Rev.Acad.Canar.Cienc., VIII (Núm. 1), 19-29 (1996)

BOOTSTRAP ESTIMATION OF POPULATION SPECTRUM WITH REPLICATED TIME SERIES.

C. N. Hernández Flores, J. Artiles Romero and P. Saavedra Santana

Department of Mathematics.

University of Las Palmas de Gran Canaria.

35017 Las Palmas de Gran Canaria, Canary Islands

e-mail: cflores@dma.ulpgc.es

Abstract

In this work a set with r objets is considered and a stationary process $X_i(t)$ with spectral distribution absolutely continuous is observed for each of them at same time interval. Each spectral density function $f_i(\omega)$ may be considered it a realization of a stationary process $R(\omega)$.The spectrum population $f(\omega)$ is defined by $E[R(\omega)]$.We estimate $f(\omega)$ by means of a bootstrap method and we proof the asymptotic validity when the number of objects r tend to infinity.

Keywords Spectrum Estimator, Bootstrap, Replicated Time Series, Population Spectrum, Periodogram .

1. Introduction.

Although spectral analysis is a very highly developed methodology, almost all of this development has been in the context of a single, long time series. This perhaps the fact reflects that the origins of the subject were signal processing and the physical sciences. However, the usefulness of time series methodology is becoming more widely accepted in biomedical sciences, where replicated experiments are the rule rather than the exception. Diggle and Al-Wasel (1993) studied replicated time series of measurements of the concentration of luteinizing hormone in blood samples. First, they considered each time serie as the realization of a stationary process with spectral density function absolutely continuous $f(\omega)$. Nevertheless, the strong variability between subjects lead to inconsistent data with this model and they proposed the alternative model $I_i(\omega_j) = f_i(\omega_j) \cdot U_{ij}$ i=1,...,r; j=1,..., [$N/2$] , being $I_i(\omega_j)$ the periodogram at the frequency ω_j for the ith individual, $f_i(\omega_j)$ the subject specific spectrum and U_{ij} mutually independent, unit mean exponential variates with common pdf e^{-u}, $u \geq 0$. Suppose now that r units are selected at random from a given population. Then we regard the $f_i(\omega_j)$ as independent realizations of a random function $R(\omega)$, and set $f(\omega)=E[R(\omega)]$, where the expectation is defined with respect to the population of subjects and $f(\omega)$ is called the population spectrum. They estimated $f(\omega)$ supposing certain parametrizations for the processes involved in the model and they obtained the average periodogram as a maximum likehood estimator of the population spectrum. Obviously, this estimator is unbiased and consistent for the number of objects.

We have considered a more generalised model than the one developped by Diggle and Al-Wasel without making any parametrizations and we haved used a bootstrap method for the estimation of the population spectrum. Moreover, we have compared the confidence intervals obtained by means of the bootstrap method with those obtained using the normal approximation when r is large and the central limit theorem is taken into account. Efron and Tibshirani (1986) used bootstrap for estimation of a parametric time series model. Franke and Härdle (1992) worked with bootstrap too for estimating the spectral density function when there is only one realization of a stationary process. They proved the theoretical asymptotic validity of the bootstrap principle according to Bickel and Freedman (1981), Freedman (1981) and Romano (1988). We have also proved the validity of the bootstrap principle for a fixed number of observations by object and an increased number of objects. The simulations we present illustrate that our procedure works for moderate sample sizes by object and a large number of objects.

2. The model.

Let $\{X_{it}: t=1,\ldots,N\}$ be r time series, each one with spectral density function $f_i(\omega_j)$. According to Diggle and Al-Wasel we consider the above mentioned model $I_i(\omega_j)=f_i(\omega_j)U_{ij}, i=1,\ldots,r; j=1,\ldots,[N/2]$, being $I_i(\omega_j)$ the periodogram at the frequency ω_j for the ith subject, $f_i(\omega_j)$ the specific spectrum and U_{ij} mutually independent, unit mean variates; we omit the hypothesis of Diggle and Al.-Wasel that the U_{i_j} are exponential variates. We also assume that $f_i(\omega)=f(\omega)Z_i(\omega)$ where $\{Z_i(\omega)\}$: $i=1,\ldots,r$ are independent copies of a stochastic process $\{Z(\omega_j)\}$ with $E[Z(\omega_j)]=1$, for all ω_j, where $f(\omega)$ is called the population spectrum. So, $I_i(\omega_j)=f(\omega_j)W_{ij}$, being $W_{ij}=Z_i(\omega_j)U_{ij}$, random variables i.i.d. for each fixed j with distribution function F_j, unit mean and finite variance $\sigma_j^2 \le \sigma^2$.

We propose as an estimator of the population spectrum $f(\omega)$ the average periodogram

$$\hat{f}(\omega)=\frac{1}{r}\sum_{i=1}^{r} I_i(\omega)=\bar{I}(\omega)$$

The following procedure gives a bootstrap aproximation for, $\hat{f}(\omega)$

Step 1. The variables W_{ij} are estimated as:

$$\hat{W}_{ij}=\frac{I_i(\omega_j)}{\bar{I}(\omega_j)} \quad , \ i=1,\ldots,r; \ j=1,\ldots,[N/2]$$

Obviously, for each frequency ω_j , $\sum_i \hat{W}_{ij}/r=1$.

Step 2. B bootstrap samples $\{\hat{W}^*_{1j},\ldots,\hat{W}^*_{rj}\}$ are drawn from $\{\hat{W}_{1j},\ldots,\hat{W}_{rj}\}$. For each frequency ω_j, we consider the bootstrap periodogram computed as follows:

$$\hat{I}_i^*(\omega_j)=\hat{I}_i^*(-\omega_j)=\hat{f}(\omega_j)\hat{W}_{ij}^*$$

Step 3. Finally, the bootstrap estimation of the spectrum population is computed as follows:

$$\hat{f}^{*}(\omega_j)=\frac{1}{r}\sum_{i=1}^{r}\hat{I}_i^{*}(\omega_j)$$

The $\hat{W}_{ij}^{*}$ obtained through a boostrap resampling from $\{\hat{W}_{1j},\ldots,\hat{W}_{rj}\}$ have unit mean with regard to the empirical distribution.

Obviously the average periodogram is unbiased and a consistent estimator for $f(\omega_j)$ when r increase to infinity. Within the bootstrap context, we obtain $E^{*}[\hat{f}^{*}(\omega_j)]=\frac{1}{r}\sum_{i=1}^{r}E[I_i^{*}(\omega_j)]=\hat{f}(\omega_j)$; therefore, the bootstrap estimator is unbiased for $\hat{f}(\omega_j)$. According to Härdle and Bowman (1988) it is not necessary to correct the pivotal quantity.

3.- The bootstrap principle holds.

The basic idea of bootstrapping, as applied to the population spectrum estimation context, is to infer properties of the distribution of the estimator $\hat{f}(\omega)$ from the conditional distribution of its bootstrap approximations $\hat{f}^{*}(\omega)$, given the original data. To prove the theorical validity of this bootstrap principle, we follow Bickel and Freedman (1981) and consider the Mallows distance between the pivotal quantity $\sqrt{r}(\hat{f}(\omega_j)-f(\omega_j))$ and its bootstrap approximation $\sqrt{r}(\hat{f}^{*}(\omega_j)-\hat{f}(\omega_j))$. Here, the Mallows distance between distributions F and G is defined as

$$d_2(F,G)=\inf\left\{E\left[|X-Y|^2\right]^{1/2}\right\}$$

where the infimum is taken over all pairs of random variables X and Y having marginal distributions F and G, respectively. We adopt the convention that where random variables appear as arguments of d_2, they represent the corresponding distributions. In particular, bootstrap quantities represent their conditional distribution given the original data.

Theorem.

Under the preceding conditions, the bootstrap principle holds in the following form:

$$d_2\left(\sqrt{r}\left(\hat{f}(\omega_j)-f(\omega_j)\right),\sqrt{r}\left(\hat{f}^*(\omega_j)-\hat{f}(\omega_j)\right)\right)\underset{r\to\infty}{\longrightarrow}0 \text{ in probability} \quad (1)$$

Proof:

By definition $\sqrt{r}\left(\hat{f}(\omega_j)-f(\omega_j)\right)=\frac{f(\omega_j)}{\sqrt{r}}\sum_{i=1}^{r}(W_{ij}-1)$ having distribution $\psi(F_j)$ and

$\sqrt{r}\left(\hat{f}^*(\omega_j)-\hat{f}(\omega_j)\right)=\frac{\hat{f}(\omega_j)}{\sqrt{r}}\sum_{i=1}^{r}\left(\hat{W}_{ij}^*-1\right)$ having distribution $\hat{\psi}\left(\hat{F}_{jr}\right)$. Therefore, replacing in (1), we

finally get

$$d_2^2\left(\psi(F_j),\hat{\psi}\left(\hat{F}_{jr}\right)\right)\to 0 \text{ in probability} \quad (2)$$

Since d_2 is a metric, and we obtain

$$d_2^2\left(\psi(F_j),\hat{\psi}\left(\hat{F}_{jr}\right)\right)\le 2d_2^2\left(\psi(F_j),\psi(F_{jr})\right)+2d_2^2\left(\psi(F_{jr}),\hat{\psi}(F_{jr})\right)+2d_2^2\left(\hat{\psi}(F_{jr}),\hat{\psi}\left(\hat{F}_{jr}\right)\right)$$

We will prove that each term of the second member in the inequality converges to zero in probability.

Firstly, $d_2^2\left(\psi(F_j),\psi(F_{jr})\right)\to 0$ in probability

Let F_j denote the distribution function of W_{ij} and F_{jr} denote the empirical distribution function of $\{W_{1j},\ldots,W_{rj}\}$. Let W'_{ij} be random variables with a distribution function F_{jr}.

$$d_2^2\left(\psi(F_j),\psi(F_{jr})\right)=d_2^2\left(\frac{f(\omega_j)}{\sqrt{r}}\sum_{i=1}^{r}(W_{ij}-1),\frac{f(\omega_j)}{\sqrt{r}}\sum_{i=1}^{r}(W'_{ij}-1)\right)$$

An application of Bickel and Freedmans's (1981) lemma 8.6 lead to

$$d_2^2\left(\frac{f(\omega_j)}{\sqrt{r}}\sum_{i=1}^{r}(W_{ij}-1),\frac{f(\omega_j)}{\sqrt{r}}\sum_{i=1}^{r}(W'_{ij}-1)\right)\leq\frac{f^2(\omega_j)}{r}\sum_{i=1}^{r}d_2^2(W_{ij}-1,W'_{ij}-1) \quad (3)$$

$\forall l,k \;\; d_2^2(W_{lj}-1,W'_{lj}-1)=d_2^2(W_{kj}-1,W'_{kj}-1)$, using it in (3)

$$\frac{f^2(\omega_j)}{r}\sum_{i=1}^{r}d_2^2(W_{ij}-1,W'_{ij}-1)=f^2(\omega_j)d_2^2(W_{ij}-1,W'_{ij}-1)$$

by lemma 8.3 of Bickel and Freedman's (1981) $d_2^2(W_{ij},W'_{ij})\xrightarrow[a.s.]{r\to\infty}0$

Secondly, we have to prove that $d_2^2(\psi(F_{jr}),\hat{\psi}(F_{jr}))\to 0$ in probability for $r\to\infty$.

$$d_2^2(\psi(F_{jr}),\hat{\psi}(F_{jr}))=d_2^2\left(\frac{f(\omega_j)}{\sqrt{r}}\sum_{i=1}^{r}(W'_{ij}-1),\frac{\hat{f}(\omega_j)}{\sqrt{r}}\sum_{i=1}^{r}(W'_{ij}-1)\right)$$

$$\leq\frac{1}{r}\sum_{i=1}^{r}d_2^2\left(f(\omega_j)(W'_{ij}-1),\hat{f}(\omega_j)(W'_{ij}-1)\right)=d_2^2\left(f(\omega_j)(W'_{ij}-1),\hat{f}(\omega_j)(W'_{ij}-1)\right)$$

By definition of the Mallows metric

$$d_2^2\left(f(\omega_j)(W'_{ij}-1),\hat{f}(\omega_j)(W'_{ij}-1)\right)=\left|f(\omega_j)-\hat{f}(\omega_j)\right|^2E[W'_{ij}-1]^2 \quad (4)$$

$\hat{f}(\omega_j)$ is a consistent estimator for $f(\omega_j)$ and converges in mean square for $r\to\infty$

$E[W'_{ij}-1]^2\leq\sigma_j^2$, applying this in (4), we obtain:

$$\left|f(\omega_j)-\hat{f}(\omega_j)\right|^2E[W'_{ij}-1]^2\xrightarrow{p}0$$

Finally, we have to prove that $d_2^2(\hat{\psi}(F_{jr}),\hat{\psi}(\hat{F}_{jr}))\to 0$ in probability for $r\to\infty$.

$$d_2^2(\hat{\psi}(F_{jr}),\hat{\psi}(\hat{F}_{jr}))=d_2^2\left[\frac{\hat{f}(\omega_j)}{\sqrt{r}}\sum_{i=1}^{r}(W'_{ij}-1),\frac{\hat{f}(\omega_j)}{\sqrt{r}}\sum_{i=1}^{r}(\hat{W}^*_{ij}-1)\right]\leq\frac{\hat{f}^2(\omega_j)}{r}\sum_{i=1}^{r}d_2^2(W'_{ij}-1,\hat{W}^*_{ij}-1)$$

$$=\hat{f}^2(\omega_j)d_2^2(W'_{ij},\hat{W}^*_{ij})=\hat{f}^2(\omega_j)d_2^2(F_r,\hat{F}_r)$$

By definition of the Mallows metric $d_2^2(F_r, \hat{F}_r)$, we may consider the joint distribution of $\{W_{ij}\}$ and $\{\hat{W}_{ij}\}$, which assigns probability $1/_r$ at each $(w_{ij}, \hat{w}_{ij})$ for each j fixed to establish that

$$d_2^2(W_{ij}^*, \hat{W}_{ij}^*) \le \frac{1}{r}\sum_{i=1}^{r}(W_{ij} - \hat{W}_{ij})^2 = \frac{1}{r}\sum_{i=1}^{r}\left(\frac{I_i(\omega_j)}{f(\omega_j)} - \frac{I_i(\omega_j)}{\hat{f}(\omega_j)}\right)^2$$

$$= \frac{1}{r}\sum_{i=1}^{r}\left(\frac{1}{f(\omega_j)} - \frac{1}{\hat{f}(\omega_j)}\right)^2 I_i^2(\omega_j) = \left(\frac{\hat{f}(\omega_j) - f(\omega_j)}{\hat{f}(\omega_j) f(\omega_j)}\right)^2 \frac{1}{r}\sum_{i=1}^{r} I_i^2(\omega_j)$$

Since we know $\hat{f}(\omega_j)$ is a consistent estimator and $\hat{f}(\omega_j) \xrightarrow{r\to\infty} f(\omega_j)$ converges in mean square.

On the other hand, it can be proved that $\frac{1}{r}\sum_{i=1}^{r} I_i^2(\omega_j) = \frac{f^2(\omega_j)}{r}\sum_{i=1}^{r} W_{ij}^2 \xrightarrow{p} f^2(\omega_j)(\sigma_j^2 + 1)$.

Therefore, we can write:

$$\left(\frac{\hat{f}(\omega_j) - f(\omega_j)}{\hat{f}(\omega_j) f(\omega_j)}\right)^2 \frac{1}{r}\sum_{i=1}^{r} I_i^2(\omega_j) \xrightarrow{p} 0$$

Collecting together the three terms of the Mallows metric, we now have that this converges in probability to 0.

4. Simulation

In this section, a simulation study illustrates the performance of our bootstrap approach. We have considered r subjects and for each one, a moving average process (MA(2)) has been simulated at same time intervals, where the coefficients (ϕ_1, ϕ_2) were chosen at random from a bivariate normal distribution with mean vector *(0.2, -4)*, $Var(\phi_1)=Var(\phi_2)=0.01$ and $Cov(\phi_1,\phi_2)=-0.007$.

We have represented simultaneously for each frequency the population spectrum and the confidence band obtained by means of the bootstrap estimation proposed. For large values of r, we have represented the confidence intervals obtained after approximating the pivot $\sqrt{r}\left(\hat{f}(\omega_j) - f(\omega_j)\right)$

by a normal distribution $N\left(0, f\left(\omega_j\right)\sigma_j^2\right)$, having replaced σ_j^2 by its estimator $\hat{\sigma}_j^2 = \sum_{i=1}^{r}\left(\hat{W}_{ij} - 1\right)^2 \Big/ (r-1)$.

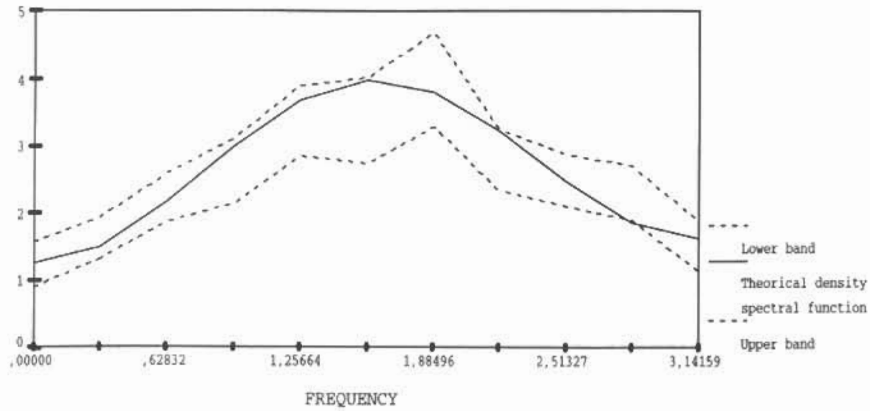

Fig. 1. N=20 r=100

In Fig. 1, due to the fact that the interpolations are made with a low number of frequencies used to estimate the population spectrum, one can think of the existence of a possible bias of the estimator in the maximum of the population spectrum.

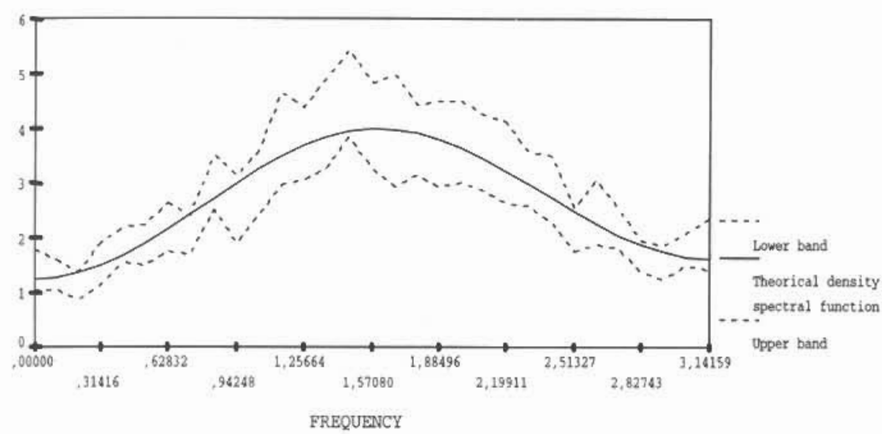

Fig. 2. N=60 r=100

In Fig 2. the increase of frequencies produces as a consequence a rise in the amount of peaks of the confidence bands that could lead to overvalue the contribution of certain frequencies to the spectral power.

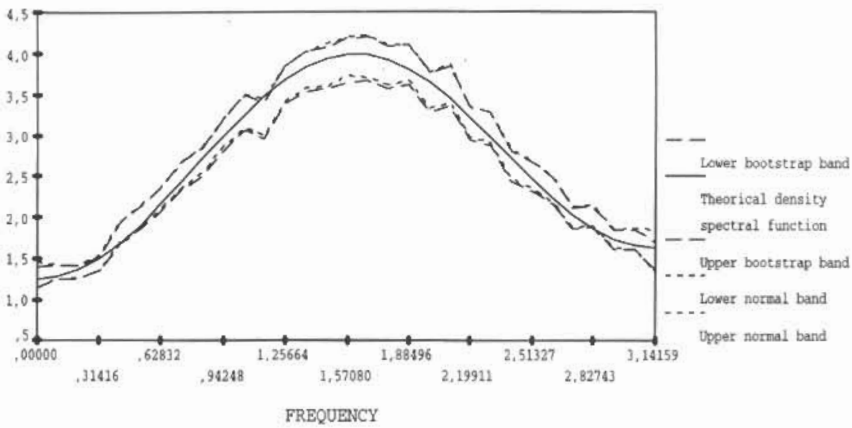

Fig 3 N=60 r=1000

Fig 3 shows that for large values of *r* the bootstrap approximation coincides practically with the normal approximation.

5. Discussion.

The graphics obtained suggest that for large values of *N*, it is more suitable to estimate the population spectrum by means of average periodogram smoothing. So, a more adequate perception is obtained since the peaks in the estimator suggest that some frequencies have negligible contributions to the spectral power. In the same way, a smaller variance for the estimator of population spectrum is obtained although a bias is introduced. Hernández-Flores (1996) proposes some estimators for population spectrum based in smoothing of the average periodogram by means of kernel estimators, for a large *N*.

However, in real life problems, specially in medical sciences, only a small number of observations by individual is available, although a large number of individuals can be analysed. Obviously, it is not adequate to make estimations through smoothings in this case.

In comparison to Diggle and Al-Wasel's model(1993), ours allows the $Var(Z(\omega_j))$ to change with the frequency ω_j . The parametrization of the model they introduced is not justified completely and it produces an estimator $\hat{f}(\omega_j)$ whose variance does not take into account the number N of observations by individual as one could expect from a parametric procedure. Despite the fact of using too the mean periodogram as the spectral density estimator and not using a parametric model, the results we have obtained show that the variance does not increase. As mentioned in the above section, we have used a bootstrap procedure that not only satisfies the asymptotic validity conditions following Mallows metric, but it also achieves confidence regions similar to the ones obtained using a normal approximation when we are dealing with a large number of individuals.

6. References

[1] Bickel, P. and Freedman, D. (1981). "Some Asymptotic Theory for the Bootstrap". The Annals of Statistics. 9: 1196-1217.

[2] Diggle, P. J. and Al-Wasel, I. (1993). "On Periodogram-Based Spectral Estimation for Replicated Time Series". Developments in Time Series Analysis. Chapman & Hall.

[3] Franke, J. and Härdle, W. (1992). "On Bootstraping Kernel Spectral Estimates". The Annals of Statistical. 20:121-145.

[4] Freedman, D. (1981) "Bootstrapping Regression Models". The Annals of Statistics. Vol9, nº 6, 1218-1228.

[5] Härdle, W. and Bowman, A. (1988). "Bootstrapping in Nonparametric Regression: Local Adaptive Smoothing and Confidence Bands". Journal of the American Statistical Association. 83:102-110.

[6] **Hernández-Flores, C. N.** (1996) In preparation. Ph D Thesis, Las Palmas de Gran Canaria University.

[7] **Priestley, M.B.** (1981). " Spectral Analysis and Time Series". Wiley.

[8] **Romano, J**. (1988). Bootstrapping the Mode. Inst. Statist. Math. 40: 565-586.

[9] **Saavedra, P. y Hernández, C.N.** (1994). "Un problema de estimación espectral con series replicadas". XXI Congreso Nacional de Estadística e Investigación Operativa. Calella, 1994.

[10] **Saavedra, P., Artiles, J. y Hernández, C.N.** (1995). "Un problema de estimación espectral utilizando técnicas bootstrap". XXII Congreso Nacional de Estadística e Investigación Operativa. Sevilla, 1995.

Recibido: 17 Mayo 1996