_{l\bullet j} = (1/v)\sum_{i=1}^{r_l} Y_{lij}$.

ii. Let define the sequential test as follows: Choose two numbers $a, b > 0$, then accept H_0 if $J_2 < a$, take m new time series of each group if $a \le J_2 \le b$, and reject H_0 if $J_2 > b$.

Distribution of the sequential statistical test under the null and alternative hypothesis.

Under the null hypothesis, H_0, the statistical test can be expressed by:

$$J_2 = \frac{1}{v}\sum_{j=1}^{v}\left(\overline{\xi}_{1.j} - \overline{\xi}_{2.j}\right)^2 \qquad (2)$$

where $\overline{\xi}_{l\bullet j} = (1/r_l)\sum_i \xi_{jij}$. By means of elementary numerical methods, we obtain that $E[\xi_{lij}] = C = -0.57721$ (*C* is called the Euler constant) and $\mathrm{var}(\xi_{lij}) = \sigma^2 = 1.648124$. Obviusly, under H_0, $E[J_2] = 0$. For larger values of r_1 and r_2, we can consider the aproximation $\overline{\xi}_{1\bullet j} - \overline{\xi}_{2\bullet j} \approx N(0, \theta)$, being $\theta = \sigma\sqrt{(1/r_1) + (1/r_2)}$. Given that ξ_{lij} are independent random variables for *j*, then it follows that $\sum_j \left(\overline{\xi}_{1\bullet j} - \overline{\xi}_{2\bullet j}\right)^2 / \theta^2 \approx \chi^2(v)$, where this last notation represents the chi-square distribution with v degrees of freedom. Let $J_2 = \frac{\theta^2 \cdot Z}{v}$, being $Z \cong \chi^2(v)$.

For an specificied difference $H_1: d(\omega) = \mu_1(\omega) - \mu_2(\omega)$, we can obtain in analogous way an approximate distribution for J_2 under H_1. In fact,

$$\overline{Y}_{1\bullet j} - \overline{Y}_{2\bullet j} = d(\omega_j) + \overline{\xi}_{1\bullet j} - \overline{\xi}_{2\bullet j} \approx N\left(d(\omega_j), \theta\right) \qquad (3)$$

so,

$$\frac{1}{\theta^2}\sum_{j=1}^{v}\left(\overline{Y}_{1\bullet j} - \overline{Y}_{2\bullet j} - d(\omega_j)\right)^2 \approx \chi^2(v) \qquad (4)$$

From this formula, we can find:

$$J_2 = \frac{\theta^2 \cdot Z}{v} - \frac{1}{v}\sum_{j=1}^{v} d(\omega_j)^2 + \frac{2}{v}\sum_{j=1}^{v}\left(\overline{Y}_{1\bullet j} - \overline{Y}_{2\bullet j}\right) \cdot d(\omega_j) \qquad (5)$$

being $Z \cong \chi^2(v)$. But, since $\overline{Y}_{1\bullet j} - \overline{Y}_{2\bullet j}$ is an unbiased estimator of $d(\omega_j)$, it results in the approximation:

$$J_2 \approx \frac{\theta^2 \cdot Z}{v} + \frac{1}{v}\sum_{j=1}^{v} d(\omega_j)^2 \qquad (6)$$

Fixed α and β, the Type I and Type II errors respectively, we can determine the critical values a and b solving the ecuations $\alpha = P(J_2 > b \geq |H_0)$ and $\beta = P(J_2 < a|d(\omega))$. In fact,

$$\alpha = P(J_2 > b \geq |H_0) = P\left(\sum_j (\bar{\xi}_{1\bullet j} - \bar{\xi}_{2\bullet j})^2 / \theta^2 > b \cdot v / \theta^2\right) \quad (7)$$

so we can obtain $b = \theta^2 \cdot \chi^2_\alpha(v)/v$, being $\chi^2_\alpha(v)$ the quantil 1-α of the distribution $\chi^2(v)$. Analogously, we can obtain:

$$a = \frac{1}{v}\sum_{j=1}^{v} d(\omega_j)^2 + \frac{\theta^2 \cdot \chi^2_{1-\beta}(v)}{v} \quad (8)$$

Estimation of the parameter *d*

The natural estimation of the parameter $d(\omega)$ is obtained by estimating each $\mu_l(\omega)$ parameter, for l = 1, 2. Such estimation can be carried out, efficiently, from the model $Y_{lij} = \mu(\omega_j) + e_{lij}$, where $e_{lij} = \xi_{lij} - C = \log U_{lij} - C$ and considering that U_{lij}, l=1,2; i = 1, ..., r_l; $j = 1, \ldots, v$ are independent e identically distributed random variables with $\exp(1)$. It is easy to see that $E[e_{lij}] = 0$. By using the principle of the local likelihood of [2], we consider for each $\omega \in [-\pi, \pi]$ the local likelihood:

$$\sum_{i=1}^{r_l}\sum_{j=1}^{v}\left\{Y_{lij} - \gamma_{l0} - \gamma_{l1}(\omega_j - \omega) - \exp(Y_{lij} - \gamma_{l0} - \gamma_{l1}(\omega_j - \omega))\right\}\cdot\frac{1}{h}K\left(\frac{\omega_j - \omega}{h}\right) \quad (9)$$

where $K(z)$ is a kernel function and h the corresponding bandwidth. Notice that the parameter $\mu_l(\omega_j)$ has been aproximated linearly on a neighbourhood of ω. Finally, we consider the estimate $\hat{d}(\omega) = \hat{\mu}_1(\omega) - \hat{\mu}_2(\omega)$.

Simulation

We have considered, for the lth population, that the signals are generated by the stationary moving average process:

$$X_l(t) = b_{l0} \cdot \varepsilon_l(t) + b_{l1} \cdot \varepsilon_l(t-1) + b_{l2} \cdot \varepsilon_l(t)\,; l = 1,2 \quad (10)$$

being $\{\varepsilon_l(t)\}$ i.i.d. random variables $N(0, \sigma_\varepsilon(l))$. Its spectral density function is:

$$f_l(\omega) = \frac{\sigma_\varepsilon(l)^2}{2\pi}\left|b_{l0} + b_{l1} \cdot \exp(i\omega) + b_{l2} \cdot \exp(2i\omega)\right|^2 \quad (11)$$

Figure 1 shows the parameter $\mu_l(\omega) = \log f_l(\omega)$, for l=1, $\sigma_\varepsilon(1) = 1$, $b_{10} = 3$, $b_{11} = 2$ and $b_{12} = 1$. It is shown jointly with its local polynomial estimator of degree 1, for N =120, and $r_1 = r_2 = 50$.

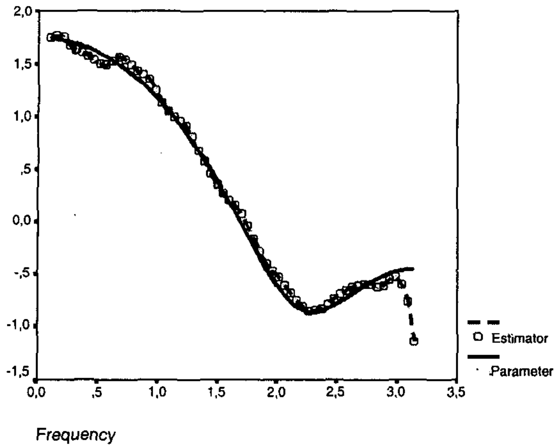

Fig. 1

Now let us suppose that in population 2, the stationary moving average process has the same coeficients than that in population 1, but being $\sigma_{\varepsilon}(2) = 1.1$. Figure 2 show us the log-spectrals of both populations. For the values N, r_1 and r_2, specified above, the test detects diferences, in the initial step, 88 times from 100 simulations.

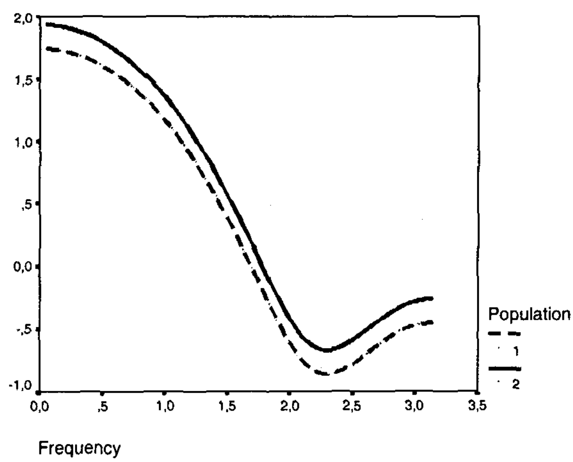

Fig. 2

References.

[1] P.J. Diggle and I. Al-Wasel, "On Periodogram-Based Spectral Estimation for Replicated Time Series". Subba Rao (Ed), *Developments in Time Series Analysis.* (Chapman and Hall, Great Britain), (1993), 341-354.

[2] J. Fan and I. Gijbels, *Local Polynomial Modelling and its Aplications*, Chapman and Hall, 1996.
[3] C.N. Hernández-Flores, J. Artiles-Romero and P. Saavedra-Santana, "Estimation of the Population Spectrum with Replicated Time Series". *Comp. Stat. and Data Anal.* **30**, (1999), 271-280.
[4] M. B. Priestley, *Spectral Analysis and Time Series*, Wiley, New York, 1981.
[5] P. Saavedra, C.N. Hernández and J. Artiles, "Spectral Analysis with Replicated Time Series". *Comm. Stat. Theory and Methods.* **29**,(2000), *Issue 10.*
[6] J. Whitehead, *The Desing and Analysis of sequential Clinical Trials*,(Revised 2nd edition) Chichester:Wiley, 1997.

1 Departament of Mathematics. University of Las Palmas de Gran Canaria. Campus Universitario de Tafira. 35017 Las Palmas de Gran Canaria. Spain

2 Departament of Mathematics. University of Las Palmas de Gran Canaria. Campus Universitario de Tafira. 35017 Las Palmas de Gran Canaria. Spain