UNIFORM CONFIDENCE BANDS FOR THE SPECTRUM BASED ON SUBSAMPLES

by

Dimitris N. Politis and Joseph P. Romano
Purdue University Stanford University
Linfeng You
Stanford University

Technical Report #93-44

Department of Statistics
Purdue University

August 1993

◇

Uniform Confidence Bands For The Spectrum Based On Subsamples

Dimitris N. Politis
Department of Statistics
Purdue University
W. Lafayette, IN 47907

Joseph P. Romano
Department of Statistics
Stanford University
Stanford, CA 94305

Linfeng You
Department of Statistics
Stanford University
Stanford, CA 94305

Abstract

In this article, the problem of constructing uniform confidence bands for the spectral distribution function is considered. A subsampling method, based on recomputing the spectrum over subseries of the data, is presented and shown to be asymptotically valid. In fact, the method and theory applies quite generally to the problem of constructing uniform confidence bands for parameters taking values in a function space. A small simulation is also presented.

Keywords. *Jackknife, sampling distribution, spectral distribution, subsamples.*

1 Introduction

The goal of the present work is to present an asymptotically valid method for the construction of uniform confidence bands for the spectral distribution function. The method is based on a jackknife approach in that the basis of the construction is the recomputing of the estimated spectrum over certain subseries of the data. The jackknife has been used predominantly to estimate and remove bias of an estimator, or to estimate the standard error of the estimator. See Carlstein (1986) in the stationary time series context. Here, we utilize the estimate recomputed over subsets of the data to actually approximate an entire sampling distribution. A general theory for such an approach for the construction of confidence intervals for real-valued parameters is presented in Politis and Romano (1992), both in the case of i.i.d. data and in the context of homogeneous random fields. The use of the jackknife to estimate a sampling distribution was first considered in Wu (1990) in the i.i.d. case for statistics that are asymptotically linear. Here, the theory is generalized to accomodate parameters that are functions so that confidence regions for the unknown function may be constructed. Also, very little is assumed about the structure (such as asymptotic linearity) of the statistic. Attention is focused on the stationary time series case here, though generalizations can be made.

In section 2, the method is described and some general theory is presented. The problem is then specialized to the case of the spectral distribution function in section 3. Some numerical work is also presented.

2 Construction and general theory.

Let $X_1, \ldots, X_n$ be observations from a stationary time series. The underlying probability mechanism generating the process is denoted P. The X_i may take values in an arbitrary sample space S. Interest focuses on some unknown parameter $\theta = \theta_P$. We will assume θ_P takes values in a normed linear space Θ, with norm denoted $\|\cdot\|$. The special case where Θ is an appropriate function space endowed with the supremum norm will result in uniform confidence bands for the unknown function; this will be made explicit in section 3 and such a specialization is not necessary at this point.

Let T_n be an estimate of θ_P, also taking values in Θ. A confidence region for θ_P could be constructed if the distribution of $\|T_n - \theta_P\|$ were known. Let $H_n(P)$ denote the distribution of $\tau_n\|T_n - \theta_P\|$ under the true model P. Here, τ_n serves as an appropriate sequence of normalizing constants. Also, let $H_n(\cdot, P)$ be the corresponding c.d.f. of $H_n(P)$. The following will be assumed.

Assumption A. $H_n(P)$ converges weakly to a limit law $H(P)$ as $n \to \infty$.

Assumption A follows if $\tau_n[T_n - \theta_P]$, regarded as a random element of Θ, has a weak limit. Here, Θ must be endowed with an appropriate σ-field so that $\tau_n[T_n - \theta_P]$ is measurable and an appropriate weak convergence theory ensues. We avoid such issues by considering the sequence of real-valued random variables $\tau_n \|T_n - \theta_P\|$ initially.

To describe the method, fix an integer $b < n$ and let $S_{n,i}$ be equal to the statistic T_b evaluated at the subseries of size b given by $\{X_i, X_{i+1}, \ldots, X_{i+b-1}\}$. Thus, $S_{n,i}$ is defined for $i = 1, 2, \ldots, n-b+1$. The approximation to $H_n(x, P)$ we study is given by

$$\hat{H}_n(x) = (n-b+1)^{-1} \sum_{i=1}^{n-b+1} 1\{\tau_b \|S_{n,i} - T_n\| \le x\}.$$

The motivation behind the method is the following. For any i, the data set $\{X_i, X_{i+1}, \ldots, X_{i+b-1}\}$ is a sample of size b under the stationary model P. Hence, for any i, $1 \le i \le n-b+1$, the exact distribution of $\tau_b(S_{n,i} - \theta_P)$ is $H_b(P)$. Hence, the empirical distribution of these $n-b+1$ should serve as a good approximation to $H_b(P)$, assuming the underlying process is weakly dependent. If b and n are large, Assumption A will imply that $H_b(P)$ and $H_n(P)$ are both near each other because they are both near $H(P)$. However, θ_P is unknown, so it is replaced by its estimate T_n, which is asymptotically permissible because $\tau_b \|T_n - \theta_P\|$ is of order τ_b/τ_n in probability. Assumptions on b will ensure τ_b/τ_n tends to zero.

Recall some standard notation. For a stationary time series $X = \{X_n, n \in \mathbf{Z}^+\}$, define Rosenblatt's α-mixing coefficent by $\alpha_X(j) = \sup_{A,B} |P(AB) - P(A)P(B)|$, where A and B vary over events in the σ-fields generated by $\{X_n, n \le k\}$ and $\{X_n, n \ge j+k\}$, respectively. The sequence X is said to be α-mixing if $\alpha_X(j) \to 0$ as $j \to \infty$.

Theorem 1. Assume $X_1, \ldots, X_n$ is generated from a stationary time series which is α-mixing. Assume Assumption A. Also assume $\tau_b/\tau_n \to 0$, $b \to \infty$ and $b/n \to 0$ as $n \to \infty$. Let x be a continuity point of $H(\cdot, P)$, the limit c.d.f. corresponding to $H(P)$. Then, $\hat{H}_n(x) \to H(x, P)$ in probability. If $H(\cdot, P)$ is continuous, then

$$\sup_x |\hat{H}_n(x) - H_n(x, P)| \to 0 \qquad (1)$$

in probability. Let $h_n(1-\alpha) = \inf\{x : \hat{H}_n(x) \ge 1-\alpha\}$. Correspondingly, define $h(1-\alpha, P) = \inf\{x : H(x, P) \ge 1-\alpha\}$. If $H(\cdot, P)$ is continuous at $h(1-\alpha, P)$, then

$$Prob\{\tau_n \|T_n - \theta_P\| \le h_n(1-\alpha)\} \to 1-\alpha \qquad (2)$$

as $n \to \infty$. Thus, the asymptotic coverage probability under P of the set

$$\{\theta \in \Theta : \tau_n \|T_n - \theta\| \le h_n(1-\alpha)\}$$

is $1-\alpha$.

Proof. Let $\tilde{H}_n(x)$ be defined by

$$\tilde{H}_n(x) = (n-b+1)^{-1} \sum_{i=1}^{n-b+1} 1\{\tau_b \|S_{n,i} - \theta_P\| \le x\}.$$

We claim $\tilde{H}_n(x) \to H(x, P)$ in probability, if x is a continuity point of $H(\cdot, P)$. To see why, let $Y_{n,i} = 1\{\tau_b \|S_{n,i} - \theta_P\| \le x\}$. so that $\tilde{H}_n(x)$ is an average of (a triangular array of) stationary, bounded, weakly dependent random variables. Moreover, $E[\tilde{H}_n(x)] = H_b(x, P) \to H(x, P)$ as $n \to \infty$. So, it suffices to show $\mathrm{var}[\tilde{H}_n(x)] \to 0$ as $n \to \infty$. Set $N = n-b+1$. Then, by stationarity,

$$\mathrm{var}[\tilde{H}_n(x)] = \frac{1}{N}\mathrm{var}(Y_{n,1}) + \frac{2}{N}\sum_{i=1}^{N}(1-\frac{i}{N})\mathrm{cov}(Y_{n,1}, Y_{n,1+i}).$$

Bound $\mathrm{cov}(Y_{n,1})$ by one, and for $i = 1, \ldots, b-1$, bound $\mathrm{cov}(Y_{n,1}, Y_{n,1+i})$ by one. So,

$$\mathrm{var}[\tilde{H}_n(x)] \le (2b+1)/N + \frac{2}{N}\sum_{i=b}^{N} |\mathrm{cov}(Y_{n,1}, Y_{n,1+i})|.$$

But, $Y_{n,1}$ is a function of $\{X_1, \ldots, X_b\}$ and $Y_{n,1+i}$ is a function of $\{X_{1+i}, \ldots, X_{i+b}\}$, so that $Y_{n,1}$ and $Y_{n,1+i}$ are determined by X_j's separated by $1+i-b$. So, by the well-known mixing inequality (c.f. Ibragimov (1962)),

$$|\mathrm{cov}(Y_{n,1}, Y_{n,1+i})| \le 4\alpha_X(1+i-b),$$

for $i = b, \ldots, N$. Hence,

$$\mathrm{var}[\tilde{H}_n(x)] \le (2b+1)/N + \frac{8}{N}\sum_{i=b}^{N} \alpha_X(1+i-b)$$

$$\le (2b+1)/N + \frac{8}{N}\sum_{j=1}^{N} \alpha_X(j).$$

The assumptions on b imply $(2b+1)/N \to 0$ as $n \to \infty$. Strong mixing implies $N^{-1}\sum_{j=1}^{N} \alpha_X(j) \to 0$. Thus, $\tilde{H}_n(x) \to H(x, P)$ in probability, if x is a continuity point of $H(\cdot, P)$.

To deduce the same for $\hat{H}_n(x)$, let E_n be the event $\{\tau_b \|T_n - \theta_P\| < \epsilon\}$. Then, assumption A and $\tau_b/\tau_n \to 0$ imply $P(E_n) \to 1$ as $n \to \infty$. So, by the triangle inequality,

$$\tilde{H}_n(x-\epsilon) \le \hat{H}_n(x) \le \tilde{H}_n(x+\epsilon)$$

with probability tending to one. Hence, if $x \pm \epsilon$ are continuity points of $H(\cdot, P)$, the above argument implies

$\tilde{H}_n(x \pm \epsilon) \rightarrow H(x \pm \epsilon, P)$ in probability. Letting $\epsilon \rightarrow 0$ allows one to conclude $\hat{H}_n(x) \rightarrow H(x, P)$ in probability.

The rest of the argument is fairly routine. The result (1) follows by a subsequence argument; see the proof of Theorem 1 in Politis and Romano (1992). The result (2) holds easily if $H(\cdot, P)$ is also assumed strictly increasing at $h(1-\alpha, P)$, in which case one can also deduce that $h_n(1-\alpha) \rightarrow h(1-\alpha, P)$ in probability. Without this assumption, (2) remains true: the argument is similar to the proof of Theorem 1 of Beran (1984) given $\hat{H}_n(x)\ toH(x, P)$ in probability.

Remark 2.1. Since an i.i.d. sequence is α-mixing, the theorem applies to this setting. In the i.i.d. context, however, it is more natural to use all $\binom{n}{b}$ subsamples of size b from the original data. If the underlying process is i.i.d., this would be more efficient. On the other hand, slight deviations from the assumption of independence can easily lead to invalid inferences, so it may be desirable from a robustness point of view to not assume independence.

Remark 2.2. If n is quite large, it may be unnecessary to recompute the statistic over all $n-b+1$ subseries of length b. Instead, one can introduce a *lag* variable h, and compute the statistic over subseries of length b beginning at indices i of the form X_{hj+1} as j runs from $0, \ldots, [(n-b)/h]+1$. Taking $h=1$ is obviously most efficient; see section 3.4 of Politis and Romano (1992).

3 Uniform Confidence Bands For The Spectrum.

The general theory in section 2 was motivated by the problem of constructing a uniform confidence band for the spectral distribution function, now denoted $F(\cdot)$. Here, $\theta = F(\cdot)$. Borrowing notation from Dahlhaus (1985), let $I_n(\lambda)$ denote the periodogram with tapered data, defined by

$$I_n(\lambda) = [2\pi H_{n,2}(0)]^{-1} d_n(\lambda) d_n(-\lambda),$$

where

$$d_n(\lambda) = \sum_{t=1}^{n} h[t/(n+1)] X_t exp[-i\lambda t]$$

and

$$H_{n,k}(\lambda) = \sum_{t=1}^{n} h^k[t/(n+1)] exp[-i\lambda t].$$

The data taper h is assumed of bounded variation and square integrable on $[0,1]$. Let $\hat{F}_n(\cdot)$ be the corresponding integrated periodogram given by

$$\hat{F}_n(\lambda) = \frac{2\pi}{n} \sum_{0<2\pi s/n \leq \lambda} I_n(2\pi s/n).$$

Take $\tau_n = n^{1/2}$ and regard $Y_n(\cdot) = n^{1/2}[\hat{F}_n(\cdot) - F(\cdot)]$ as a random element of $D[0,\pi]$ endowed with the sup norm $\|\cdot\|$. Under suitable weak dependence conditions, the process $Y_n(\cdot)$ converges weakly to a mean zero Gaussian process $Y(\cdot)$ with covariance

$$cov[Y(\lambda), Y(\mu)] = 2\pi G(min\{\lambda, \mu\}) + 2\pi F_4(\lambda, \mu),$$

where

$$G(\lambda) = \int_0^\lambda f^2(\beta) d\beta$$

and

$$F_4(\lambda, \mu) = \int_0^\lambda \int_0^\mu f_4(\alpha, -\alpha, -\beta) d\alpha d\beta;$$

here, f is the spectral denstiy and f_4 is the fourth order cumulant spectrum (see e.g., Brillinger (1975)). For various sets of conditions for this weak convergence to hold, see Anderson (1991), Brillinger (1975), and Dahlhaus (1985). In summary, under weak dependence conditions of the underlying process, our Assumption A holds. Moreover, the limiting distribution has a continuous distribution (a condition needed in Theorem 1); see Tsirel'son (1975). Thus, the subseries method yields asymptotically valid uniform confidence bands for the spectral distribution function. Since the limit distribution is that of the supremum of a certain Gaussian process whose covariance structure depends on intricate fourth order properties of the underlying stationary process, analytical approximations to this limit law would be difficult to obtain; but, see Anderson (1991).

Remark 3.1. The above arguments apply to the case where θ is the standardized spectral distribution function. Consider the process $Z_n(\cdot) = n^{1/2}[\hat{F}_n(\cdot)/\hat{F}_n(\pi) - F(\cdot)/F(\pi)]$. The weak convergence properties of Z_n can be deduced from that of Y_n, so that Assumption A holds here as well.

In fact, the argument can be generalized to get uniform confidence bands for the spectral density itself, which is a harder problem. Here, assumption A must be weakened so that it is assumed $\tau_n \|T_n - \theta_P\| - c_n$ has a limit distribution for some c_n. This assumption holds for spectral density estimates; see Woodroofe and Van-Ness (1967).

4 Some numerical work.

In order to assess how well the method actually works on the problem of constructing a uniform confidence band for the spectral distribution function, a small simulation was done. In all cases considered, the sample size for the data was taken to be $n = 1000$ and the underlying simulated processes were Gaussian. For a given situation, 200 simulated data sets were generated (each having sample size 1000), and the subseries estimated sampling distribution was computed. Confidence bands were constructed at a nominal level of 0.95 in all cases. For example, for an underlying white noise process with block size $b = 30$, in 188 out of the 200 simulations did the resulting confidence band entirely contain the true spectrum, resulting in an estimated coverage probability of 0.94. For $b = 40$ and $b = 60$, the estimate coverages were 0.94 and 0.91. The next situation tried was a moving average process of order one: $X_t = \epsilon_t + 0.5\epsilon_{t-1}$. For $b = 20, 30, 40, 60$, the estimated coverage probabilities were $0.92, 0.93, 0.92, 0.91$, respectively. Finally, a moving average process of order 4 was simulated:

$$X_t = \epsilon_t + 0.75\epsilon_{t-1} + 0.5\epsilon_{t-2} + 0.25\epsilon_{t-3} + 0.5\epsilon_{t-4}.$$

For $b = 20, 30, 40, 60, 100$, the estimated coverages were $0.91, 0.93, 0.93, 0.925, 0.90$, respectively. Although the sample size was large, we feel the method performed adequately, especially given the lack of competing methods for this problem. The choice of block size does not seem to matter too much, and a block size of 30 or 40 seems to do well. Future work will address the choice of block size.

To see how the method performs on a real data set, the Laser Pulsations Data, a data set used in the Time Series Prediction Competition in the fall of 1991, was utilized. The data set is available in an archive on the Santa Fe Institute computer network. To get access into the data files, type *ftp santafe.edu*, login as *anonymous*, and then type: *cd pu/Time – Series/competition*. A full description of the laser intensity data can be found in Hubner, Abraham and Weiss (1989). The log of the data is plotted in Figure 1, and the periodogram is given in Figure 2. A 95 percent uniform confidence band for the spectrum is displayed in Figure 3, having used $b = 30$.

References

[1] Anderson, T. (1991). Goodness of fit tests for spectral distributions. Technical Report 27, Department of Statistics, Stanford University.

[2] Beran, R. (1984). Bootstrap methods in statistics. *Jber. d. Dt. Math.-Verein*, **86**, 14-30.

[3] Brillinger, D. (1975). *Time Series: Data Analysis and Theory*, Holt, Rinehart and Winston: New York.

[4] Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.*, 14, 1171-1179.

[5] Dahlhaus, R. (1985). On the asymptotioc distribution of Bartlett's U_p-statistic. *Journal of Time Series Analysis*, **6**, 213-227.

[6] Hubner, U., Abraham, N. and Weiss, C. (1989). Dimensions and entropies of chaotic intensity pulsations in a single-mode far-infrared NH_3 laser. *Physical Review A*, **40**, 6354-6365.

[7] Ibragimov, I. (1962). Some limit theorems for stationary processes. *Theory of Probability and its Applications*, **7**, 349-382.

[8] Politis, D.N., and Romano, J.P. (1992). A general theory for large sample confidence regions based on subsamples under minimal assumptions. Technical Report 399, Department of Statistics, Stanford University.

[9] Tsirel'son, V. (1975). Density of distribution of the maximum of a gaussian process (in Russian). *Teor. Verojatnost Primenen*, **20**, 856-73.

[10] Woodroofe, M. and VanNess, J. (1967). The maximum deviation of sample spectral densities. *Annals of Mathematical Statistics*, **38**, 1558-1569.

[11] Wu, C.F.J. (1990). On the asymptotic properties of the jackknife histogram. *Ann. Statist.* **18**, 1438-1452.

Figure1: Laser Pulsations Data

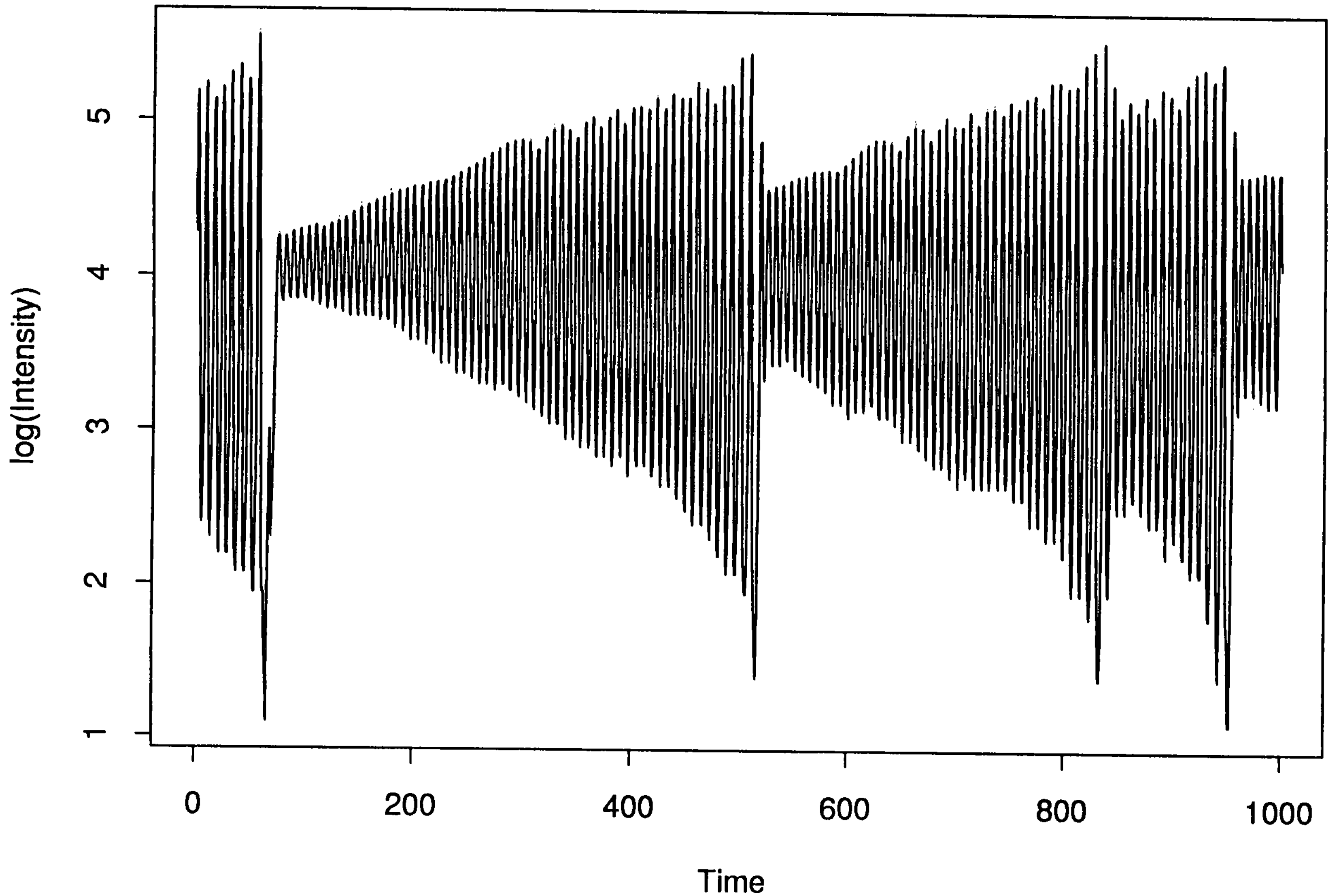

Figure2: The Periodogram of Log Laser Pulsations Time Series

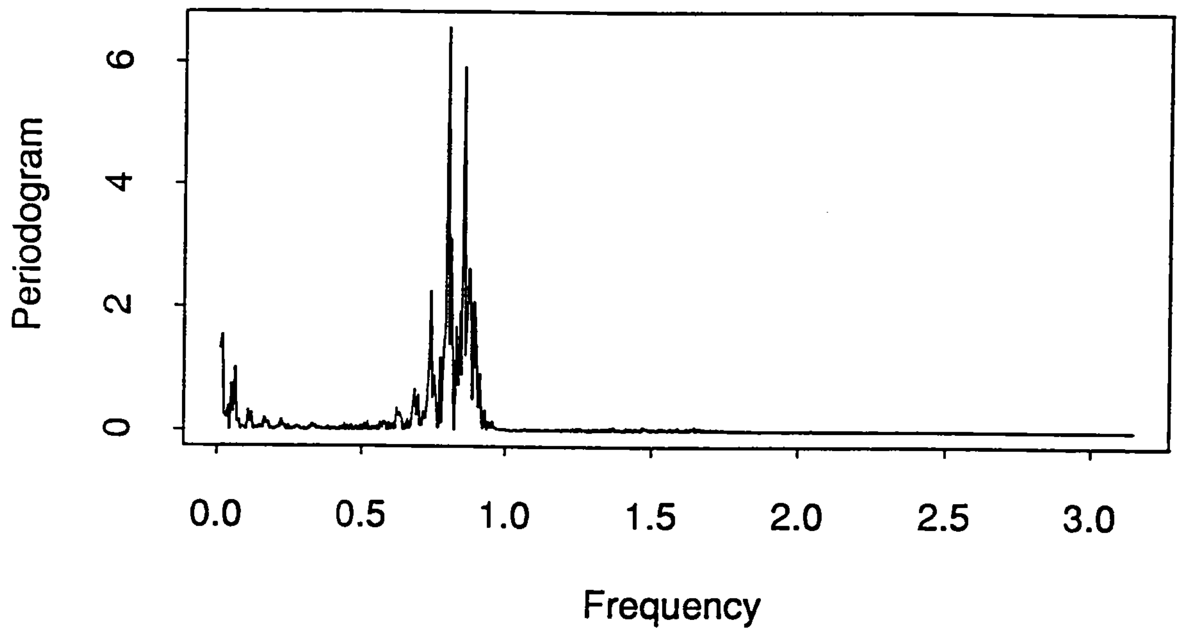

Figure3: The Spectral Estimate Function of Log Laser Pulsations Data