AL- Rafidain
University College

PISSN: (1681-6870); EISSN: (2790-2293)

Journal of AL-Rafidain
University College for Sciences

Available online at: https://www.jrucs.iq

JRUCS
Journal of AL-Rafidain
University College
for Sciences

The Spectral Envelope Estimate via Dirichlet Kernel Function

Tahir R. Dikheel	Ali J. Alkenani
tahir.dikheel@qu.edu.iq	ali.alkenani@qu.edu.iq
Department of Statistics, College of Administration and Economics, University of AL-Qadisiyah, AL-Qadisiyah, Iraq	
Ali G. Abood	
alialmoaly88@gmail.com	
Department of Statistics, College of Administration and Economics, University of AL-Qadisiyah, AL-Qadisiyah, Iraq	

Article Information

Article History:
Received: February, 15, 2024
Accepted: April, 12, 2024
Available Online: December, 31, 2024

Keywords:
Spectral envelope, Dirichlet kernel, SNR, Kullback-Leibler divergence, Tree-Based Adaptive Segmentation.

Correspondence:
Tahir R. Dikheel
tahir.dikheel@qu.edu.iq

Abstract

In this article the spectral envelope function for a nonstationary time series is estimated by using the Dirichlet kernel function. we divide the time series into a number of segmentations, then the nonstationary problem is solved to reduce its effect. The simulation and the sequence of the nitrogenous bases of the DNA of Escherichia coli bacteria are used to implement the goal of this article.

DOI: https://doi.org/10.55562/jrucs.v56i1.44

Introduction

Statistical analysis is one of the most important methods of understanding various phenomena whether physical, biological, or other phenomena. One of the sub-sciences of statistics is time series which is a series of data points indexed in time order. Time series can be analyzed either within the time domain or the frequency domain, and time series are analyzed within the frequency domain in several ways such as spectral analysis or wavelet analysis, but in this article we relied on the spectral envelope analysis method.

There are some time series which its data is categorical, where the categorical data includes all types of data except numbers. time series analysis is complicated if these series are not static and have no trend, therefore, methods have been found to be used to eliminate the trend, in this article have relied on the method of dividing time series into local series to find local spectrum envelopes to reduce the effect of the trend, we have used the sequence of the nitrogenous bases of bacterial

DNA in this article. We have analyzed a categorical time series within the frequency domain, also we have estimated the spectral envelope function, it should be noted that estimating the spectral envelope function requires a weight function, where the weight function was used is the Dirichlet function, noting that efficiency is measured by signal to noise ratio (SNR).

The Fourier Transform

Some phenomenon changes over time in a certain pattern, this called a periodic time series, and the fundamental change on time series can be seen in the pattern, but outside the patterns rang, it is a repetition of a pattern in that rang, any repeat in this time series called frequency which can be one of the domains in series analysis.

Analyzing a time series based on frequency requires periodic signal, this would be a problem if the signal being analyzed was non-periodic, so the Fourier transform is integration make the signals periodic. the Fourier transform mechanism starts by assuming the aperiodic signal is a periodic signal, and the period time approaches infinity as following [8]:

$$F(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)\, e^{-i\omega t} dt$$

Where $F(\omega)$ Fourier transform, ω the frequency, t the time, $f(t)$ the signal at t time.

We can write the orthogonal trigonometric sine and cosine function as:

$$x_t = \sum_{k=0}^{n/2}\left(a_k cos\left(\frac{2\pi kt}{n}\right) + b_k sin\left(\frac{2\pi kt}{n}\right)\right) \qquad t = 1, \dots, n \tag{1}$$

This equation are call Fourier series, we can find the Fourier coefficients a_k and b_k by :

$$a_k = \begin{cases} \frac{1}{n}\sum_{t=1}^{n} x_t cos\left(\frac{2\pi kt}{n}\right) & k = 0 \; and \; k = n/2 \text{ if n is even} \\ \frac{2}{n}\sum_{t=1}^{n} x_t cos\left(\frac{2\pi kt}{n}\right) & k = 1, 2, \dots, \frac{n-1}{2} \text{ if n is odd} \end{cases}$$

$$\mathrm{b_k} = \frac{2}{\mathrm{n}}\sum_{t=1}^{n} x_t \, sin\left(\frac{2\pi kt}{n}\right) \qquad k = 1, 2, \dots, \frac{n-1}{2}$$

Now we can write the above formula in split-complex formula as following:

$$x_t = \begin{cases} \sum_{k=-(n-1)/2}^{(n-1)/2} c_k e^{i\omega_k t} & \text{if } n \text{ is odd} \\ \sum_{k=-n/2+1}^{n/2} c_k e^{i\omega_k t} & \text{if } n \text{ is even} \end{cases}$$

when:

$$\omega_k = \frac{2\pi k}{n} \tag{2}$$

and the Fourier coefficient c_k can be founded by:

$$c_k = \frac{1}{n}\sum_{t=1}^{n} x_t e^{i\omega_k t}$$

By the equations (1) and (2) we get:

$$x_t = \sum_{k=0}^{n/2}\left(a_k cos(\omega_k t) + b_k sin(\omega_k t)\right)$$

$$= \begin{cases} \sum_{k=-(n-1)/2}^{(n-1)/2} c_k e^{i\omega_k t} & \text{if } n \text{ is odd} \\ \sum_{k=-(n/2)+1}^{n/2} c_k e^{i\omega_k t} & \text{if } n \text{ is even} \end{cases}$$

Now we defined a function f(t) as follow:

$$x_t = f(t), \qquad -\frac{p}{2} \le t \le \frac{p}{2}$$

$$x_{t+jp} = x_t \qquad j = \pm 1, \pm 2, \dots$$

that mean x_t are periodic with p, so [11]:

$$x_t = \sum_{k=-\infty}^{\infty} c_k e^{ik\omega_0 t} \tag{3}$$

The estimating of spectral density in equation (3) will not be consistent because of the autocorrelation δk be flexible the higher value of $|k|$ as the following:

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-T}^{T} \delta_k w(k) \cos(k\omega)$$

The accuracy of the autocorrelation function is low because of the lack of observations, thus the effect of the tails of the autocorrelation function increases, so we have to truncate some values of N which T will be the point of truncate, since this truncation causes the neglect part of the information, so we need a weight function for weighting the value of the autocorrelation function [1].

The Dirichlet Kernel estimator is a general stat of the beta Kernel estimator. it is an asymmetric Kernel which is basically derived from Dirichlet distribution, Dirichlet distribution is derived from beta distribution. The Dirichlet Kernel function defined as follow [8]:

$$k_{D,n}(\theta) = \frac{1}{2\pi}\left(1 + 2\sum_{k=1}^{n} \cos(k\theta)\right)$$

The Spectral Envelope

It is a technique which depend on a frequency-based principal component applied to a multivariate time series, now when we suppose that x_t, t=0,±1,±2,… is a stationary categorical time series with finite state-space which is $c = \{c_1, c_2, \dots, c_k\}$ values, then we suppose the numeric value α_j for any c_j. so α is a vector of the real values, α=(α_1,α_2,…,α_k)′, and $p_j = Pr(x_t = c_j) > 0$, and h(x_t) be stationary time series with the real value.

Now we defined Y_t as the flowing:

$$Y_t \begin{cases} z_j \; if \; x_t = c_j \; for \; j = 1,2,3 \dots, k-1 \\ \qquad o \; if \; x_t = c_j \end{cases}$$

When z_j is a vector of k items, all its items are zero except jth row are one, and o are k×1 vector, all its items are zero, then we come collusion $h(x_t) = \boldsymbol{\alpha}' Y_t$, all so $h(x_t) = \alpha_j$

In our application the probability space are the alphabet of the DNA which is {A, C, G, T}, so we can design Y_t as following [9]:

$Y_t = (1, 0, 0)', when\; X_t = A$
$Y_t = (0, 1, 0)', when\; X_t = C$
$Y_t = (0, 0, 1)', when\; X_t = G$
$Y_t = (0, 0, 0)', when\; X_t = T$

So the goal is chose the best value to α so that maximize the power at each frequency ω as follow:

$$\lambda(\omega) = \max_{\alpha} \frac{f_{xx}(\omega{:}\alpha)}{\sigma^2(\alpha)}$$

where $\lambda(\omega)$ is the power of the frequency, $f_{xx}(\omega{:}\alpha)$ is the spectral density and $\sigma^2(\alpha) = var\{h(x_t)\}$ [6, 11].

Suppose the vector process y_t has a continuous spectral density denoted by $f_{yy}(\omega)$ and for each ω there are $k \times k$ complex-valued Hermitian matrix, as we have $h(x_t) = \boldsymbol{\alpha}' Y_t$ suggest $f_{xx}(\omega{:}\alpha) = \boldsymbol{\alpha}' f_{yy}(\omega)\boldsymbol{\alpha}$. now when $f_{yy}^{re}(\omega)$ are the real part of $f_{yy}(\omega)$ and $f_{yy}^{im}(\omega)$ the imaginary part of $f_{yy}(\omega)$, and as $f_{yy}^{im}(\omega)$ is skew-symmetric, so $f_{yy}^{im}(\omega)' = -f_{yy}^{im}(\omega)$ and $x = x^{re} + ix^{im}$, so $\boldsymbol{\alpha}' f_{yy}(\omega)\boldsymbol{\alpha} = \boldsymbol{\alpha}' f_{yy}^{re}(\omega)\boldsymbol{\alpha}$

$$\lambda(\omega) = \frac{\boldsymbol{\alpha}' f_{yy}^{re}(\omega)\boldsymbol{\alpha}}{\boldsymbol{\alpha}' \boldsymbol{V}\boldsymbol{\alpha}}$$

where $\boldsymbol{V}$ is the variance–covariance matrix of y_t, where $p = (p_1, p_2, \dots, p_k)'$, $\boldsymbol{V} = \boldsymbol{D} - \boldsymbol{pp}'$, and $\boldsymbol{D}$ IS $\mathrm{K} \times \mathrm{K}$ diagonal matrix, $\boldsymbol{D} = diag\{p_1, p_2, \dots, p_k\}$. by assumption $p_j > 0$, $j = 1, \dots, k$; so the rank$(V) = k - 1$ with the null space of being spanned by I_k for any $k \times (k-1)$ full rank matrix $\boldsymbol{Q}$

whose columns are linearly independent of I_K. And $\acute{\boldsymbol{Q}}\boldsymbol{V}\boldsymbol{Q}$ is a $(k-1)\times(k-1)$ positive symmetric matrix.
When $f_{yy}(\omega)$ is a consistent each $j = 1, \dots, J$ the largest root of $f_{yy}^{re}(\omega)$ is distinct, then:

$$\left\{\frac{\eta_n\left[\hat{\lambda}(\omega_j)-\lambda(\omega_j)\right]}{\lambda(\omega_j)},\ \eta_n\left[\hat{\propto}(\omega_j)-\propto(\omega_j)\right]; j=1,\dots,J\right\} \tag{4}$$

converges jointly in distribution to independent zero-mean, normal distribution as $n \to \infty$
It can be noted that the value of η_n in the equation (4) depends on the type of the estimator.
In this article the smoothed periodogram matrix [11]

$$I_n(\omega_j) = \hat{f}_{xx} = \sum_{l=-m}^{m} h_l I_n(\omega_j + l/n)$$

When we use the smoothed periodogram matrix with weight h_l then $\eta_n^{-2} = \sum_{l=-m}^{m} h_l^{\,2}$
applying the following approximations, the peak search can be aided in estimating the smooth spectral envelope, using first order Taylor expansion, we have:

$$\log\hat{\lambda}(\omega) \approx \log\lambda(\omega) + \frac{\hat{\lambda}(\omega)+\lambda(\omega)}{\lambda(\omega)}$$

hence $E\left[\log\hat{\lambda}(\omega)\right] \approx \log\lambda(\omega)$ and $\operatorname{var}\left[\log\hat{\lambda}(\omega)\right] \approx \eta_n^{-2}$ because $\eta_n\left(\log\hat{\lambda}(\omega)-\log\lambda(\omega)\right)$ is standard normal [5].

Time Series Stationarity

A nonstationary time series is one that has a trend, and the trend is any systematical change in the level of a time series, the presence of a trend within the series will cause us problems in time series analysis which is the inability to calculate the mean, variance and autocorrelations, therefore the trend must be eliminated or its impact reduced, the method of dividing a time series into a set of segments one of the methods followed, because the effect of the trend decreases with each segment [7].

The algorithm that will be followed in the partitioning process is a Tree-Based Adaptive Segmentation, the series is divided by two segmentations which is the level one, then each segmentation are divided by two in level two till k-times in level k which is the deeper level, when T is the length of the entire series then length of each block are $T/2^k$, we will denote the block $B(k,l)$, l is the l-th block in k level, and N_k is the length of blocks in k level, then if $D(k,l)$ be between two adjacent blocks, $B(K+1,2l)$ and $B(K+1,2l-1)$ compute the estimates of the distances $D(k,l)$ [2].

Let's assume $k\times 1$ a vector value of pricewise stationary process, $\{Y_{s,T}\}_{s=0}^{T-1}$ for $T \geq 1$, is given as:

$$Y_{s,t} = \sum_{b=1}^{B} Y_{s,b} I(\mathrm{s/T}, \mathrm{U}_b)$$

where $Y_{s,b}$ are stationary processes with continuous $k\times k$ spectral matrices $f_{s,b}(\omega)$. And $\mathrm{U}_b = [\mathrm{u}_{b-1}\mathrm{u}_b,) \subset [0,1)$ is the interval, and $I(\mathrm{s/T},\mathrm{U}_b)$ is an indicator which be $equal\ to\ 1\ if\ s/b \in \mathrm{U}_b$, and 0 otherwise…
now let rescale time in each block, so [4]:

$$\{Y_{s,b}: s/t \in \mathrm{U}_b\} \to \{Y_{t,b}: t = 0,\dots,M_b-1\}$$

and the number of observations in segment b is M_b, and $\sum_{b=1}^{B} M_b = T$, this rescaling of time represents a simple time shift to the origin where $Y_{s,b} \to Y_{t,b}$ for $s/t \in \mathrm{U}_b$ with $t = s - \sum_{i=1}^{b-1} M_i$
It can say that a categorical time series, $\{x_{s,T}\}$, on nonzero marginal and a finite state-space is pieceunise stationary when the corresponding $k\times 1$point process, $\{Y_{s,T}\}$, is piecewise stationary, to be sure that more observations considered within each stationary segment (or block) when it sampling the process $x_{s,T}$, it assumed that the lower bound, M, for the number of observations in each block, b, satisfies $M\to\infty$ as $T\to\infty$.

It can define the local spectral envelope as follows when $x_{s,T}$ is a piecewise stationary categorical time series. the local analogue of the optimality criterion:

$$\lambda_b(\omega) \underset{\boldsymbol{\alpha} \propto I_k}{sub} = \frac{\boldsymbol{\alpha}' f_{yy}^{re}(\omega)\boldsymbol{\alpha}}{\boldsymbol{\alpha}' \boldsymbol{v_b}\boldsymbol{\alpha}}$$

for $b = 1,2,\dots,B$ where $\boldsymbol{v_b}$ is the variance-covariance matrix of $Y_{t,b}$ and $\lambda_b(\omega)$ the local spectral envelope and the corresponding eigenvector $\boldsymbol{\alpha}_b(\omega)$ to be the local optimal scaling of block b and frequency ω.

now it can be show some asymptotic $T \to \infty$ results for estimators of the local spectral envelope and the corresponding local scaling vectors.

Next, let $\boldsymbol{Q}$ which it's columns are linearly independent of I_K, $\boldsymbol{Q} = [I_{K-1}|\acute{0}]$, and let $\widehat{\boldsymbol{V}}_b$ be the sample variance-covariance matrix obtained from the data in segment $b, \{Y_{s,T}: s/T \in U_b\}$, or equivalently, $\{Y_{s,b}: t = 0,1,\dots,M_b - 1\}$, so:

$$Y_{s,b} \overset{def}{=} \acute{\boldsymbol{Q}} Y_{s,b}$$

This process effected by removing the k-th element from $Y_{t,b}$ so that it is now a $(k-1) \times 1$ vector, we denote in this case:

$$\widehat{\boldsymbol{V}}_b \overset{def}{=} \acute{\boldsymbol{Q}} \widehat{\boldsymbol{V}}_{\boldsymbol{b}} \boldsymbol{Q} \quad and \quad \hat{f}_{Y,b}(\omega) \overset{def}{=} \acute{\boldsymbol{Q}} \hat{f}_{Y,b}(\omega)\boldsymbol{Q}$$

It can see that $\widehat{\boldsymbol{V}}_b$ and $\hat{f}_{Y,b}(\omega)$ become the upper $(k-1) \times (k-1)$ blocks of the previously defined $\widehat{\boldsymbol{V}}_b$ and $\hat{f}_{Y,b}(\omega)$ matrices respectively, in addition it can use the same convention for the population values $\widehat{\boldsymbol{V}}_b$, and $\hat{f}_{Y,b}(\omega)$.

In order to more clarification while maintaining generality it determined the buoyant cover of the local sample, $\hat{\lambda}_b(\omega)$ to assimilate the largest eigenvalue of $\hat{g}_b^{re}(\omega)$ where:

$$\hat{g}_b = \widehat{\boldsymbol{V}}_b^{-1\backslash 2} \hat{f}_{Y,b} \widehat{\boldsymbol{V}}_b^{-1\backslash 2}$$

The local sample optimal scaling $\widehat{\boldsymbol{\alpha}}_b(\omega)$ defined by $\widehat{\boldsymbol{\alpha}}_b(\omega) = \widehat{\boldsymbol{V}}_b^{-1\backslash 2} \hat{u}_b(\omega)$, where $\hat{u}_b(\omega)$ is the eigenvector of $\hat{g}_b^{re}(\omega)$ related with the root hat $\hat{\lambda}_b(\omega)$ we fixed the scale corresponding to k-th at zero, in addition, let $\hat{u}_b(\omega)$ be normalized so $\acute{\hat{u}}_b \hat{u}_b(\omega) = 1$, we saw that with the first nonzero entry of $\hat{u}_b(\omega)$ considered positive.

In order to let the application of a general theory to obtain asymptotic distributions for the estimates of the local spectral density $f_{Y,b}(\omega)$ we assume in this section that $Y_{t,b}$ is fixed stationary for each of block b, and all local cumulant spectra of all orders be exist for each series $Y_{t,b}$, the issu of assuming the existence of all local cumulant spectra is not concern only the categorical case because the elements of $Y_{t,b}$ can be only take two values: zero or one, instead of entering excessive notation.

The local periodogram of the data $\{Y_{s,T}: s/T \in U_b\}$ in black b is given by:

$$I_b(\omega) = d_b(\omega) d_b^*(\omega)$$

where

$$d_b(\omega) = M_b^{-1\backslash 2} \sum_{t=0}^{M_b - 1} Y_{t,b} e^{-2\pi i \omega t}$$

is the finite Fourier transform of the data $\{Y_{s,T}: s/T \in U_b\}$.

where [6]:

$$\hat{f}_{Y,b} = (2m+1)^{-1} \sum_{i=-m}^{m} I_b(\omega + i/M_b)$$

Although increasing the blocks reduces the effect of the trend, it causes complexity in the algorithm and calculations, so there is a way to reduce the number of blocks without increasing the effect of the trend which is to merge two adjacent blocks when they show similar behavior in the spectral envelope, and we'll use an algorithm Kullback-Leibler divergence to measure the amount of convergence in behavior between blocks, as follow:

$$I\big(p(X), q(x)\big) = \sum \left(\log \frac{p(x)}{q(x)}\right) p(x) \geq 0$$

where p(x) and q(x) denote the probability density functions of random variable x [10].
Finally, we calculated the amount of its efficiency by calculating the SNR which is Signal-to-noise ratio where the higher the ratio, the lower the efficiency of the function, and vice versa. and the SNR is given as [9]:

$$SNR = 10 \log \frac{P_s^2}{p_n^2}$$

The Simulation

First we tested our algorithm before used it in a real data, we used the following equation to generate it:

$$X_1(t) = 2\cos\left(\frac{2\pi t}{10}\right) + \cos\left(\frac{2\pi t}{3}\right) + 0.3\epsilon_1(t) \tag{5}$$

$$X_2(t) = \cos\left(\frac{2\pi t}{3}\right) + 0.01\epsilon_2(t) \tag{6}$$

where $\epsilon_1(t)$ and $\epsilon_2(t)$ are Gaussian white noise and with unit variance, we repeat the experiment 500 times to reach the stability in the results, we set the deepest level at K = 4 to get best segmentation of the data set simulated [5].
It can have obtained that the results as follows table;

Table (1) : This table shows the results of the SNR of the simulation

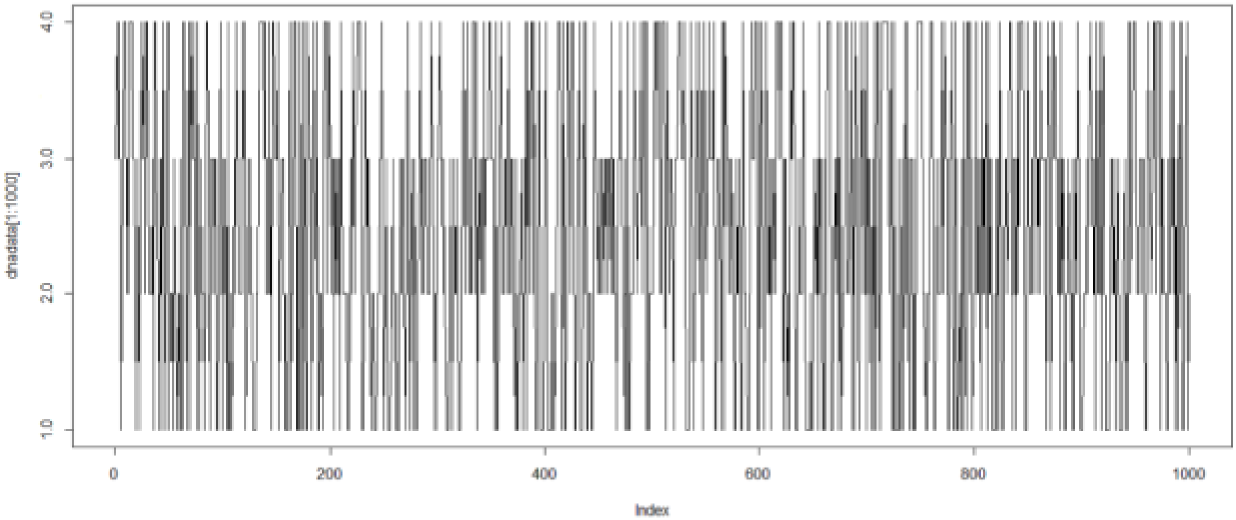
ϵ1	ϵ2	T=32	T=64	T=128	T=256	T=512
0.1	**0.1**	0.3093892	0.3716449	0.4916296	0.6388364	0.8528351
	0.2	0.3095647	0.3843005	0.4875641	0.6353982	0.8610108
	0.3	0.3032232	0.3844411	0.4903628	0.6299271	0.8643097
	0.4	0.3070692	0.3856573	0.4934298	0.6518823	0.8574658
	0.5	0.30329	0.377933	0.4915778	0.6514229	0.859195
0.2	**0.1**	0.6208312	0.7628395	0.9920309	1.304429	1.718979
	0.2	0.6149923	0.7661904	0.9787744	1.270372	1.686205
	0.3	0.6224691	0.7700758	0.9760111	1.297633	1.737542
	0.4	0.6109164	0.7713775	0.9666547	1.297604	1.711009
	0.5	0.6150353	0.7532902	0.9622879	1.28379	1.705533
0.3	**0.1**	0.8981534	1.149613	1.44927	1.919325	2.622079
	0.2	0.9446555	1.147468	1.476582	1.938994	2.574888
	0.3	0.9329392	1.163584	1.460798	1.972616	2.587633
	0.4	0.9139074	1.174593	1.464646	1.9409	2.603265
	0.5	0.9307467	1.165727	1.475798	1.926553	2.574234
0.4	**0.1**	1.242426	1.552434	1.955448	2.607022	3.446554
	0.2	1.232675	1.524964	1.966233	2.545621	3.449007
	0.3	1.236753	1.52105	1.968392	2.599238	3.438864
	0.4	1.225208	1.51946	1.919457	2.578736	3.463009
	0.5	1.220586	1.526252	1.927068	2.580245	3.424694
0.5	**0.1**	1.563	1.890793	2.448022	3.244801	4.295082
	0.2	1.504903	1.931691	2.444091	3.217919	4.280181
	0.3	1.532101	1.925021	2.44182	3.194905	4.299988
	0.4	1.525421	1.887476	2.434137	3.230507	4.324139
	0.5	1.542217	1.904484	2.450165	3.260982	4.303323

In the table above, we calculated the SNR value when $\epsilon_1(t) = 0.1, 0.2, 0.3, 0.4, 0.5$ and $\epsilon_2(t) = 0.1, 0.2, 0.3, 0.4, 0.5$, also at different lengths of the time series T, when $T = 32, 64, 128, 256, 512$ we noticed that the SNR value increases with the increase of the $\epsilon_1(t)$

value and the length of the chain, at the same time it is not affected by the $\epsilon_2(t)$value.

Escherichia Coli DNA

In this article, the application made to the DNA of Escherichia coli strain Iso00225 chromosome, this data has been collected from the website NCBI.
It is one of the most important types of bacteria that commonly found in the intestines of mammals. it was discovered by German-Austrian pediatrician and biologist Theodor Escherich.



In general, the effects of this bacterium are harmless and may even be beneficial, as it is an important source of vitamin K, However, there are many strains that may cause disease in mammals that are infected especially in humans like Urinary tract infection or internal bleeding, this will be the length of the string $T = 8192$.
Representing the first 1000 data in the form of a chart, where the number 1 represents the nuclide A, the number 2 represents the nuclide C, the number 3 represents the nuccludite G, and the number 4 represents the nuccludite T
By applying the scenario to the data, we obtained the result of SNR = 13.97456

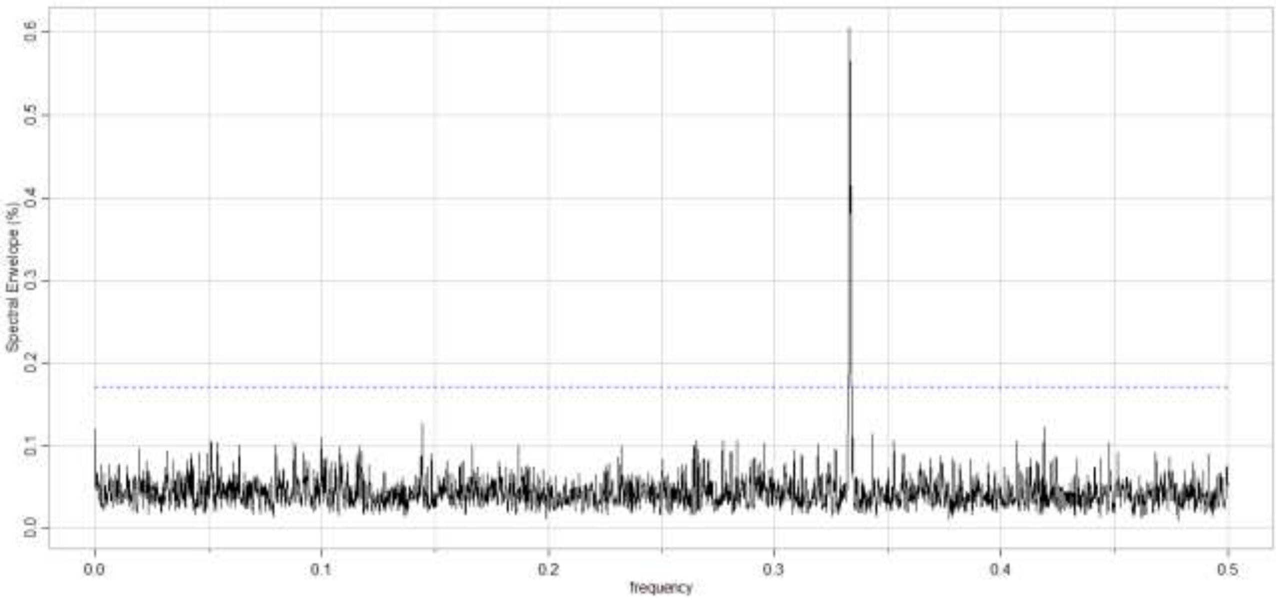


Figure (1): The spectral envelop estimation under the Dirichlet kernel function

Here we notice a peak at frequencies $\omega_{2730} = 0.33325$ which coincide to the period of $P = 18.85427$ where $P = 2\pi/\omega$. This shows that the data has an approximate eighteen sequence cycle. This peak indicates the possibility of a genetic mutation. Where during a DNA test, it is stimulated to replicate and reproduce itself.

Conclusions

Through simulations, we reached the following result, which is that the SNR increases with two factors: the first is the length of the time series, and the second is the $\epsilon_1(t)$ value, while $\epsilon_2(t)$ did not affect the value of the SNR due to the small coefficient of $\epsilon_2(t)$in Equation (6), where $\epsilon_1(t)$ and $\epsilon_2(t)$ are Gaussian white noise and with unit variance. Furthermore, the DNA of Escherichia coli strain Iso00225 chromosome data has an approximate eighteen sequence cycle. This peak indicates the possibility of a genetic mutation.

References

[1] K. S. Chan, and J. D. Cryer, “Time series analysis with applications in R”, springer publication, 2008.

[2] H. Jeong, “The spectral analysis of nonstationary categorical time series using local spectral envelope”, Doctoral dissertation, University of Pittsburgh, 2012.

[3] J. Lin, “On the Dirichlet distribution”, Department of Mathematics and Statistics, Queens University, 10-11, 2016.

[4] S D. S.toffer, H. C. Ombao, and D. E. Tyler, “Local spectral envelope: an approach using dyadic tree-based adaptive segmentation”, Annals of the Institute of Statistical Mathematics, 54, 201-223. 2002.

[5] D. S. Stoffer, “Detecting common signals in multiple time series using the spectral envelope”, Journal of the American Statistical Association, 94(448), 1341-1356, 1999.

[6] D. S. Stoffer, D. E. Tyler, and A. J. McDougall, “Spectral analysis for categorical time series: Scaling and the spectral envelope”, Biometrika, 80(3), 611-622, 1993.

[7] W. Vandaele, “Applied time series and Box-Jenkins models”, 時系列入門, 67-123, 1988.

[8] W. W. WEI, “Time series analysis: univariate and multivariate methods “, USA, Pearson Addison Wesley, Segunda edicion. Cap 10: 212-235, 2006.

[9] Papadopoulos, P., Tsiartas, A., & Narayanan, S. (2016). Long-term SNR estimation of speech signals in known and unknown channel conditions. IEEE/ACM Transactions on audio, speech, and language processing, 24(12), 2495-2506.

[10] Pérez-Cruz, F. (2008, July). Kullback-Leibler divergence estimation of continuous distributions. In 2008 IEEE international symposium on information theory (pp. 1666-1670). IEEE.

[11] Shumway, R. H., Stoffer, D. S., & Stoffer, D. S. (2000). Time series analysis and its applications (Vol. 3). New York: springer.

مجلة كلية الرافدين الجامعة للعلوم (2024)؛ العدد 56؛ 497 -505

AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

مجلة كلية الرافدين الجامعة للعلوم

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

تقدير الغلاف الطيفي باستخدام دالة Dirichlet kernel

طاهر ريسان دخيل	علي جواد الكناني
tahir.dikheel@qu.edu.iq	ali.alkenani@qu.edu.iq
قسم الاحصاء، كلية الادارة والاقتصاد، جامعة القادسية، العراق	
علي غانم عبود	
alialmoaly88@gmail.com	
قسم الاحصاء، كلية الادارة والاقتصاد، جامعة القادسية، العراق	

المستخلص

تم في هذا البحث تقدير دالة الغلاف الطيفية لسلسلة زمنية نوعية غير مستقرة باستخدام دالة Dirichlet kernel حيث نقوم بتقسيم السلسلة الزمنية إلى عدد من الاجزاء، ومن ثم يتم حل مشكلة عدم الاستقرارية بتقليل تأثيرها. تم استخدام المحاكاة وبيانات تمثل سلسلة القواعد النيتروجينية للحمض النووي DNA لبكتيريا الإشريكية القولونية لتحقيق هدف هذا البحث.

معلومات البحث

تواريخ البحث:

تاريخ تقديم البحث:15/2/2024
تاريخ قبول البحث:12/4/2024
تاريخ رفع البحث على الموقع: 31/12/2024

الكلمات المفتاحية:
الغلاف الطيفي، نواة ديريشليت، نسبة الإشارة إلى الضوضاء (SNR)، تباعد كولباك-ليبلر، التجزئة التكيفية القائمة على أساس الشجرة.

للمراسلة:
طاهر ريسان دخيل
tahir.dikheel@qu.edu.iq

DOI: https://doi.org/10.55562/jrucs.v56i1.44