# TESTES PARA COMPONENTES PERIÓDICAS EM SÉRIES TEMPORAIS

Jacira Guiro Carvalho da Rocha

DISSERTAÇÃO APRESENTADA

AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU DE MESTRE

EM

ESTATÍSTICA

ÁREA DE CONCENTRAÇÃO: ESTATÍSTICA

ORIENTADOR: PROF. DR. PEDRO ALBERTO MORETTIN

Este trabalho foi parcialmente financiado pela CAPES e UFPe

- SÃO PAULO, AGOSTO DE 1983 -

À

Enivaldo

e

Leonardo

## AGRADECIMENTOS

Ao terminar este trabalho queremos agradecer a todos que contribuíram para a sua realização, em particular:

- Ao Professor Doutor Pedro Alberto Morettin, pela paciência e dedicação com que nos orientou e transmitiu seus conhecimentos.

- A Enivaldo Carvalho da Rocha, pela assessoria na elaboração dos programas de computação e pelo incentivo constante.

- A todos que direta ou indiretamente colaboraram e nos apoiaram no decorrer do trabalho.

- A Vera Lucia Martins, pelo eficiente trabalho de datilografia.

# Í N D I C E

- o -

# CAPÍTULO 1

## INTRODUÇÃO

Para muitos pesquisadores, em áreas tais como oceanografia, astronomia, administração e outras, a possibilidade de detectar periodicidades "ocultas" cuja presença não seja sequer suspeitada antes de uma análise específica, poderia ser uma primeira motivação para o uso da análise espectral.

Se k termos periódicos de frequências $\omega_1, \omega_2, \ldots, \omega_k$ estão presentes nos dados, então um estimador espectral mostrará picos nestas frequências, e uma simples análise visual nos possibilitará identificá-los.

Ao realizar uma análise espectral é normal associar periodicidades a picos do periodograma, mas nem todos os picos deste estimador espectral são necessariamente causados pela existência de termos harmônicos, os mesmos podendo ocorrer devido à flutuações aleatórias.

Torna-se então necessário, na prática, decidir se um pico do periodograma implica a presença de um termo estritamente periódico na série observada, ou não.

Surge então a utilização de testes de hipótese estatísticos para detectar a existência de tais componentes.

Os primeiros a se preocuparem com este problema foram Schuster (1898) e Fisher (1929).

Desde então vários outros testes foram propostos, dentre os quais se notabilizaram os de Whittle (1952), Hannan (1961) e Priestley (1962).

No Capítulo 2, abordamos alguns conceitos básicos da teoria da análise espectral.

No Capítulo 3, apresentamos um teste para verificar a existência de componentes harmônicas numa série de tempo que tem espectro puramente discreto.

No Capítulo 4, discutiremos a separação de componentes espectrais discretas e contínuas e apresentaremos três testes.

No Capítulo 5, desenvolveremos o método do correlograma, proposto por Priestley (1962), que ao contrário dos anteriores não se baseia na análise do periodograma.

No Capítulo 6, apresentaremos algumas aplicações que consistem na simulação de séries de tempo com estrutura conhecida, cujos programas se encontram no Apêndice.

No presente trabalho não será abordado o caso multivariado. Mac Neill (1974, 1977) trata do problema de testar se várias séries temporais têm uma periodicidade comum.

# CAPÍTULO 2

# CONCEITOS FUNDAMENTAIS DA TEORIA ESPECTRAL

## 2.1 - Introdução

Neste capítulo apresentaremos brevemente um resumo da análise espectral, que será utilizada no desenvolvimento dos testes para detectar periodicidades ocultas nos capítulos que seguem.

O conteúdo deste capítulo pode ser visto em Koopmans (1974) e Brillinger (1975).

## 2.2 - Definições

Definição 2.2.1 - Seja T um conjunto arbitrário. Um processo estocástico é uma família $\mathcal{X}=\{X(t,\omega),\ t\in T,\ \omega\in\Omega\}$, tal que para cada $t\in T$, $X(t,\omega)$ é uma variável aleatória definida num espaço de probabilidades $(\Omega,\mathcal{A},P)$.

Se T é finito ou infinito enumerável, temos um processo estocástico a parâmetro discreto. Se T é um intervalo de $\mathbb{R}=(-\infty,\infty)$, temos um processo estocástico a parâmetro contínuo.

Para cada $\omega$ fixo, $X(t,\omega)$ é uma série temporal, isto é, uma particular realização de um processo estocástico.

No que segue denotaremos $X(t,\omega)$ simplesmente por $X(t)$.

Definição 2.2.2 - Um processo estocástico $\{X(t),\ t \in T\}$ diz-se estritamente estacionário se a distribuição conjunta de $(X(t_1), X(t_2),\ldots,X(t_k))$ é idêntica à distribuição conjunta de $(X(t_1+h), \ldots,X(t_k+h))$ para quaisquer $t_1,\ldots,t_k,h \in T$.

Definição 2.2.3 - Um processo estocástico $\{X(t),\ t \in T\}$ é dito estacionário de 2ª ordem (ou simplesmente estacionário), se e somente se:

i) $E[X(t)] = \mu$, constante, $\forall t \in T$

ii) $E[X^2(t)] < \infty$, $\forall t \in T$

iii) $Cov[X(t),X(s)]$ é uma função apenas de $t-s$, $t,s \in T$.

Como $E[X(t)]$ é constante, podemos, sem perda de generalidade, considerá-la igual a zero, e então pode-se escrever:

$$\gamma(k) = Cov[X(t),X(t+k)] = E[X(t)X(t+k)], \qquad (2.2.1)$$

onde $\gamma(k)$ é denominada função de autocovariância do processo estacionário de 2ª ordem, tendo as seguintes propriedades:

i) $\gamma(0) \geq 0$

ii) $\gamma(-k) = \gamma(k)$

iii) $|\gamma(k)| \leq \gamma(0)$

iv) $\gamma(k)$ é positiva definida, isto é,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j\gamma(k_i-k_j) \geq 0, \quad \forall \alpha_1,\ldots,\alpha_n \in \mathbb{R} \text{ e } k_1,\ldots,k_n \in T.$$

Definição 2.2.4 - Uma série $\{\xi(t),\ t \in Z\}$ de variáveis aleatórias independentes e identicamente distribuídas, tal que $E[\xi(t)]=0$, $E[\xi^2(t)]=\sigma^2$ e $E[\xi(t)\xi(s)]=0$, $\forall t\neq s$ é chamada ruído branco.

Usualmente $\xi(t)\sim N(0;\sigma^2)$, mas não necessariamente.

Teorema 2.2.1 - Se o processo $\{X(t),\ t\in\mathbb{R}\}$ tem função de autocovariância $\gamma(\tau)$, então $\gamma(\tau)$ pode ser representada por:

$$\gamma(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau}\, dF(\omega), \quad \tau\in\mathbb{R} \tag{2.2.2}$$

onde $F(\omega)$ é uma função não negativa, não decrescente, limitada e denominada função de distribuição espectral de $X(t)$.

Teorema 2.2.2 - Seja $\{X(t),\ t\in\mathbb{R}\}$ um processo estacionário de 2ª ordem, então está associado a ele um processo $Z(\omega)$ de incrementos ortogonais, tal que $X(t)$ pode ser representado na forma:

$$X(t) = \int_{-\infty}^{\infty} e^{i\omega t}\, dZ(\omega) \tag{2.2.3}$$

onde $Z(\omega)$ é denominado processo espectral associado a $X(t)$ e:

i) $E[Z(\omega)] = 0$

ii) $E[|Z(\omega)|]^2 = F(\omega)$

iii) $E[dZ(\omega)\overline{dZ(\lambda)}] = \begin{cases} 0, & \text{se } \omega\neq\lambda \\ dF(\omega), & \text{se } \omega=\lambda \end{cases}$

Se $F(\omega)$ é derivável, com derivada $f(\omega)$, então (2.2.3) pode ser escrita como:

$$\gamma(\tau) = \int_{-\infty}^{\infty} f(\omega) e^{i\omega\tau}\, d\omega, \quad -\infty<\tau<\infty \tag{2.2.4}$$

onde $f(\omega)$ é a função de densidade espectral de $X(t)$.

Se $\gamma(\tau)$ é absolutamente integrável, então $f(\omega)$ pode ser representada como a transformada de Fourier, dada por:

$$f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \gamma(\tau) e^{-i\omega\tau} d\tau, \quad -\infty<\omega<\infty \tag{2.2.5}$$

Se a série $X(t)$ é discreta, $t\in\mathbb{Z}$, os limites de integração de (2.2.2) serão $-\pi$ e $\pi$, obtendo-se:

$$\gamma(k) = \int_{-\pi}^{\pi} f(\omega) e^{i\omega k} d\omega, \quad k\in\mathbb{Z}=\{0,\pm1,\ldots\} \tag{2.2.6}$$

e se $\gamma(k)$ é absolutamente somável, então:

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-i\omega k}, \quad -\pi\leqslant\omega\leqslant\pi \tag{2.2.7}$$

## 2.3 - Estimadores Espectrais

Consideremos agora a série de observações $X(1),\ldots,X(n)$ de um processo estacionário de 2ª ordem; então a transformada finita de Fourier é definida por:

$$T_\omega^{(n)} = \frac{1}{\sqrt{2\pi n}} \sum_{j=1}^{n} X(j) e^{-i\omega j}, \quad -\infty<\omega<\infty \tag{2.3.1}$$

onde $T_\omega^{(n)}$ tem período $2\pi$ e $T_{-\omega}^{(n)}=T_\omega^{(n)}$.

Portanto, basta considerar $\omega\in(-\pi,\pi)$ e $T_\omega^{(n)}$ será calculada nas frequências de Fourier, dadas por:

$$\omega_\ell = \frac{2\pi\ell}{n}, \quad -[\tfrac{n-1}{2}] \leqslant \ell \leqslant [\tfrac{n}{2}]$$

Então (2.3.1) fica:

$$T_{\omega_\ell}^{(n)} = \frac{1}{\sqrt{2\pi n}} \sum_{j=1}^{n} X(j)e^{-i\omega_\ell j}, \quad -[\tfrac{n-1}{2}] \leqslant \ell \leqslant [\tfrac{n}{2}] \tag{2.3.2}$$

Substituindo (2.2.3) em (2.3.2) temos:

$$T_{\omega_\ell}^{(n)} = \int_{-\pi}^{\pi} \Delta(\omega-\omega_\ell)dZ(\omega),$$

onde $|\Delta(\omega-\omega_\ell)|^2$ é o núcleo de Féjer (ver Figueiredo, 1977) e portanto se comporta como uma função delta de Dirac quando $n\to\infty$.

Então, pode-se verificar que:

$$E\{|T_{\omega_\ell}^{(n)}|^2\} \simeq f(\omega_\ell),$$

fato que nos sugere um estimador assintoticamente não viciado para $f(\omega_\ell)$, denominado periodograma e denotado por $I(\omega_\ell)$, ou mais simplesmente por $I_\ell$, dado por:

$$I_\ell = |T_{\omega_\ell}^{(n)}|^2 = \frac{1}{2\pi n} |\sum_{j=1}^{n} X(j)e^{-i\omega_\ell j}|^2 \tag{2.3.3}$$

Pode-se demonstrar (ver Koopmans, 1974) que os $T_{\omega_\ell}^{(n)}$ são assintoticamente independentes e com distribuição aproximadamente normal complexa com média zero e variância $f(\omega_\ell)$, se $\ell\neq 0, \frac{n}{2}$ e normal real para $\ell=0, \frac{n}{2}$.

Consequentemente, os $I(\omega_\ell)$ são assintóticamente independentes e tem distribuição assintótica $\chi^2_{(2)}$ para $\ell\neq 0, \frac{n}{2}$ e $\chi^2_{(1)}$ para $\ell=0, \frac{n}{2}$, ou seja:

$$I(\omega_\ell) \sim \begin{cases} \frac{1}{2} f(\omega_\ell)\chi^2_{(2)}, & \ell \neq 0, \frac{n}{2} \\ f(\omega_\ell)\chi^2_{(1)}, & \ell = 0, \frac{n}{2} \end{cases}$$

Ver Koopmans (1974) para detalhes.

Observamos então que o periodograma tem as seguintes propriedades:

i) $\lim\limits_{n\to\infty} E[I(\omega_\ell)] = f(\omega_\ell)$

ii) $\lim\limits_{n\to\infty} Var[I(\omega_\ell)] = f^2(\omega_\ell)$

iii) $\lim\limits_{n\to\infty} Cov[I(\omega_j), I(\lambda_j)] = 0, \quad \omega_j \neq \lambda_j.$

Das propriedades acima vemos que, a menos que $f(\omega)$ seja zero, a $Var[I(\omega)]$ não tende a zero quando n tende a infinito, ou seja, o periodograma é um estimador não consistente de $f(\omega)$.

Da expressão (2.2.7), um estimador natural de $f(\omega)$ será obtido substituindo $\gamma(k)$ por $\hat{\gamma}(k)$, definido por:

$$\hat{\gamma}(k) \doteq \begin{cases} \frac{1}{n} \sum\limits_{t=1}^{n-|k|} X(t)X(t+|k|), & |k| \leqslant n-1 \\ 0, & |k| > n-1 \end{cases} \tag{2.3.4}$$

onde podemos supor $E[X(t)]=0$, sem perda de generalidade.

Logo, o estimador de $f(\omega)$ será:

$$\hat{f}(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-i\omega k}\hat{\gamma}(k) \tag{2.3.5}$$

Um estimador não viciado de $\gamma(k)$ será obtido se substituirmos n por n-k no denominador de (2.3.4), porém ele apresenta um erro quadrático maior do que o estimador definido anteriormente.

Pode-se demonstrar que o estimador de $f(\omega)$ dado por (2.3.5) é o próprio periodograma e portanto tem as mesmas propriedades.

Com o objetivo de eliminar a inconsistência do periodograma e conservar as outras propriedades, vamos ponderar a função de auto-covariância através de uma sequência de pesos.

Consequentemente, a transformada de Fourier de $\hat{\gamma}(k)$ ponderada será denominada estimador suavizado de covariâncias, e será dada por:

$$\hat{f}(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-i\omega k} w_s(k)\hat{\gamma}(k) \qquad (2.3.6)$$

onde, para um inteiro s<n, a sequência de pesos $w_s(k)$, $k \in \mathbb{Z}$ tem as seguintes propriedades:

i) $0 \leqslant w_s(k) \leqslant w_s(0) = 1$

ii) $w_s(-k) = w_s(k), \quad \forall k$

iii) $w_s(k) = 0, \quad |k|>s$ (2.3.7)

onde s é denominado ponto de truncamento ou "lag number", a função peso $w_s(k)$ é denominada "lag window" e a transformada de Fourier de $w_s(k)$ é denominada janela espectral, e é dada por:

$$W_s(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-i\lambda k} w_s(k), \qquad (2.3.8)$$

De (2.3.7) segue que:

i) $W_s(-\lambda) = W_s(\lambda), \quad \forall\lambda$

ii) $\int_{-\pi}^{\pi} W_s(\lambda)d\lambda = 1$

Como $\hat{f}(\omega)$ é a transformada de Fourier do produto $\omega_s(k).\hat{\gamma}(k)$, podemos obter $\hat{f}(\omega)$ através da convolução das transformadas de Fourier de $\omega_s(k)$ e $\hat{\gamma}(k)$, ou seja:

$$\hat{f}(\omega) = \int_{-\pi}^{\pi} W_s(\omega-\lambda)I(\lambda)d\lambda \tag{2.3.9}$$

Aproximando a integral pela soma de Riemann, o estimador

$$\hat{f}(\omega) = \frac{2\pi}{n} \sum_{\nu=-[\frac{n-1}{2}]}^{[\frac{n}{2}]} W_s(\omega-\omega_\nu)I(\omega_\nu) \tag{2.3.10}$$

tem a mesma distribuição assintótica que o estimador suavizado de covariâncias. Tais estimadores são denominados estimadores suavizados de periodograma, e são expressos por:

$$\bar{f}(\omega) = \sum_{\nu=-[\frac{n-1}{2}]}^{[\frac{n}{2}]} W(\omega-\omega_\nu)I(\omega_\nu) \tag{2.3.11}$$

## 2.4 - Largura de Faixa

Pode-se demonstrar, a partir da covariância assintótica entre os estimadores suavizados nas diferentes frequências, que (ver Jenkins e Watts, 1969):

$$\mathrm{Var}[\hat{f}(\omega)] = f^2(\omega)\,\frac{2\pi}{n}\int_{-\pi}^{\pi} W_s^2(\lambda)d\lambda, \qquad (2.4.1)$$

Seja $w(k/s)=w_s(k)$, tal que $w(\lambda)=w(-\lambda)$, $0 \leqslant w(\lambda) \leqslant w(0)=1$, $|\lambda|<1$; pela relação de Parseval:

$$\mathrm{Var}[\hat{f}(\omega)] = f^2(\omega).\,\frac{I}{n},$$

onde

$$I = s\int_{-1}^{1} w^2(\lambda)d\lambda$$

e $\frac{I}{n}$ mede a redução da variância de acordo com a janela espectral utilizada.

Um parâmetro importante na análise espectral é a largura de faixa, ou largura da base da janela espectral utilizada.

Há várias maneiras de definí-la; uma delas seria tomarmos uma janela centrada na frequência zero e projetarmos no eixo das frequências a metade da energia nessa frequência. A esse intervalo obtido dá-se o nome largura de faixa. Ou seja, é o domínio de valores de $\omega$ aos quais é dado um peso consideravelmente alto.

Outra maneira de definirmos a largura de faixa é como a largura da base de uma janela retangular que forneceria a mesma variância que a do estimador suavizado pela janela espectral utilizada, ou seja:

$$B = \frac{1}{I} = \frac{1}{s\int_{-1}^{1} w^2(\lambda)d\lambda}, \qquad (2.4.2)$$

Temos então, que à medida que s aumenta, a largura de faixa diminui.

Como a variância do estimador espectral suavizado é inversamente proporcional à largura da faixa, a uma maior largura de faixa corresponde uma menor variância; entretanto, aumentando a largura da faixa aumentará também o vício do estimador, e vice-versa.

Portanto deve existir um compromisso entre resolução e estabilidade, que se reflete na escolha do ponto de truncamento s.

# CAPÍTULO 3

# O TESTE DE FISHER E SUA EXTENSÃO

## 3.1 - Introdução

Em alguns casos notamos a presença de componentes periódicas num certo conjunto de dados e nos interessamos em determinar as amplitudes a elas associadas.

Neste capítulo e nos seguintes tentaremos verificar a existência de componentes periódicas numa série de tempo.

Observamos que o periodograma frequentemente tem um formato muito irregular e vários pesquisadores foram tentados a atribuir significado real a muitos picos do periodograma.

Para que essa tentativa não seja mal dirigida surge a utilização de testes para verificar a existência de periodicidades em uma dada série temporal. Fisher (1929) propôs um teste de significância para o maior pico do periodograma e nos fornece uma tabela de valores críticos para vários comprimentos de série.

Em estudos posteriores, Nowroozi (1967) descreve o uso do Teste de Fisher e publica tabelas a serem usadas com o teste. Ele aplica o teste e rejeita a um certo nível de confiança toda amplitude abaixo de um certo valor teórico, não levando em conta que o teste se refere somente à maior amplitude.

Vários autores estenderam o teste para incluir o segundo maior pico e outros, dentre eles Whittle (1952) e Shimshoni (1971).

No presente capítulo apresentamos o teste de Fisher, e uma extensão dele, a qual surgiu das idéias de estatística de ordem, e foi desenvolvida por Grenander e Rosemblatt (1957).

## 3.2 - O Teste de Fisher

Fisher (1929) desenvolveu um teste de significância, utilizado quando queremos fazer a análise harmônica de uma série.

Seja X(i), i=1,...,n (=2m+1), uma amostra aleatória de uma população normalmente distribuída. Quaisquer funções lineares $A=\sum_{i=1}^{n} a_i X_i$ e $B=\sum_{i=1}^{n} b_i X_i$ serão também normalmente distribuídas, e suas médias serão zero, se $\sum_{i=1}^{n} a_i=0$ e $\sum_{i=1}^{n} b_i=0$, a variância será igual à da população se $\sum_{i=1}^{n} a_i^2=1$ e $\sum_{i=1}^{n} b_i^2=1$ e serão independentes se $\sum_{i=1}^{n} a_i b_i=0$

No caso de independência temos que $X = \frac{A^2}{\sigma^2} + \frac{B^2}{\sigma^2}$ terá distribuição qui-quadrado com 2 graus de liberdade, e a probabilidade de exceder qualquer valor x é $e^{-x/c}$, onde c é duas vezes a variância amostral e $\sigma^2$ é a variância da população.

Pode-se verificar facilmente que os coeficientes:

$$a_i = \sqrt{\frac{2}{n}} \cos \frac{2k\pi i}{n} \quad \text{e}$$

$$b_i = \sqrt{\frac{2}{n}} \operatorname{sen} \frac{2k\pi i}{n}$$

satisfazem às condições acima para todos os valores de $k \in Z$.

A decomposição de $X(i)$ em suas componentes harmônicas é dada por:

$$X(i) = \frac{a_0}{2} + \sum_{k=1}^{m} \left[a_k \cos \frac{2k\pi i}{n} + b_k \operatorname{sen} \frac{2k\pi i}{n}\right],$$

onde

$$a_0 = \frac{2}{n} \sum_{i=1}^{n} X(i)$$

$$a_k = \frac{2}{n} \sum_{i=1}^{n} X(i) \cos \frac{2k\pi i}{n}$$

$$b_k = \frac{2}{n} \sum_{i=1}^{n} X(i) \operatorname{sen} \frac{2k\pi i}{n}, \quad k=1,\dots,n \qquad (3.2.1)$$

onde m é o número de coeficientes harmônicos.

A amplitude harmônica $c_k$ do k-ésimo harmônico é definida como:

$$c_k = (a_k^2+b_k^2)^{1/2}, \quad k=1,\dots,m \qquad (3.2.2)$$

A estatística do teste proposto por Fisher é a maior das ordenadas do periodograma nas frequências de Fourier dividida pela soma dessas ordenadas, ou seja: $m = \frac{n}{2}$

$$g^{(1)} = \frac{I^{(1)}}{I_1+\dots I_m} \qquad (3.2.3)$$

onde $I^{(1)}=\max\{I_j,\ j=1,\dots,m\}$.

Podemos reescrever (3.2.3) em função das amplitudes dos harmônicos do seguinte modo:

$$g^{(1)} = \frac{\max(c_k^2,\ k=1,\ldots,m)}{\sum_{j=1}^{m} c_j^2} \qquad (3.2.4)$$

Dado que as ordenadas do periodograma nas frequências de Fourier são aproximadamente independentes, elas flutuam bastante e mostram muitos picos e vales, e como sua distribuição é exponencial, a maior ordenada tende a ser grande comparada com as adjacentes e pode indicar a existência de periodicidade razoavelmente forte.

Para propósitos práticos, Jeffreys (1929) mostrou que:

$$\sum_{j=1}^{m} c_j^2 = \frac{2}{2m+1} \sum_{i=1}^{m} \left[X(i) - \frac{a_0}{2}\right]^2, \qquad (3.2.5)$$

pode ser usada a fim de evitar que todos os $c_j$ sejam calculados para se obter a soma de seus quadrados, caso seja necessário.

Assim (3.2.4) pode ser escrita como:

$$g^{(1)} = \frac{\max(c_k^2,\ k=1,\ldots,m)}{\frac{2}{2m+1} \sum_{i=1}^{m} \left[X(i) - \frac{a_0}{2}\right]^2} \qquad (3.2.6)$$

Como vimos $\frac{X}{\sigma^2} \sim \chi^2_{(2)}$, então a probabilidade de que qualquer particular termo X seja excedido é

$$P[X \geq x] = e^{-x/c} \qquad (3.2.7)$$

Mas se P for definido como a probabilidade de que X exceda o maior dos n valores independentes, temos, de (3.2.7) que:

$$P = 1-(1-e^{-x/c})^n \tag{3.2.8}$$

a qual determina o menor valor de n a ser julgado significante.

<u>Proposição 3.1</u>

Suponha que $I_1,\ldots,I_m$ sejam independentes e exponencialmente distribuídas entre si, isto é:

$$P[I_j \leqslant x] = 1-e^{-x}, \quad j=1,\ldots,m$$

Então (sendo $\ell n\ m = \log_e n$):

i) $P[I^{(1)} \leqslant x] = (1-e^{-x})^m$ e

$P[I^{(1)} \leqslant x+\ell n\ m] \simeq \exp\{-e^{-x}\}$, para m grande.

ii) Para m grande:

$$Y_m = I_1+\ldots+I_m \simeq m,$$

no sentido de que a distribuição de $\frac{Y_m}{m}$ se torna concentrada arbitrariamente próximo do valor 1.

iii) Se $G_m = \frac{I^{(1)}}{Y_m}$ é a estatística de Fisher, então:

$$P[m\ G_m \leqslant x+\ell n\ m] \simeq \exp\{-e^{-x}\}.$$

Demonstração:

i) $P[I^{(1)} \leqslant x] = P[I_1 \leqslant x, \ldots, I_m \leqslant x] = (P[I_1 \leqslant x])^m = (1-e^{-x})^m$

e

$$P[I^{(1)} \leqslant x+\ell n\ m] = (1-e^{-x-\ell n\ m})^m = (1-\frac{1}{m}e^{-x})^m$$

Passando ao limite:

$$\lim_{m\to\infty} P[I^{(1)} \leqslant x+\ell n\ m] = \lim_{m\to\infty} (1-\frac{1}{m}e^{-x})^m = \exp\{-e^{-x}\}$$

Portanto:

$$P[I^{(1)} \leqslant x+\ell n\ m] \simeq \exp\{-e^{-x}\},$$

para m grande, e isto implica que $I^{(1)}$ é provavelmente próximo de $\ell n\ x$.

ii) Sabemos que $I_j \sim \varepsilon(1)$, $j=1,\ldots,m$ então $I_j \sim \Gamma(1,1)$, $j=1,\ldots,m$.

Seja $Y_m = I_1 + \ldots + I_m \sim \Gamma(m,1)$, então $\frac{Y_m}{m} \sim \Gamma(m,m)$.

Mas $\frac{Y_m}{m} \xrightarrow[m\uparrow\infty]{P} \mu$, onde $\mu$ é a média de uma $\Gamma(m,m)$.

Logo $\frac{Y_m}{m} \xrightarrow[m\uparrow\infty]{P} 1$.

Portanto $Y_m \simeq m$, para m grande.

iii) Se $G_m = \frac{I^{(1)}}{Y_m}$, então $m\ G_m = \frac{I^{(1)}}{Y_m/m}$.

De (ii) temos que $\frac{Y_m}{m} \xrightarrow[m\uparrow\infty]{P} 1$, então pelo Teorema de

Slustky:

$$\frac{I^{(1)}}{Y_m/m} \xrightarrow[m\uparrow\infty]{L} Z_k,$$

onde

$$P[Z_k \leqslant z+\ell n\ k] \simeq \exp\{-e^{-z}\},$$

ou seja, $\frac{I^{(1)}}{Y_m/m}$ tem aproximadamente a mesma distribuição que $I^{(1)}$, para m grande.

Portanto:

$$P[m\ G_m \leqslant x+\ell n\ m] \simeq \exp\{-e^{-x}\},$$

para m grande.

Fisher (1929) mostrou que a probabilidade, P, de que $g^{(1)}$ exceda o parâmetro g é:

$$P = m(1-g)^{m-1} - \binom{m}{2}(1-2g)^{m-1} + \ldots + (-1)^{k-1}(1-kg)^{m-1}, \qquad (3.2.9)$$

onde k é o maior inteiro menor que $\frac{1}{g}$ e que denotaremos por $k=[\frac{1}{g}]$.

Na demonstração apresentada por Fisher é utilizado um argumento geométrico.

Apresentaremos aqui uma demonstração analítica (ver Grenander e Rosemblatt (1957) e Hannan (1970)). Uma demonstração alternativa é dada por Anderson (1971).

Para simplificar a notação na demonstração, seja:

$$g^{(1)} = \frac{I^{(1)}}{\sum_{j=1}^{m} I_j} = \frac{y_1}{\sum_{j=1}^{m} y_j}$$

Seja $\phi_1$ a função característica de $\frac{1}{g^{(1)}}$, dada por:

$$\phi_1(t) = m \int_{y_1=0}^{\infty} \int_{y_2=0}^{y_1} \cdots \int_{y_m=0}^{y_1} e^{it\frac{\sum_{j=1}^{m} y_j}{y_1} - \sum_{j=1}^{m} y_j} dy_m \ldots dy_1$$

$$= m \int_{y_1=0}^{\infty} \int_{y_2=0}^{y_1} \cdots \int_{y_m=0}^{y_1} e^{it\frac{\sum_{j=1}^{m-1} y_j}{y_1} - \sum_{j=1}^{m-1} y_j} e^{-(1-\frac{it}{y_1})y_m} dy_m \ldots dy_1$$

$$= m \int_{y_1=0}^{\infty} \int_{y_2=0}^{y_1} \cdots \int_{y_{m-1}=0}^{y_1} e^{it\frac{\sum_{j=1}^{m-1} y_j}{y_1} - \sum_{j=1}^{m-1} y_j} \frac{(1-e^{it-y_1})}{(1-\frac{it}{y_1})} dy_{m-1} \ldots dy_1$$

Integrando sucessivamente até $dy_2$, temos:

$$\phi_1(t) = m \int_{y=0}^{\infty} e^{(it-y)} \frac{(1-e^{it-y})^{m-1}}{(1-\frac{it}{y})^{m-1}} dy$$

Para $m>2$ a expressão acima é absolutamente integrável e, então, pela fórmula de inversão da função característica, temos que a função de frequência de $\frac{1}{g^{(1)}}$ é:

$$f_{\frac{1}{g^{(1)}}}(t) = \frac{1}{2\pi} m \int_{-\infty}^{\infty} \int_{0}^{\infty} \frac{e^{\{-\frac{it}{g^{(1)}} + it-y\}}}{(1-\frac{it}{y})^{m-1}} (1-e^{it-y})^{m-1} dy dt$$

Desenvolvendo $(1-e^{it-y})^{m-1}$, temos:

$$(1-e^{it-y})^{m-1} = \sum_{j=0}^{m-1} \binom{m-1}{j} (-1)^j (e^{it-y})^j$$

Então:

$$f_{\frac{1}{g^{(1)}}}(t) = \frac{1}{2\pi} m \sum_{j=0}^{m-1} \binom{m-1}{j} (-1)^j \int_{-\infty}^{\infty} \int_0^{\infty} \frac{e^{-\frac{it}{g^{(1)}} + (j+1)(it-y)}}{(1 - \frac{it}{y})^{m-1}} dydt$$

$$= \frac{1}{2\pi} m \sum_{j=0}^{m-1} \binom{m-1}{j} (-1)^j \int_{-\infty}^{\infty} [\int_0^{\infty} \frac{e^{it\{j+1 - \frac{1}{g^{(1)}}\}}}{(1 - \frac{it}{y})^{m-1}} dt]\, e^{-y(j+1)} dy$$

Como $t=-iy$ é um polo de ordem $m-1$, a integral se anula para $j > \frac{1}{g^{(1)}} - 1$, portanto:

$$f_{\frac{1}{g^{(1)}}}(t) = m \sum_{j=0}^{[\frac{1}{g^{(1)}} - 1]} (-1)^j \binom{m-1}{j} \int_0^{\infty} \frac{y^{m-1} (\frac{1}{g^{(1)}} - j+1)}{m-2}\, e^{-\frac{y}{g^{(1)}}} dy$$

Fazendo $k=j+1$, temos:

$$f_{\frac{1}{g^{(1)}}}(t) = m \sum_{k=1}^{[\frac{1}{g^{(1)}}]} (-1)^{k-1} \binom{m-1}{k-1} (\frac{1}{g^{(1)}} - k)^{m-2} \int_0^{\infty} \frac{y^{m-1} e^{-\frac{y}{g^{(1)}}}}{(m-2)!} dy$$

$$= m(m-1) \sum_{k=1}^{[\frac{1}{g^{(1)}}]} (-1)^{k-1} \binom{m-1}{k-1} \left(\frac{1}{g^{(1)}} - k\right)^{m-2} \left(\frac{1}{g^{(1)}}\right)^{m}$$

$$= m(m-1) \sum_{k=1}^{[\frac{1}{g^{(1)}}]} (-1)^{k-1} \binom{m-1}{k-1} (1-g^{(1)}k)^{m-2} (g^{(1)})^{2}.$$

Multiplicando pelo Jacobiano, $\left(\frac{1}{g^{(1)}}\right)^2$, e calculando a probabilidade de $g^{(1)}$ exceder g, temos:

$$P[g^{(1)}>g] = m(m-1) \int_g^1 \sum_{k=1}^{[\frac{1}{g^{(1)}}]} (-1)^{k-1} \binom{m-1}{k-1} (1-g^{(1)}k)^{m-2} dg^{(1)} =$$

$$= m(m-1) \sum_{k=1}^{[\frac{1}{g}]} (-1)^{k-1} \binom{m-1}{k-1} \frac{(1-kg)^{m-1}}{k(m-1)} =$$

$$= \sum_{k=1}^{[\frac{1}{g}]} (-1)^{k-1} \binom{m}{k} (1-kg)^{m-1}.$$

Fisher (1929) publicou uma tabela para P=0,05 e dá os valores de g para várias situações, baseado em (3.2.9) e os obtidos utilizando somente o primeiro termo de (3.2.9), que diferem do valor exato somente a partir do 4º dígito (ver Tabela 3.1).

Os valores críticos tabelados são percentis da distribuição da estatística, sob a hipótese nula de que a série consiste de variáveis aleatórias independentes.

Tabela 3.1

Tabela para P=0,05 e m=5(5)50

| m | g (pela fórmula exata) | g (apenas pelo primeiro termo) |
|---|---|---|
| 5 | 0.68377 | 0.68377 |
| 10 | 0.44495 | 0.44495 |
| 15 | 0.33462 | 0.33463 |
| 20 | 0.27040 | 0.27046 |
| 25 | 0.22805 | 0.22813 |
| 30 | 0.19784 | 0.19794 |
| 35 | 0.17513 | 0.17525 |
| 40 | 0.15738 | 0.15752 |
| 45 | 0.14310 | 0.14324 |
| 50 | 0.13135 | 0.13149 |

Fonte: Fisher, Proc. R.Soc. A, 125, 1929

Os valores críticos que não se encontram tabelados podem ser obtidos por interpolação linear.

Em 1963, Fisher estendeu essa tabela para m=5(1)50 e acrescentou P=0,01.

## 3.3 - Aplicações do Teste de Fisher

Como um exemplo de aplicação do teste original, consideremos os dados apresentados na Tabela A.1, os quais são a média mensal de 24 leituras diárias da temperatura de São Paulo, obtidas pelo IAG-USP.

A Figura 3.1 mostra o gráfico da série original, onde podemos observar a existência de um período aproximado de doze meses.

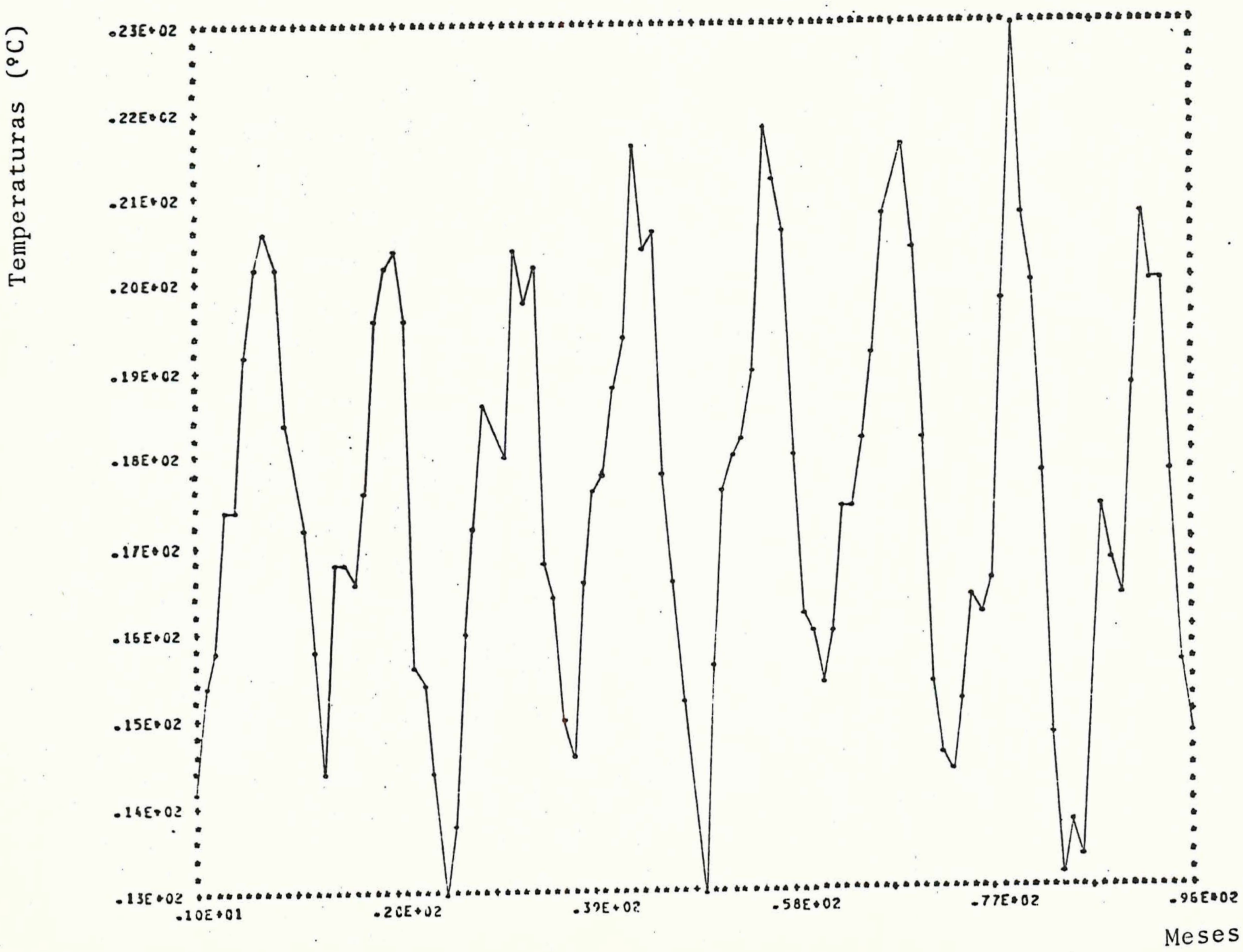



Figura 3.1 - Média mensal de 24 leituras diárias da temperatura de São Paulo

A Figura 3.2 mostra o periodograma da série de temperaturas e nos evidencia a presença de um pico no harmônico 8.

O valor da estatística de Fisher é 0,77, e ocorre na 8ª ordenada.

A entrada mais próxima na Tabela de Fisher (ver Tabela 3.1) é para 50 ordenadas do periodograma, a qual ao nível de 95% de significância é 0,13135.

Então, como $g^{(1)} > 0,13135$, o pico, correspondente a um período de 96/8=12 meses, tem grande significância estatística, o que nos leva a rejeitar a hipótese de que a série consiste de variáveis aleatórias independentes.

Num outro exemplo de aplicação foram utilizados os dados relativos à extinção dos linces canadenses (Lynx Data) apresentados por Moran (1953 a,b), referentes aos anos de 1821 a 1934, dados na Tabela A.2.

A Figura 3.3 mostra o gráfico da série original, onde observa-se a existência de um período de cerca de 9 anos.

A Figura 3.4 apresenta o periodograma da série do "Lynx Data" e confirma a existência de um pico pronunciado no harmônico 12.

O valor da estatística de Fisher é 0,505, e ocorre na 12ª ordenada.

A entrada mais próxima na Tabela de Fisher (por interpolação linear) é para 60 ordenadas, a qual, ao nível de 95% de significância vale 0,12.

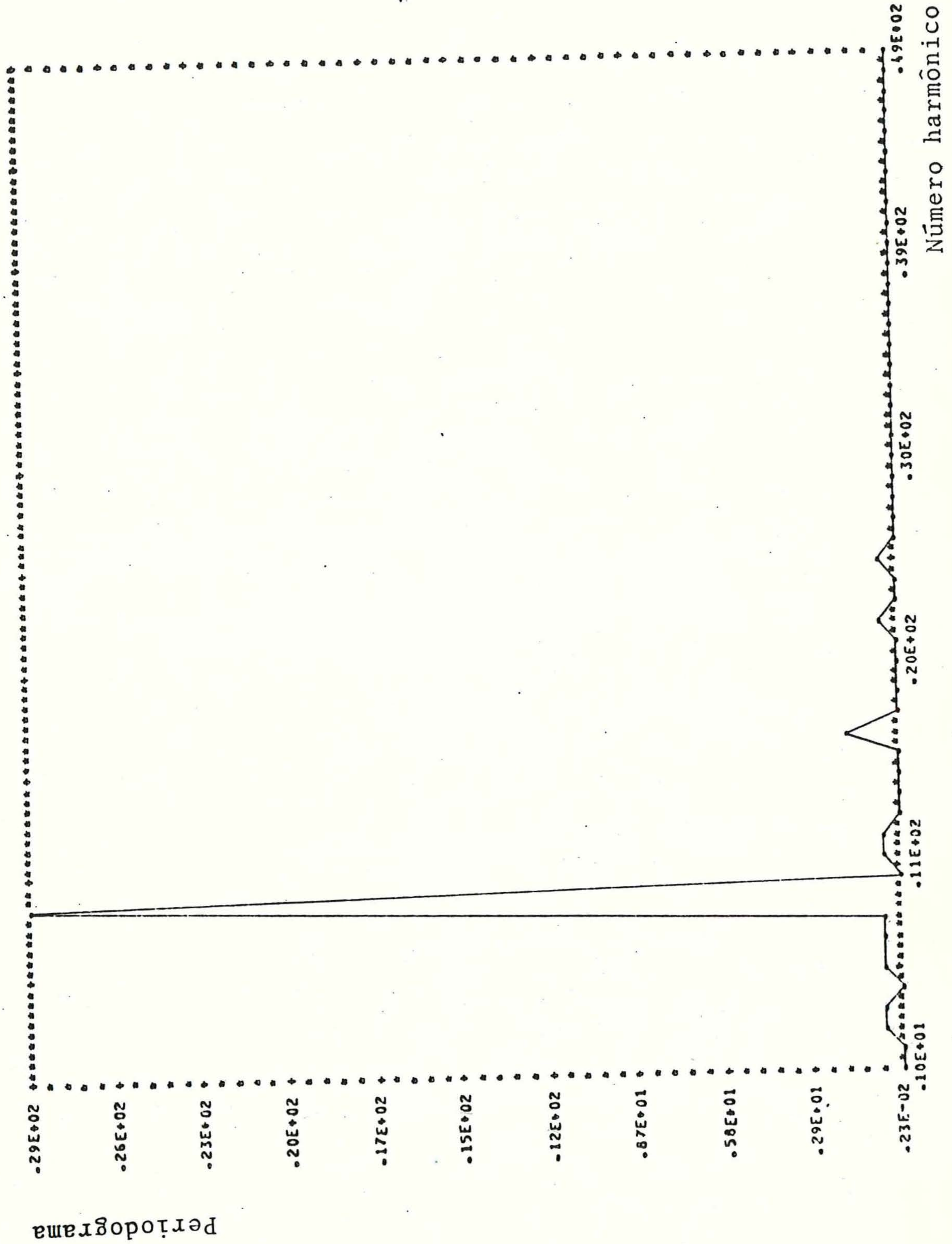



Figura 3.2 - Periodograma da série de temperaturas

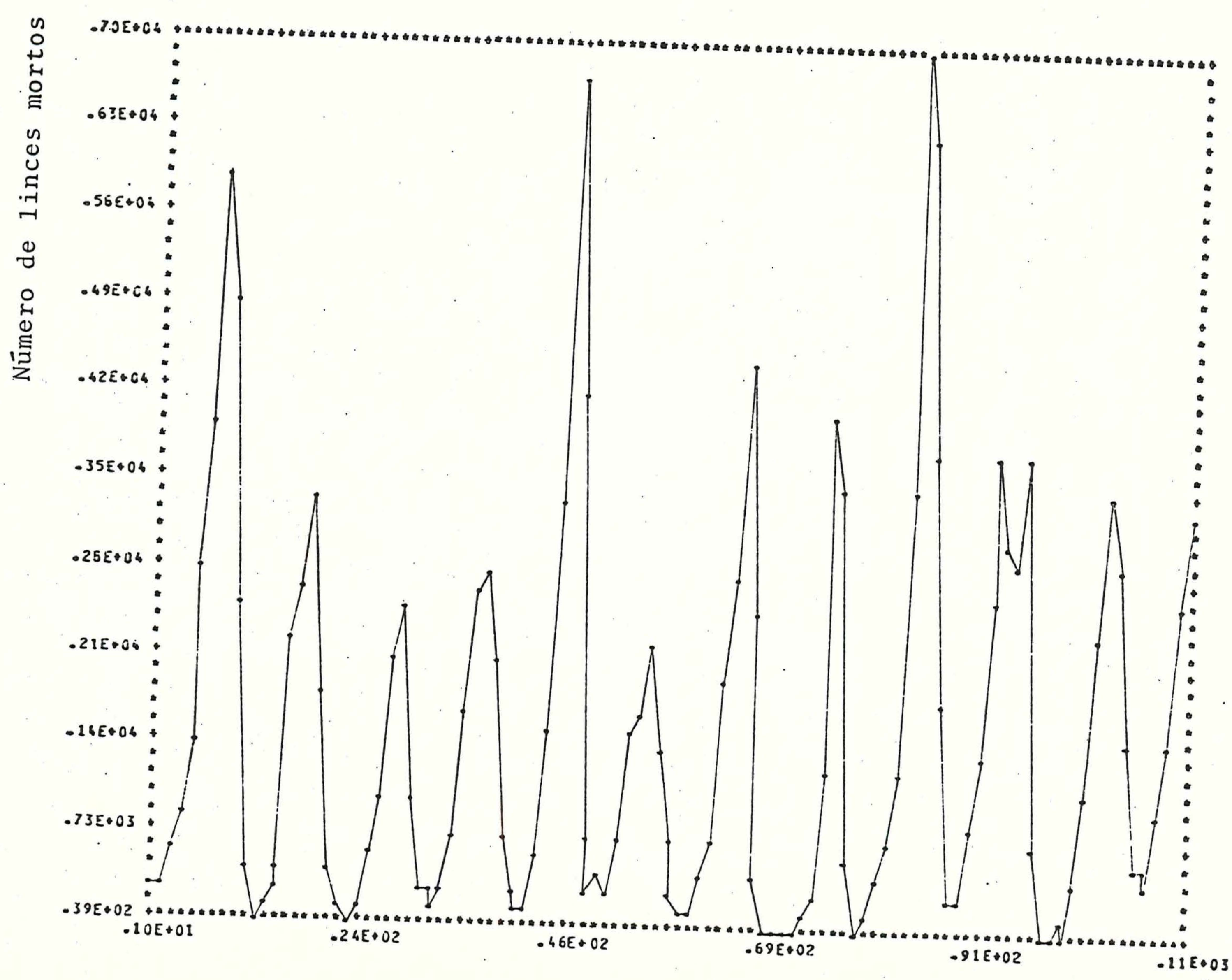
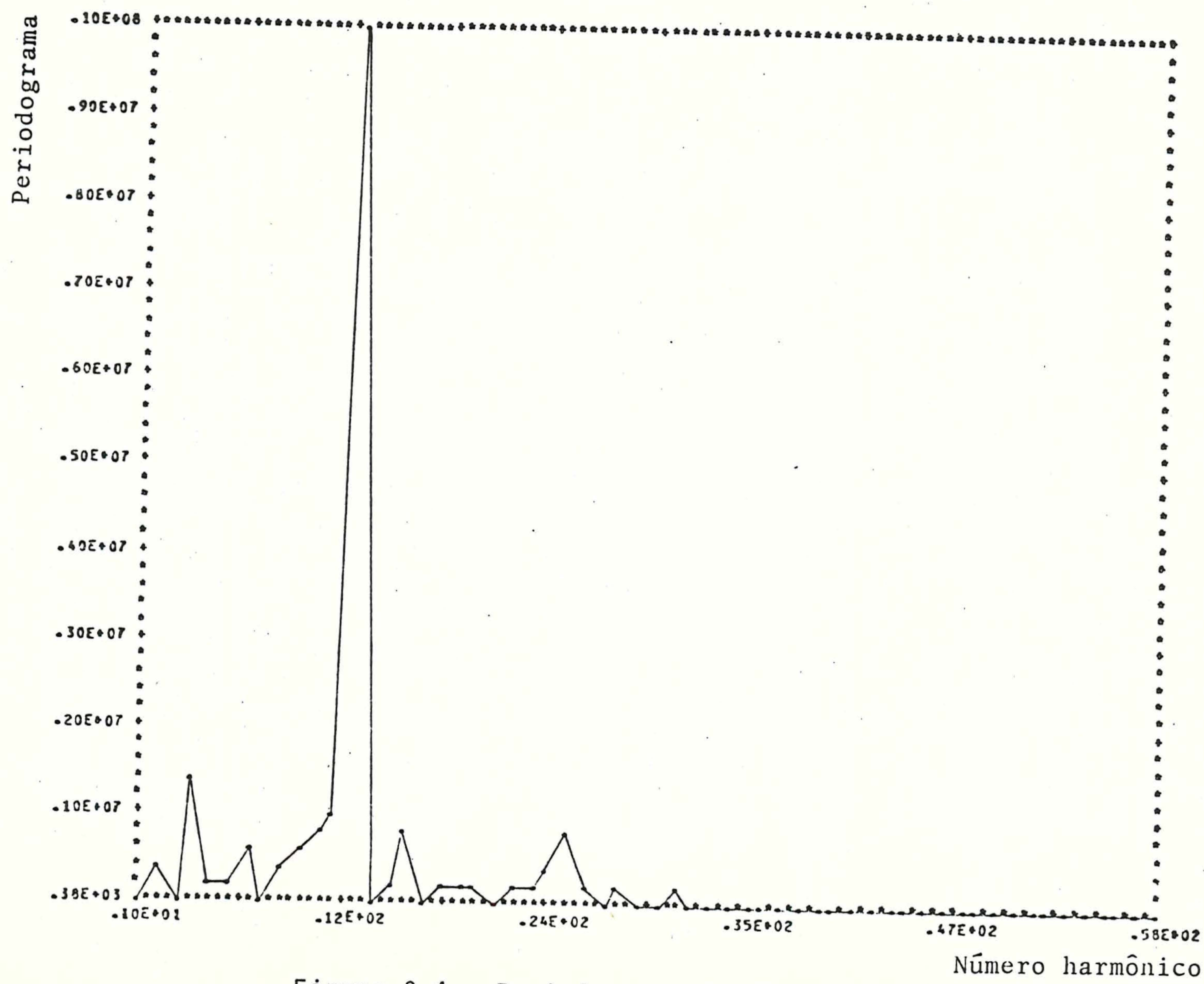



Figura 3.3 - Série original da Lynx Data




<u>Figura</u> <u>3.4</u> - Periodograma do "Lynx Data"

Então, como $g^{(1)}>0,12$, o pico, correspondente a um período de $\frac{114}{12}=9,5$ anos, tem grande significância estatística, nos levando a rejeitar a hipótese de que a série consiste de variáveis aleatórias independentes.

3.4 - Extensão do Teste de Fisher

Whittle (1952) estendeu o teste de Fisher para detectar a existência de r componentes periódicas numa série de tempo.

Sejam $I_1,\ldots,I_m$ as ordenadas do periodograma nas frequências de Fourier de uma série de tempo $X_t$, $t=1,\ldots,n(=2m+1)$, e seja $I^{(r)}$ o r-ésimo maior termo dessa sequência; então podemos obter o teste para a segunda maior ordenada pela omissão do termo $I^{(1)}$ no denominador da estatística $g^{(1)}$, dada por (3.2.4), e substituindo m por (m-1) na distribuição de $g^{(1)}$.

Então, se a maior ordenada $I^{(1)}$ é significante, nós podemos testar a segunda maior ordenada $I^{(2)}$ usando a estatística:

$$g^{(2)}=\frac{I^{(2)}}{\left[\sum\limits_{j=1}^{m} I_j\right]-I^{(1)}}, \tag{3.4.1}$$

e nos referirmos à distribuição $g^{(1)}$ de Fisher com (m-1) graus de liberdade.

Se a segunda maior ordenada for significante nós continuamos com o procedimento acima para testar a terceira maior ordenada, e assim por diante até que não se obtenha um resultado significante.

Dessa maneira, nós podemos detectar a existência de s componentes periódicas numa série de tempo.

Grenander e Rosemblatt (1957) e Hannan (1970), mostraram que a probabilidade $P_r$, de que $g^{(r)}$ exceda o parâmetro g é:

$$P_r = \frac{m!}{(r-1)!} \sum_{j=r}^{k} \frac{(-1)^{j-r}(1-jg)^{m-1}}{j\,(m-j)!\,(j-r)!} \qquad (3.4.2)$$

onde $k=[\frac{1}{g}]$.

Notamos que (3.2.9) é um caso particular de (3.4.2) para r=1, e sua demonstração pode ser obtida de maneira análoga.

Shimshoni (1971) fornece tabelas para vários valores de m e P=0,01 e P=0,05 (ver Tabela 3.2 e 3.3).

Tabela 3.2

Parâmetros de Significância para Componentes Periódicas
P=0,05 e m=5(5)50

| m/r | 1 | 2 | 5 | 7 | 10 |
|---|---|---|---|---|---|
| 5 | 0.68377 | | | | |
| 10 | 0.44495 | 0.26511 | | | |
| 15 | 0.33461 | 0.21016 | 0.10738 | | |
| 20 | 0.27040 | 0.17547 | 0.09559 | 0.07324 | |
| 25 | 0.22805 | 0.15139 | 0.08612 | 0.06768 | 0.05008 |
| 30 | 0.19784 | 0.13360 | 0.07846 | 0.06275 | 0.04777 |
| 35 | 0.17513 | 0.11986 | 0.07215 | 0.05847 | 0.04540 |
| 40 | 0.15738 | 0.10890 | 0.06687 | 0.05475 | 0.04315 |
| 45 | 0.14310 | 0.09993 | 0.06238 | 0.05150 | 0.04108 |
| 50 | 0.13135 | 0.09244 | 0.05851 | 0.04865 | 0.03918 |

Tabela 3.3

Parâmetros de Significância para Componentes Periódicas
P=0,05 e m=100(100)3000

| m/r | 1 | 2 | 5 | 10 | 25 | 50 |
|---|---|---|---|---|---|---|
| 100 | 0.07378 | 0.05425 | 0.03704 | 0.02702 | 0.01584 | 0.00827 |
| 200 | 0.04074 | 0.03098 | 0.02234 | 0.01726 | 0.01150 | 0.00760 |
| 300 | 0.02861 | 0.02211 | 0.01635 | 0.01296 | 0.00909 | 0.00644 |
| 400 | 0.02222 | 0.01735 | 0.01303 | 0.01048 | 0.00757 | 0.00557 |
| 500 | 0.01825 | 0.01435 | 0.01090 | 0.00886 | 0.00652 | 0.00492 |
| 600 | 0.01552 | 0.01228 | 0.00940 | 0.00770 | 0.00575 | 0.00441 |
| 700 | 0.01353 | 0.01075 | 0.00829 | 0.00683 | 0.00516 | 0.00401 |
| 800 | 0.01201 | 0.00958 | 0.00743 | 0.00615 | 0.00469 | 0.00368 |
| 900 | 0.01081 | 0.00865 | 0.00674 | 0.00561 | 0.00431 | 0.00341 |
| 1000 | 0.00984 | 0.00790 | 0.00617 | 0.00516 | 0.00398 | 0.00318 |
| 1100 | 0.00904 | 0.00727 | 0.00570 | 0.00478 | 0.00371 | 0.00298 |
| 1200 | 0.00836 | 0.00674 | 0.00530 | 0.00445 | 0.00348 | 0.00280 |
| 1300 | 0.00778 | 0.00628 | 0.00496 | 0.00417 | 0.00327 | 0.00265 |
| 1400 | 0.00728 | 0.00589 | 0.00466 | 0.00393 | 0.00310 | 0.00252 |
| 1500 | 0.00684 | 0.00554 | 0.00440 | 0.00372 | 0.00294 | 0.00240 |
| 1600 | 0.00645 | 0.00524 | 0.00416 | 0.00353 | 0.00279 | 0.00229 |
| 1700 | 0.00611 | 0.00497 | 0.00395 | 0.00336 | 0.00267 | 0.00219 |
| 1800 | 0.00580 | 0.00472 | 0.00377 | 0.00320 | 0.00255 | 0.00210 |
| 1900 | 0.00553 | 0.00450 | 0.00360 | 0.00306 | 0.00245 | 0.00202 |
| 2000 | 0.00527 | 0.00430 | 0.00344 | 0.00294 | 0.00235 | 0.00195 |
| 2100 | 0.00505 | 0.00412 | 0.00330 | 0.00282 | 0.00226 | 0.00188 |
| 2200 | 0.00484 | 0.00396 | 0.00318 | 0.00271 | 0.00218 | 0.00181 |
| 2300 | 0.00465 | 0.00380 | 0.00306 | 0.00261 | 0.00211 | 0.00175 |
| 2400 | 0.00447 | 0.00366 | 0.00295 | 0.00252 | 0.00204 | 0.00170 |
| 2500 | 0.00431 | 0.00353 | 0.00285 | 0.00244 | 0.00197 | 0.00165 |
| 2600 | 0.00416 | 0.00341 | 0.00275 | 0.00236 | 0.00191 | 0.00160 |
| 2700 | 0.00402 | 0.00330 | 0.00267 | 0.00229 | 0.00186 | 0.00155 |
| 2800 | 0.00389 | 0.00320 | 0.00258 | 0.00222 | 0.00180 | 0.00151 |
| 2900 | 0.00377 | 0.00310 | 0.00251 | 0.00216 | 0.00175 | 0.00147 |
| 3000 | 0.00365 | 0.00301 | 0.00243 | 0.00210 | 0.00171 | 0.00144 |

A distribuição dada por (3.4.2) nos dá um teste apropriado para o caso em que nós sabemos que se a hipótese nula é falsa, $X_t$ deve conter exatamente s componentes periódicas, sendo s conhecido a priori.

Entretanto, se o número de termos periódicos é desconhecido, nós não podemos simplesmente aplicar o teste sequencialmente, desde que a distribuição dada por (3.4.2) permanece válida somente sob a hipótese nula e se a máxima ordenada do periodograma é significante, a hipótese nula para a r-ésima maior ordenada do periodograma $(r>1)$ não é mais apropriada.

## 3.5 - Aplicação

Como um exemplo de aplicação do teste descrito na seção anterior, consideremos os dados da Tabela A.3, referentes à série de precipitação anual de chuvas em Fortaleza, Ceará, no período de 1849 a 1979, e vamos testar a existência de um certo número de componentes periódicas usando a proposta de Whittle.

A Figura 3.5 mostra o gráfico da série original, para a qual estudaremos a existência de tais componentes.

A Figura 3.6 apresenta o periodograma desta série.

O valor da estatística $g^{(1)}$ é 0,136, e a entrada na Tabela de Fisher (por interpolação linear) para m=65, ao nível de 95%, é 0,0961.

Então, como $g^{(1)}>0{,}0961$, o pico correspondente a um período de 131/10=13,1 anos é significante, e nos leva a rejeitar a hipótese nula de que a série consiste de variáveis aleatórias independentes.

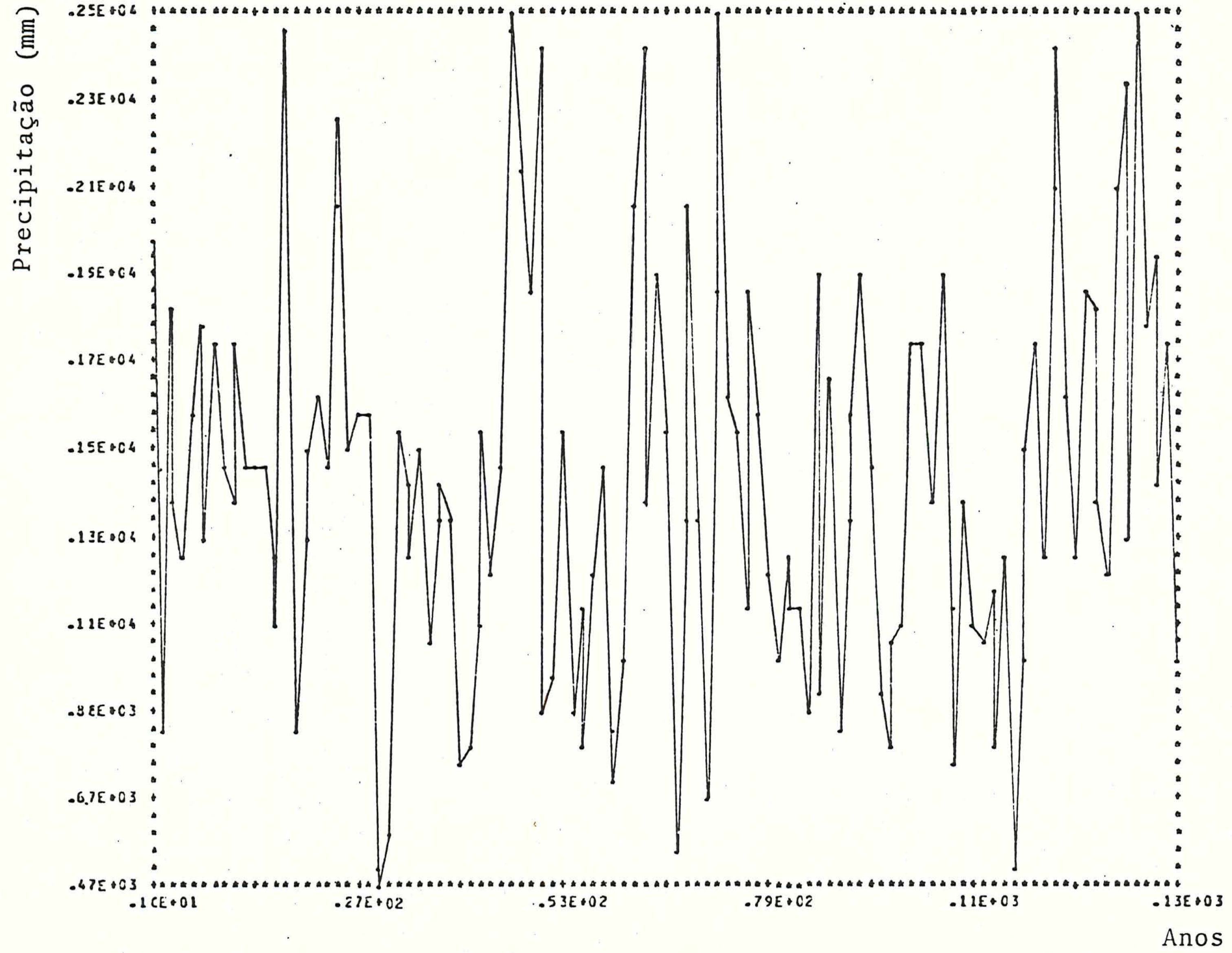



Figura 3.5 - Série de precipitações de chuvas em Fortaleza

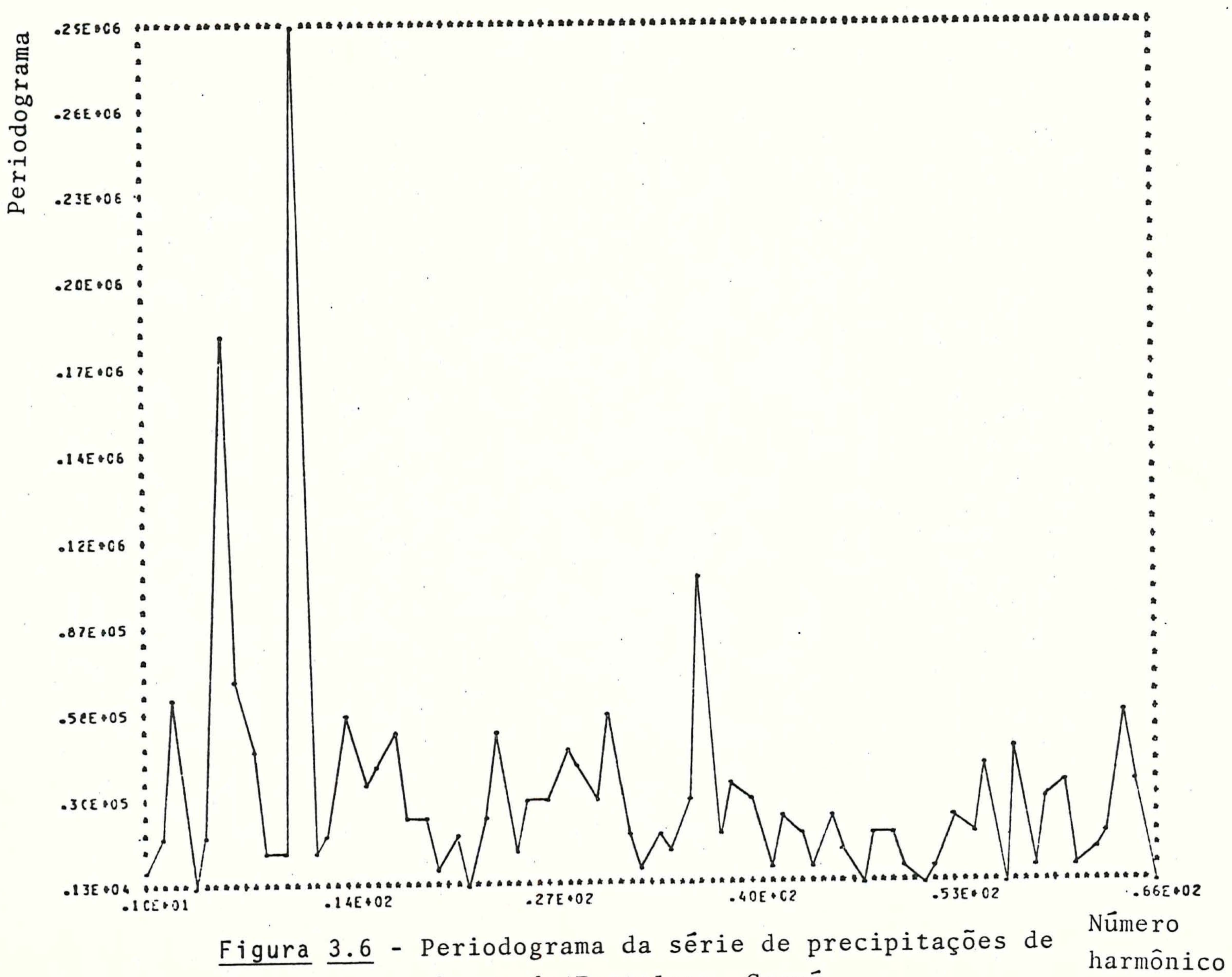



Figura 3.6 - Periodograma da série de precipitações de chuvas de Fortaleza, Ceará

Realizando o teste para a segunda maior ordenada observamos que $g^{(2)}=0,1012$.

A entrada na Tabela de Fisher para m=64, ao nível de 95%, é 0,0984, e como $g^{(2)}>0,0984$, o pico correspondente a um período de 131/5=26,2 anos é significante.

Realizando o teste para a terceira maior ordenada, observamos que $g^{(3)}=0,0623$.

A entrada na Tabela de Fisher para m=63, ao nível de 95%, vale 0,1008.

Como $g^{(3)}<0,1008$, rejeitamos a existência de um período de 131/36=3,64 anos.

Logo, de acordo com este teste, a série de Precipitações de chuvas de Fortaleza, Ceará, apresenta dois períodos, de 13,1 anos e 26,2 anos.

<u>Observação</u>: Um método alternativo proposto por Shimshoni (1971) consiste em testar sequencialmente as ordenadas do periodograma pela ordem de grandeza, sendo m fixo e r variável em função da ordenada a ser testada.

Um fato que foi possível observar, através da comparação dos dois métodos, é que o procedimento de Shimshoni aceita picos rejeitados pelo método de Whittle acima descrito.

# CAPÍTULO 4

# ANÁLISE DE ESPECTROS MISTOS

## 4.1 - Introdução

Neste capítulo discutiremos o problema de separar componentes espectrais discretas e contínuas.

O pioneiro em tratar da análise espectral foi Schuster (1898), cujo método da análise do periodograma é aplicável a espectros puramente discretos.

Posteriormente, vários métodos foram propostos ao tratar com espectros puramente contínuos, que permitissem uma análise modificada do periodograma e a estimação da função densidade espectral, dentre os quais temos Daniell(1946), Bartlett (1950), Grenander e Rosemblatt (1957) e Parzen (1957,1958).

Entretanto, a estimação da função densidade espectral é apenas um aspecto da análise espectral, pois um processo estacionário geralmente tem um espectro misto, isto é, contém componentes discretas e contínuas; e os métodos considerados pelos autores citados acima são inadequados nestas situações.

Neste capítulo, nós sugerimos dois métodos para separar e estimar componentes discretas e contínuas, sem conhecimento prévio da frequência ou amplitude das componentes harmônicas, as quais constituem o espectro discreto, mas assumindo somente que a largura de faixa da função densidade espectral é conhecida ou tem um limite inferior conhecido.

Um processo estacionário com espectro misto pode ser descrito pelo seguinte modelo:

$$X_t = Y_t + Z_t \tag{4.1.1}$$

onde $X_t$ é o processo observado, $Y_t$ é o processo estacionário com função densidade espectral absolutamente contínua e $Z_t$ é o processo estacionário com espectro discreto.

Assumimos que $Y_t$ é um processo linear da forma:

$$Y_t = \sum_{u=0}^{\infty} g_u \varepsilon_{t-u} \tag{4.1.2}$$

onde $\varepsilon_t$ é o processo ruído branco dado pela Definição 2.2.4.

Para garantir a estacionariedade do processo, os coeficientes $\{g_u\}$ devem satisfazer:

$$\sum_{u=0}^{\infty} g_u^2 < \infty \quad \text{e} \quad \sum_{u=0}^{\infty} u|g_u| < \infty \tag{4.1.3}$$

ou (equivalentemente):

$$g_u = 0\left(\frac{1}{|u|^k}\right), \quad k>2 \tag{4.1.4}$$

Seja $\gamma_Y(s)$ a função de autocovariância de $Y_t$:

$$\gamma_Y(s) = E[Y_t Y_{t+|s|}] \tag{4.1.5}$$

e f(ω) a sua função densidade espectral:

$$f_Y(\omega) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \gamma_Y(s)e^{-i\omega s}, \quad -\pi \leqslant \omega \leqslant \pi \tag{4.1.6}$$

O processo $Z_t$ é a soma de k componentes harmônicas, da forma:

$$Z_t = \sum_{i=1}^{k} A_i \cos(\omega_i t+\phi_i), \tag{4.1.7}$$

onde $\{\phi_i\}$ são independentes e uniformemente distribuídos em $(-\pi,\pi)$ e $\{A_i\}$ e $\{\omega_i\}$, i=1,...,k são constantes desconhecidas.

Seja $\gamma_Z(s)$ a função de autocovariância de $Z_t$, dada por:

$$\gamma_Z(s) = E[Z_t Z_{t+|s|}] = \frac{1}{2} \sum_{i=1}^{k} A_i^2 \cos(s\omega_i), \tag{4.1.8}$$

Finalmente, $\gamma_X(s)=E[X_t X_{t+|s|}]$ denota a função de autocovariância de $X_t$.

## 4.2 - O Teste de Whittle

Dado o conjunto de observações $X_1,\ldots,X_n$, definimos o periodograma (ligeiramente modificado em relação a (2.3.3)), por:

$$I_X(\omega) = \frac{2}{n}\left|\sum_{t=1}^{n} X_t e^{-i\omega t}\right|^2 \tag{4.2.1}$$

e $I_Y(\omega)$ e $I_\varepsilon(\omega)$ sendo definidos similarmente para $Y_t$ e $\varepsilon_t$.

Sabemos que:

$$I_Y(\omega) \sim 2\pi f_Y(\omega) I_\varepsilon(\omega) \tag{4.2.2}$$

(ver Bartlett, 1955, pg. 279).

Então, para n grande e de (4.2.2) temos que $\frac{I_Y(\omega)}{2\pi f_Y(\omega)}$ pode ser considerado como o periodograma do processo ruído branco, e a existência de componentes harmônicas pode ser testada aplicando o teste de Fisher para o periodograma normalizado $\frac{I_X(\omega)}{2\pi f_Y(\omega)}$.

Se $f_Y(\omega)$ fosse uma função conhecida, a maneira natural de tratar o espectro misto seria calcular as ordenadas de $\frac{I_X(\omega)}{2\pi f_Y(\omega)}$ nas frequências de Fourier $\omega_j$ e considerar a estatística g, dada por:

$$g = \frac{\max\limits_{j=1,\ldots,m} [I_X(\omega_j)/2\pi f_Y(\omega_j)]}{\sum\limits_{j=1}^{m} [I_X(\omega_j)/2\pi f_Y(\omega_j)]} \qquad (4.2.3)$$

Mas em geral $f_Y(\omega)$ é desconhecida e para utilizar a estatística g é necessário estimar $f_Y(\omega)$ do processo observado $X_t$.

Whittle (1952) sugere o estimador do periodograma truncado, dado por:

$$\hat{f}_Y(\omega_j) = \frac{1}{2\pi} \sum_{s=-(\ell-1)}^{\ell-1} C_X(s) \cos(\omega_j s), \quad \ell<n, \qquad (4.2.4)$$

onde

$$C_X(s) = \frac{1}{n} \sum_{t=1}^{n-|s|} X_t X_{t+|s|}$$

é a auto-covariância amostral de lag s.

Substituindo (4.2.4) em (4.2.3), temos:

$$g = \frac{\max\limits_{j=1,\ldots,m} I_X(\omega_j)/2\pi\hat{f}_Y(\omega_j)}{\sum\limits_{j=1}^{m} [I_X(\omega_j)/2\pi\hat{f}_Y(\omega_j)]} \tag{4.2.5}$$

Sob a hipótese nula de que o espectro é puramente contínuo, isto é, que $A_i=0$, $\forall i$, g ainda é assintóticamente distribuida como a estatística de Fisher com m graus de liberdade.

Entretanto, se rejeitarmos a hipótese nula, existe uma componente harmônica na frequência $\omega_i$, que acarreta uma contribuição à $E[\hat{f}_Y(\omega_i)]$.

Para determinar essa contribuição vamos calcular $E[\hat{f}_Y(\omega_i)]$.

De (4.2.4), temos:

$$E[\hat{f}_Y(\omega_i)] = E[\frac{1}{2\pi} \sum_{s=-(\ell-1)}^{\ell-1} C_X(s) \cos(s\omega_i)] =$$

$$= E[\frac{1}{2\pi} \sum_{s=-(\ell-1)}^{\ell-1} (\frac{1}{n} \sum_{t=1}^{n-|s|} X_t X_{t+|s|}) \cos(s\omega_i)] =$$

$$= \frac{1}{2\pi} \sum_{s=-(\ell-1)}^{\ell-1} \cos(s\omega_i) . \frac{1}{n} E[\sum_{t=1}^{n-|s|} X_t X_{t+|s|}]$$

Substituindo (4.1.1) temos:

$$= \frac{1}{2\pi} \sum_{s=-(\ell-1)}^{\ell-1} \cos(s\omega_i) \frac{1}{n} E[\sum_{t=1}^{n-|s|} (Y_t+Z_t)(Y_{t+|s|}+Z_{t+|s|})]$$

$$= \frac{1}{2\pi} \sum_{s=-(\ell-1)}^{\ell-1} \cos(s\omega_i) . \frac{1}{n} E[\sum_{t=1}^{n-|s|} Y_t Y_{t+|s|} + \sum_{t=1}^{n-|s|} Z_t Z_{t+|s|}]$$

$$= \frac{1}{2\pi} \sum_{s=-(\ell-1)}^{\ell-1} \gamma_Y(s) \cos(s\omega_i) + \frac{1}{2\pi} \sum_{s=-(\ell-1)}^{\ell-1} \frac{1}{n} \sum_{t=1}^{n-|s|} \sum_{j=1}^{k} \frac{1}{2} A_j^2 \cos^2(s\omega_j)$$

Como $\gamma_Y(s)$ tende a zero para lags muito afastados, temos:

$$E[\hat{f}_Y(\omega_i)] \cong f_Y(\omega_i) + \frac{1}{2\pi} \sum_{s=-(\ell-1)}^{\ell-1} \frac{1}{n} \sum_{t=1}^{n-|s|} \frac{1}{2} A_i^2 \cos^2(s\omega_i),$$

pois se rejeitarmos a hipótese nula, existe uma componente harmônica na frequência $\omega_i$, com amplitude $A_i \neq 0$.

$$= f_Y(\omega_i) + \frac{1}{4\pi} \sum_{s=-(\ell-1)}^{\ell-1} \frac{n-|s|}{n} A_i^2 \cos^2(s\omega_i)$$

Passando ao limite temos:

$$\lim_{n\to\infty} E[\hat{f}_Y(\omega_i)] = f_Y(\omega_i) + \lim_{n\to\infty} \frac{1}{4\pi} \sum_{s=-(\ell-1)}^{\ell-1} \frac{n-|s|}{n} A_i^2 \cos^2(s\omega_i)$$

$$= f_Y(\omega_i) + \frac{1}{4\pi} A_i^2 \sum_{s=-(\ell-1)}^{\ell-1} \cos^2(s\omega_i) =$$

$$= f_Y(\omega_i) + \frac{1}{4\pi} \ell A_i^2.$$

Ou seja, $E[\hat{f}_Y(\omega_i)] \simeq f_Y(\omega_i) + \frac{1}{4\pi} \ell A_i^2.$

Isto significa que a contribuição da frequência $\omega_i$ à $E[\hat{f}_Y(\omega_i)]$ é $\frac{1}{4\pi} \ell A_i^2$.

Então, o efeito de estimar $f_Y(\omega)$ por $\hat{f}_Y(\omega)$ é reduzir a altura da maior ordenada do periodograma pelo fator $O(\ell)$ e, consequentemente, diminuir o poder do teste de significância baseado em $g^{(1)}$.

Para evitar isto, Whittle (1952) sugere um estimador modificado de $f_Y(\omega)$, denotado por $\tilde{f}_Y(\omega)$, obtido a partir de $\hat{f}_Y(\omega)$ pela omissão de todos os picos de $\hat{f}_Y(\omega)$, os quais suspeitamos que sejam devidos à presença de componentes harmônicas.

Este procedimento acarreta um grande risco ao omitir picos em $\hat{f}_Y(\omega)$ e assumir que eles são devidos à termos harmônicos, antes que a existência de tais picos seja detectada pelo teste.

A maior dificuldade nesta abordagem é em distinguir picos no espectro contínuo e picos devido às componentes harmônicas.

## 4.3 - O Teste de Bartlett

Um teste alternativo sugerido por Bartlett (1957) se baseia na separação de picos espectrais em função de sua largura de faixa.

Segundo Bartlett, dado um número finito de observações e nenhuma outra informação, é difícil, qualquer que seja o método, distinguir entre picos devidos à componentes harmônicas e picos do espectro contínuo cuja largura seja arbitrariamente estreita. Esta separação depende da razão ruído/sinal.

A largura de faixa de $f_Y(\omega)$ pode ser definida como a largura do pico mais estreito, e desde que a base do teste do periodograma para componentes harmônicas está no fato de que elas podem produzir picos no periodograma, é claro que, dada uma amostra de n observações, é impossível distinguir entre componentes harmônicas estritas e picos do espectro contínuo cujas larguras são da $O(\frac{1}{n})$.

Então, para tornar o problema tratável nós faremos a suposição de que o espectro contínuo $f_Y(\omega)$ não tenha nenhum pico cuja largura seja menor que k', onde $k' \gg \frac{1}{n}$, e cuja largura de faixa seja B.

Então o teste pode ser obtido agrupando as ordenadas do periodograma.

Seja $k=[k']$ e dividimos as $\frac{n}{2}$ ordenadas do periodograma em $[\frac{n}{2k}]$ conjuntos, cada um com k ordenadas, da seguinte maneira:

$$I_1,\ldots,I_k;I_{k+1},\ldots,I_{2k};\ldots;I_{[\frac{n}{2k}]-k+1},\ldots,I_{[\frac{n}{2k}]}.$$

Seja $T_k$ dado por:

$$T_k = \frac{I_{p'}/2\pi f_Y(\omega_{p'})}{\sum\limits_{p=(\ell-1)k+1}^{\ell k} I_p/2\pi f_Y(\omega_p)}, \quad \ell = 1,\ldots,[\frac{n}{2k}].\frac{1}{k} \qquad (4.3.1)$$

onde

$$\frac{I_{p'}}{2\pi f_Y(\omega_{p'})} = \max_{(\ell-1)k+1<p<\ell k} I_p/2\pi f_Y(\omega_p)$$

A estatística $T_k$ tem assintóticamente a mesma distribuição que a estatística $g^{(1)}$ de Fisher com k graus de liberdade e pode ser usada como base do teste para o espectro misto, levando em conta o fato de termos restringido a largura de faixa de $f_Y(\omega)$, o que torna $f(\omega)$ aproximadamente constante na região considerada.

Então $T_k$ pode ser aproximada por:

$$g_k^{(B)} = \frac{I_{p'}}{\sum_p I_p} \qquad (4.3.2)$$

Então, mesmo quando $f_Y(\omega)$ é desconhecida, o teste pode ser realizado considerando $g_k^{(B)}$ como a estatística de Fisher com k graus de liberdade.

Sejam M' e m' os limites superiores e inferiores de $f(\omega)$ na região considerada, e como já assumimos que $f(\omega)$ não tem picos cujas larguras sejam menores que k', então $\frac{M'}{m'} \leqslant 2$, logo:

$$T_k\left(2 - \frac{M'}{m'}\right) \leqslant g_k^{(B)} \leqslant T_k \cdot \frac{M'}{m'} \qquad (4.3.3)$$

Mesmo assim $g_k^{(B)}$ pode diferir consideravelmente de $T_k$, podendo-se chegar a um $g_k^{(B)}$ quase duas vezes maior que $T_k$.

Para obter uma boa aproximação de $T_k$ devemos considerar a escolha de k.

A maneira de obter uma boa aproximação para $T_k$ seria tomar k muito menor que a largura de faixa de $f_Y(\omega)$ e com isso reduzir a razão $\frac{M'}{m'}$.

Entretanto, a menos que $f_Y(\omega)$ tenha uma largura de faixa grande, este procedimento nos levaria a um $g_k^{(B)}$ com um número muito pequeno de graus de liberdade.

Então, não parece haver qualquer critério sistemático para a escolha de k que consiga reter graus de liberdade suficientes para dar razoável poder ao teste e que reduza a razão $\frac{M'}{m'}$.

Mesmo assumindo que nós pudessemos obter um grau de precisão suficiente para testar $g_k^{(B)}$ usando $T_k$, restaria ajustar o nível de significância do teste de modo a proceder a escolha do subintervalo cujas ordenadas estão sendo testadas.

Se nós escolhessemos um nível de significância $\alpha$ para o teste original de Fisher, então o nível de significância apropriado para o teste baseado em $g_k^{(B)}$ seria $\alpha'=\alpha k/[\frac{n}{2}]$.

## 4.4 - O Teste de Hannan

Hannan (1961) considera o problema de testar se uma componente periódica corresponde a um salto na função distribuição espectral $F(\lambda)$.

A diferença substancial neste método é o uso de um estimador suavizado de $f(\lambda)$ para detectar picos no espectro.

Testaremos a hipótese nula de que $F(\lambda)$ é absolutamente contínua, com derivada $f(\lambda)$, a qual é uma função relativamente suave, contra a hipótese alternativa de que $F(\lambda)$ tem um salto.

Seja $X_t$, $t=1,\dots,n$ um processo linear dado por (4.1.1).

Se $f(\omega)$ é conhecida a priori e não tem zeros em $[-\pi,\pi]$, nós podemos formar as quantidades $k_n(\lambda_j)$, dadas por:

$$k(\lambda_j) = \frac{I(\lambda_j)}{2\pi f(\lambda_j)}, \quad j=1,\dots,m \tag{4.4.1}$$

Neste caso o teste será dado por:

$$S = \frac{\max\limits_{j=1,\ldots,m} k(\lambda_j)}{\sum\limits_{j=1}^{m} k(\lambda_j)} \tag{4.4.2}$$

Mas, em geral, $f(\lambda)$ não é conhecida e precisamos estimá-la.

Entretanto, sob a hipótese alternativa, para $\lambda$ na vizinhança do ponto de salto de $F(\lambda)$, este procedimento pode nos levar a um estimador de $f(\lambda)$ que é afetado pela presença do salto, cuja existência queremos testar.

Seja $\hat{f}(\lambda)$ um estimador suavizado de $f(\lambda)$, dado por:

$$\hat{f}(\lambda) = \int_{-\pi}^{\pi} W(\lambda-\theta) I(\theta) d\theta, \tag{4.4.3}$$

onde $W(\lambda-\theta)$ é uma janela espectral escolhida adequadamente.

Substituindo (4.4.3) em (4.4.1) temos:

$$\hat{k}(\lambda_j) = \frac{I(\lambda_j)}{2\pi \hat{f}(\lambda_j)}, \quad j=1,\ldots,m \tag{4.4.4}$$

Então, a estatística do teste será dada por:

$$\hat{S} = \frac{\max\limits_{j=1,\ldots,m} \hat{k}(\lambda_j)}{\sum\limits_{j=1}^{m} \hat{k}(\lambda_j)} \tag{4.4.5}$$

que tem aproximadamente a distribuição de Fisher com m graus de liberdade.

Outra maneira de estimar $f(\lambda)$ seria aplicar uma regressão em cada frequência $\lambda_j$, para todo $j$, antes de estimar $f(\lambda_j)$, que denotaremos $f^*(\lambda_j)$; mas o problema não será considerado dessa forma, pois o número de regressões seria proibitivo.

Neste caso (4.4.5) seria dada por:

$$S^* = \frac{\max\limits_{j=1,\ldots,m} k^*(\lambda_j)}{\sum\limits_{j=1}^{m} k^*(\lambda_j)} \tag{4.4.6}$$

Para n grande verificamos que:

$$\frac{\sum\limits_{j=1}^{m} k^*(\lambda_j)}{n} \xrightarrow[n\uparrow\infty]{} 1$$

com probabilidade 1.

Então (4.4.6) poderia ser escrita como:

$$S^* = \frac{\max\limits_{j=1,\ldots,m} k^*(\lambda_j)}{n} \tag{4.4.7}$$

Para evitar a influência do salto em $\hat{f}(\lambda)$, cuja existência queremos testar, Hannan (1961) sugere que $\hat{f}(\lambda_j)$ seja substituido por:

$$\tilde{f}(\lambda_j) = \frac{\hat{f}(\lambda_j) - \frac{2\pi}{n} W(0) I(\lambda_j)}{1 - \frac{8\pi^2}{n} W(0)}, \tag{4.4.8}$$

onde $W(0)$ é o valor da janela espectral na frequência zero.

Com isto o teste se torna satisfatório quando a hipótese nula não é verdadeira, devido à presença do pico, o qual será induzido em $\hat{f}(\lambda)$, tornando o estimador $\tilde{f}(\lambda)$ aproximadamente não viciado.

## 4.5 - Desvantagens dos Métodos Apresentados

Resumimos a seguir alguns aspectos que tornam insatisfatória a utilização do teste de Whittle, de Bartlett e de Hannan.

i) O teste de Whittle fica extremamente limitado ao exigir a estimação da função densidade espectral.

ii) Pode-se demonstrar que quaisquer componentes harmônicas presentes produzem picos na função estimada, os quais, se não forem removidos reduzem o poder do teste de tal maneira que ele não detectaria a existência de componentes harmônicas mesmo para amplitudes razoavelmente elevadas.

Por outro lado, a remoção destes picos antes do teste ser aplicado equivale à afirmação de que os picos são devidos à componentes harmônicas antes que o teste, cujo principal objetivo é detectá-las, seja aplicado.

iii) O teste de Bartlett em geral é mais poderoso que o teste de Whittle, mas apresenta grandes dificuldades na escolha de k;

iv) Há um compromisso entre reduzir $\frac{M'}{m'}$ e reter graus de liberdade suficientes;

v) Mesmo após a escolha de k, o erro ao substituir $T_k$ por $g_k^{(B)}$ pode ser extremamente sério.

vi) No teste de Hannan, o vício do estimador de $f(\lambda)$ nos leva a uma redução na estatística $k(\lambda_j)$ e, consequentemente, a um teste de poder menor.

vii) O salto pode ocorrer entre duas frequências $\lambda_j = \frac{2\pi j}{n}$ e $\lambda_k = \frac{2\pi k}{n}$, $j \neq k$, e com isso não seria possível remover o seu efeito na estimação de $f(\lambda)$, nos impedindo de distinguir entre um salto em $F(\lambda)$ e um pico agudo em $f(\lambda)$.

## 4.6 - Aplicações

Para verificar a validade dos três métodos apresentados, faremos uma aplicação às séries de temperaturas, "Lynx Data" e precipitações de chuvas, apresentadas no Apêndice.

### A) Teste de Whittle

O teste de Whittle usa a estatística

$$g = \frac{\max\limits_{j=1,\dots,m} [I_X(\omega_j)/2\pi\hat{f}_Y(\omega_j)]}{\sum\limits_{j=1}^{m} [I_X(\omega_j)/2\pi\hat{f}_Y(\omega_j)]}$$

Série de temperaturas

A máxima ordenada do periodograma normalizado ocorre em $j=8$ ($\omega_8 = \frac{2\pi\times 8}{96} = \pi/6$), então:

$$g = \frac{[I_8/2\pi\hat{f}_Y(\omega_8)]}{\sum\limits_{j=1}^{48} [I_j/2\pi\hat{f}_Y(\omega_j)]} = 0{,}823$$

Podemos então nos referir a g como a distribuição de

Fisher como $\frac{n}{2} = 48$ graus de liberdade, e ao nível de significância $\alpha$=5% temos que o valor crítico é 0,13135, e representa que o resultado é significante ao nível de 5%.

Logo, rejeitamos a hipótese nula de que o espectro é puramente contínuo, e concluímos que existe uma componente harmônica na frequência $\omega_8$, cujo período é 12 meses.

Série do "Lynx Data"

A máxima ordenada do periodograma normalizado ocorre em $j=12$ $(\omega_{12} = \frac{2\pi \times 12}{114} \simeq \frac{\pi}{5})$, logo g=0,520.

Então, ao nível de significância $\alpha$=5%, o valor crítico é 0,12, e como g>0,12 rejeitamos a hipótese nula de que o espectro é puramente contínuo, concluindo-se a existência de uma componente periódica, com período de 9,5 anos.

Série de Precipitações

A máxima ordenada do periodograma normalizado ocorre em $j=10$ $(\omega_{10} = \frac{2\pi \times 10}{131} \simeq \frac{\pi}{6})$, então g=0,145.

Logo ao nível de significância $\alpha$=5%, o valor crítico é g>0,0964 e como g>0,13135 aceitamos a existência de uma componente periódica com período 13,1 anos.

A segunda maior ordenada do periodograma normalizado ocorre em j=5, e o valor do teste é g=0,092.

Ao nível $\alpha$=5%, o valor crítico é 0,0699, logo como g>>0,0699 verificamos a existência de um período secundário de 26,2 anos.

A terceira maior ordenada ocorre em j=36, e o valor de g é 0,056.

O valor crítico é 0,061 e como g<0,061, rejeito a existência de um terceiro período de 3,64 anos.

B) Teste de Bartlett

Aplicamos agora o teste de Bartlett, ou do periodograma agrupado, para cada série, com k=5,10,20.

A estatística do teste é:

$$g_k^{(B)} = \max_k I_{p'} / \sum_k I_p$$

Os valores obtidos para $g_k^{(B)}$ são dados na Tabela 4.1.

Escolhendo um nível de significância $\alpha$ ao aplicar o teste de Fisher para k ordenadas, obtemos um nível de significância aproximado para $g_k^{(B)}$, dado por $\alpha' = \alpha k / [\frac{n}{2}]$.

Tomando $\alpha = 5\%$ e usando apenas o primeiro termo de (3.2.9), os valores críticos de $g_k^{(B)}$ são, aproximadamente, dados por:

$$\frac{0,05 \times k}{[\frac{n}{2}]} = k(1-g)^{k-1}$$

e se encontram na Tabela 4.1.

O valor $g_5^{(B)}$ é significante ao nível de 5% para as séries de temperaturas e Lynx Data, enquanto que os valores de $g_{10}^{(B)}$ e $g_{20}^{(B)}$ são altamente significantes para ambas as séries, fato que deve ser analisado com cautela desde que sua validade requer que a função densidade espectral seja constante numa região muito grande.

Tabela 4.1

Teste do Periodograma Agrupado

| Série | k | $g_k^{(B)}$ | g |
|---|---|---|---|
| Temperatura | 5 | 0,9517 | 0,8203 |
| | 10 | 0,8997 | 0,5337 |
| | 20 | 0,8044 | 0,3033 |
| Lynx Data | 5 | 0,8391 | 0,8279 |
| | 10 | 0,7508 | 0,5425 |
| | 20 | 0,5746 | 0,1917 |
| Precipitações | 5 | 0,5657 | 0,8335 |
| | 10 | 0,4286 | 0,5492 |
| | 20 | 0,2893 | 0,3143 |

A maior ordenada do periodograma para a série de temperatura e Lynx Data ocorrem em j=8 e j=12, respectivamente, acarretando a existência de um período de 12 meses e outro de 9,5 anos, respectivamente.

Portanto, pela natureza geral dos resultados apresentados para a série de temperaturas e Lynx Data, concluimos que os dados tem a estrutura de espectro misto.

O mesmo não ocorre para a série de precipitações, onde $g_k^{(B)}$ é inferior ao valor crítico para k=5,10 e 20, nos levando a concluir que eventualmente a estrutura da série não é a do espectro misto. Isto pode ser explicado pelo fato da série apresentar vários picos, fazendo com que a estatística $g_k^{(B)}$ não assuma valores muito altos.

Concluimos então, que o teste de Bartlett se revela mais poderoso para série que contém uma componente periódica.

Mas, para séries que contêm duas ou mais componentes seu poder cai sensivelmente, podendo não detectar a existência mesmo da primeira.

C) Teste de Hannan

O teste de Hannan usa a estatística

$$\tilde{S} = \frac{\max\limits_{j=1,\dots,m} \tilde{k}(\lambda_j)}{\sum\limits_{j=1}^{m} \tilde{k}(\lambda_j)}$$

Série de Temperaturas

A maior ordenada ocorre em $j=8$, resultando $\tilde{S}=0,823$, e ao nível de significância de 5% o valor crítico é 0,1313, resultando numa rejeição da hipótese nula, e aceitando a existência de um período de 12 meses.

Série do "Lynx Data"

A maior ordenada é em $j=12$, resultando $\tilde{S}=0,520$ e ao nível de 5%, o valor crítico é 0,12, o que nos leva a rejeitar a hipótese de que o espectro é puramente contínuo, donde a existência de uma componente periódica de período 9,5 anos.

Série de Precipitações

A maior ordenada ocorre em $j=10$ resultando $\tilde{S}=0,145$, e ao nível de 5% o valor crítico é 0,13135, nos levando a aceitar a existência de um período de 13,1 anos.

Testando os picos de menor grandeza pela ordem de magnitude observamos que a segunda maior ordenada é $\tilde{S}=0,092$, e o va lor crítico correspondente é 0,0699. Com isto aceitamos a existência de um período de 26,2 anos.

A terceira maior ordenada ocorre em j=36 e o valor de $\tilde{S}$ é 0,056 e como $\tilde{S}<0,061$, que é o valor crítico, rejeitamos a existência de um terceiro período.

Notamos que o Teste de Hannan apresenta os mesmos resultados que o Teste de Whittle, fato que já era esperado devido à escolha da janela $W(\theta)$ como a janela do periodograma truncado, e pelo fato que a melhora que Hannan apresenta ao tornar o teste não viciado é mais no sentido teórico que no prático.

# CAPÍTULO 5

# O TESTE $P(\lambda)$

## 5.1 - Introdução

No capítulo anterior discutimos alguns aspectos da análise de processos com espectros mistos, isto é, processos cujo espectro contém componentes discretas e contínuas, e consideramos três tipos de testes para detectar a presença de componentes harmônicas.

Observamos que esses testes, os quais se baseiam na análise do periodograma, apresentam várias desvantagens.

Neste capítulo desenvolveremos o método do correlograma, proposto por Priestley (1962), que se baseia na análise mais direta da função de autocorrelação.

## 5.2 - O Teste $P(\lambda)$

Consideremos um processo estacionário $X_t$ dado por (4.1.1) e com função de autocovariância dada por:

$$\gamma_X(s) = \gamma_Y(s)+\gamma_Z(s), \qquad (5.2.1)$$

onde $\gamma_Y(s)$ e $\gamma_Z(s)$ denotam as funções de autocovariância de $Y_t$ e $Z_t$ respectivamente.

Como $Y_t$ tem espectro puramente contínuo, então $\gamma_Y(s)$ pode ser expressa como a transformada de Fourier da função densidade espectral contínua, e sabemos que, neste caso, $\gamma_Y(s) \xrightarrow[s\uparrow\infty]{} 0$, isto é, valores muito afastados no tempo são aproximadamente não correlacionados.

Por outro lado, $\gamma_Z(s)$ consiste de um certo número de ondas seno com a mesma frequência que $Z_t$, e que $\gamma_Z(s) \not\to 0, s\uparrow\infty$.

Consequentemente, $\gamma_X(s)$ pode oscilar com amplitude variável na parte inicial da função, mas irá se estabilizar a uma oscilação regular, da mesma forma que $\gamma_Z(s)$ à medida que s cresce.

Então, o fato que nos permite distinguir entre um processo com espectro misto de um processo com espectro puramente contínuo é o comportamento de $\gamma_X(s)$ para s grande.

Daniels (1946) mostrou que há uma relação entre a largura de faixa da função $f_Y(\omega)$ e sua transformada de Fourier, isto é, se $\gamma_Y(s)$ é a transformada de Fourier de $f_Y(\omega)$ então quanto maior a largura de faixa de $f_Y(\omega)$ mais estreita será a de $\gamma_Y(s)$, isto é, a razão pela qual $\gamma_Y(s)$ decresce depende da largura de faixa de $f_Y(\omega)$.

No capítulo anterior vimos a necessidade de restringir a largura de faixa de $f_Y(\omega)$ em relação ao número de observações, e aqui podemos igualmente especificar essa restrição em termos da razão do decaimento da função de autocovariância $\gamma_Y(s)$.

Logo, dada uma sequência de n observações, podemos assumir que existe um inteiro $m \ll n$, tal que $\gamma_Y(s) \sim 0$ para $|s| \geq m$.

Como estamos pesquisando a existência de componentes harmônicas podemos então realizar a análise harmônica da cauda da função de autocovariância.

Seja

$$P(\lambda) = \frac{1}{2\pi} \sum_{|s|=m}^{n'} C_X(s) \cos(s\lambda), \tag{5.2.2}$$

onde $C_X(s)$ é a função de autocovariância amostral, $0 \leqslant \lambda \leqslant \pi$ e $n' < n$.

Iremos testar a hipótese nula $H_0$ de que $X_t$ tem um espectro puramente contínuo contra a hipótese alternativa de que $X_t$ contém uma componente harmônica com frequência $\lambda_0$.

Sob a hipótese nula $\gamma_Z(s)=0$, $\forall s$ e do fato que $\gamma_Y(s) \sim 0$, $|s| \geqslant m$, segue que:

$$E[P(\lambda)] \sim 0, \quad \forall\lambda \tag{5.2.3}$$

Pode-se demonstrar também (Priestley, 1957) que:

$$\operatorname{Var}[P(\lambda)] \sim \begin{cases} \dfrac{f_Y^2(\lambda)}{n}\left(\dfrac{2n'}{3} - 2m + \dfrac{2m^2}{n'}\right), & \lambda \neq 0 \\ \\ \dfrac{2f_Y(0)}{n}\left(\dfrac{2n'}{3} - 2m + \dfrac{2m^2}{n'}\right), & \lambda = 0 \end{cases} \tag{5.2.4}$$

e

$$\operatorname{cov}[P(\lambda), P(\lambda')] \sim o\left(\frac{1}{n}\right), \quad \lambda \neq \lambda' \tag{5.2.5}$$

Entretanto sob a hipótese alternativa temos que:

$$E[P(\lambda)] \simeq \begin{cases} 0, & |\lambda-\lambda_0| \gg O\left(\frac{1}{m}\right) \\ O(n'-m), & \lambda \sim \lambda_0 \end{cases}$$

Com base nestes resultados é possível construir um teste de significância baseado em $P(\lambda)$ para detectar a presença de componentes harmônicas observando os picos dessa função.

Como não há necessidade de restringirmos a discussão da função dada por (5.2.2), podemos considerar uma expressão mais geral, da forma:

$$P(\lambda) = \frac{1}{2\pi} \sum_{s=-(n'-1)}^{n'-1} [w^{(1)}_{s,n',\lambda} - w^{(2)}_{s,m,\lambda}] C_X(s), \qquad (5.2.7)$$

onde $w^{(1)}_{s,n',\lambda}$ e $w^{(2)}_{s,m,\lambda}$ são duas sequências de funções peso, mencionadas no Capítulo 2.

Notamos que (5.2.7) se reduz a (5.2.2) para $w^{(1)}_{s,n',\lambda}$ e $w^{(2)}_{s,m,\lambda}$ dadas por:

$$w^{(1)}_{s,n',\lambda} = \begin{cases} \cos(s\lambda), & |s|<n' \\ 0, & |s|\geq n' \end{cases} \qquad (5.2.8)$$

e

$$w^{(2)}_{s,m,\lambda} = \begin{cases} \cos(s\ ), & |s|<m \\ 0, & |s|\geq m \end{cases} \qquad (5.2.9)$$

Para simplificar a notação denotaremos $w^{(1)}_{s,n',\lambda}$ e $w^{(2)}_{s,m,\lambda}$ por $w^{(1)}_s$ e $w^{(2)}_s$, respectivamente, mas sem esquecer a dependência dessas funções dos parâmetros n' e m, respectivamente.

Se as duas sequências de pesos $w^{(1)}_s$ e $w^{(2)}_s$ são escolhidas de modo que $m=o(n') \xrightarrow[n\uparrow\infty]{} \infty$, então vale o seguinte teorema (Priestley, 1962 a):

<u>Teorema 5.1</u> - Seja um processo linear $X_t$ e duas sequências de pesos $w_s^{(1)}$ e $w_s^{(2)}$, tais que $m=o(n') \xrightarrow[n\uparrow\infty]{} \infty$.

Então, se $Z_t \equiv 0$, temos:

i) $\lim\limits_{n\to\infty} E[P(\lambda)] = 0$

ii) $Var[P(\lambda)] \sim \frac{2f^2(\lambda)}{n} \sum\limits_s [w_s^{(1)} - w_s^{(2)}]^2$

iii) $Cov[P(\lambda), P(\lambda')] \sim O(\frac{1}{n})$, onde $[\lambda-\lambda'] \gg \frac{1}{m}$.

Seja m' um inteiro tal que $m'=o(m) \xrightarrow[n\uparrow\infty]{} \infty$ e vamos subdividir o domínio $(0,\pi)$ em intervalos de comprimento $\frac{2\pi}{m'}$.

Vamos definir as somas acumuladas normalizadas, $J_q$, como:

$$J_q = \left(\frac{n}{m'\Lambda_{n',m}}\right)^{1/2} \sum_{p=1}^{q} P\left(\frac{2\pi p}{m'}\right), \quad q=0,\ldots,[\tfrac{1}{2}m'], \tag{5.2.10}$$

onde

$$\Lambda_{n',m} = \sum_s [w_s^{(1)} - w_s^{(2)}]^2 \sim n\, Var\left[\frac{P(\lambda)}{f^2(\lambda)}\right]$$

Em geral, $\Lambda_{n',m} = O(n)$, de modo que a correlação entre $P(\lambda)$ e $P(\lambda')$ tende a zero quando n tende a infinito.

Como os termos da soma dada por (5.2.10) são assintoticamente independentes podemos fazer uma analogia entre eles e o passeio aleatório (Bartlett, 1955), obtendo assim:

$$\lim_{n\to\infty} P\left[\max_{0\leq \frac{2\pi q}{m'}\leq \pi} (J_q)\leq\alpha_0\right] = P\left[\max_{0\leq q\leq\pi} (\eta(q))\leq\alpha_0\right] \tag{5.2.11}$$

onde $\eta(q)$ é um processo normal com média zero.

Notamos que o lado direito de (5.2.11) pode ser interpretado como a probabilidade que o passeio aleatório $\eta(q)$ não seja absorvido na barreira $\alpha_0$.

Então:

$$\lim_{n\to\infty} P\left[\frac{\max\limits_{0\leq \frac{2\pi q}{m'}\leq \pi} (J_q)}{\left[\frac{1}{2\pi}\ G(\pi)\right]^{1/2}}\leq \alpha_0\right] = 2\Phi(\alpha_0)-1 = \Delta(\alpha_0), \tag{5.2.12}$$

onde $\Phi$ é a função distribuição da normal padrão e

$$G(\pi) = \int_0^{\pi} f_Y^2(\omega)d\omega.$$

Podemos utilizar (5.2.12) para determinar se algum $A_i$ é positivo e a correspondente frequência $\omega_i$, mas para isso precisamos estimar $G(\pi)$.

Mas

$$G(\pi) = \int_0^{\pi} f_Y^2(\omega)d\omega = \frac{1}{4\pi}\sum_{s=-\infty}^{\infty}[\gamma_Y(s)]^2.$$

Logo, se $A_i=0$, $\forall i$, um estimador consistente de $G(\pi)$ é dado por

$$G^*(\pi) = \frac{1}{4\pi}\sum_{s=-(m-1)}^{m-1}[C_X(s)]^2, \tag{5.2.13}$$

Entretanto, sob a hipótese alternativa de que $A_i>0$, para pelo menos um i, $G(\pi)$ pode ser estimado por:

$$\hat{G}(\pi) = \frac{1}{4\pi}\left(\sum_{s=-(m-1)}^{m-1} [C_X(s)]^2 - 2\sum_{s=m}^{2m-1} C_X(s)]^2\right) \qquad (5.2.14)$$

Pode-se então testar que $A_i=0$, $\forall i$, contra a hipótese alternativa de que $A_i>0$ para pelo menos um i, usando a distribuição de $J_q$ mencionada acima.

Sob a hipótese nula $J_q$ satisfaz a equação (5.2.12), para n grande.

Entretanto sob a hipótese alternativa $J_q=O(\sqrt{\frac{nn'}{m'}})$, onde $\frac{2\pi q}{m'} \sim \omega_i$ e assumindo $\Lambda_{n',m}=O(n)$, (ver Priestley, 1962 a).

Então, com alta probabilidade, $J_q$ irá cruzar a barreira $\alpha_0$.

Notamos que precisamos considerar apenas a barreira simples, uma vez que sob a hipótese alternativa $J_q$ tende a ser positivo e grande.

Então, o teste de significância pode ser construido esboçando-se o gráfico de $J_q$ contra q e determinando se $J_q$ cruza a barreira $\alpha_0$, onde $\alpha_0$ é determinado pelo nível de significância do teste, isto é, $\alpha_0$ é obtido pelos pontos da distribuição normal padrão que dão uma porcentagem bicaudal com $\frac{1-\Delta(\alpha_0)}{2}$ em cada cauda.

## 5.3 - Algumas Janelas Especiais

Muitas das janelas conhecidas poderiam ser escolhidas para $w_s^{(1)}$ e $w_s^{(2)}$, não havendo nenhuma razão especial para que elas sejam tomadas na mesma classe de janelas.

Mas, do ponto de vista de reduzir o vício de $P(\lambda)$, uma "boa" escolha seria:

$$w_s^{(1)} = \begin{cases} [1-\frac{|s|}{n'}]\cos(s\lambda), & |s|<n' \\ 0, & |s|\geqslant n' \end{cases} \tag{5.3.1}$$

que corresponde a $W^{(1)}(\lambda)=F_{n'}(\lambda)$, ou seja, o núcleo de Féjer e

$$w_s^{(2)} = \begin{cases} [1-\frac{|s|}{n'}]\cos(s\lambda), & |s|<m \\ 0, & |s|\geqslant m \end{cases} \tag{5.3.2}$$

Como $m=o(n')$, (5.3.2) é equivalente a:

$$w_s^{(2)} = \begin{cases} \cos(s\lambda), & |s|<m \\ 0, & |s|\geqslant m \end{cases} \tag{5.3.3}$$

que corresponde a $W^{(2)}(\lambda)=D_m(\lambda)$, ou seja, o núcleo de Dirichlet.

Com as janelas dadas em (5.3.1) e (5.3.3) podemos escrever (5.2.7) como:

$$P(\lambda) = \frac{1}{\pi}\sum_{s=m}^{n'-1}[1-\frac{|s|}{n'}]C_X(s)\cos(s\lambda) \tag{5.3.4}$$

e neste caso $P(\lambda)$ é independente das m primeiras covariâncias.

Assim $\Lambda_{n',m}$ será dado por:

$$\Lambda_{n',m} \sim \begin{cases} \frac{2}{3}n' - 2m + \frac{2m^2}{n'}, & \lambda\neq 0,\pi \\ \frac{4}{3}n' - 4m + \frac{4m^2}{n'}, & \lambda=0,\pi \end{cases} \tag{5.3.5}$$

Uma vantagem na escolha destas duas janelas é que nós podemos remover a restrição $m'=o(m)$.

Como vimos, o valor de m é determinado pela razão do decaimento de $\gamma_Y(s)$, ou seja, pela largura de faixa de $f_Y(\omega)$, e $m \ll n' < n$.

Mas, se a magnitude do pico em $P(\lambda)$ devido às componentes harmônicas é muito pequena comparada com a variância de $Y_t$, será necessário usar o correlograma completo a fim de detectar esta componente.

Neste caso (5.2.7) será dada por:

$$P_N(\lambda) = I_{N,X}(\lambda) - \frac{1}{2\pi} \sum_{s=-(m-1)}^{m-1} C_X(s) \cos(s\lambda), \qquad (5.3.6)$$

onde $W_{N,\lambda}^{(1)}$ e $W_{N,\lambda}^{(2)}$ são núcleos de Dirichlet.

Sob a hipótese alternativa, as amplitudes $A_i$, $i=1,\dots,k$ podem ser estimadas por:

$$\hat{A}_i^2 = \frac{8\pi}{n'-2m} P(\omega_i), \quad i=1,\dots,k \qquad (5.3.7)$$

As vantagens do teste $P(\lambda)$ consistem na utilização de uma estatística cujo valor esperado é assintóticamente independente de $f_Y(\omega)$ e no fato de não requerer o cálculo do correlograma completo, na maioria dos casos.

## 5.4 - O Poder Assintótico do Teste $P(\lambda)$

Antes de discutirmos o poder assintótico do teste $P(\lambda)$, consideremos brevemente o poder assintótico do teste clássico de

Fisher.

A estatística $g^{(1)}$ utilizada por Fisher é dada por:

$$g^{(1)} = \frac{I^{(1)}}{I_1+\ldots+I_{[\frac{n}{2}]}}.$$

Sob a hipótese nula de que o processo é independente, g tem distribuição assintótica dada por:

$$P[g \geqslant x_0] \sim 1-(1-e^{-x_0/2})^{n/2}. \qquad (5.4.1)$$

Se nós fixarmos o nível de significância $\alpha$, o valor crítico de $x_0$ é aproximadamente:

$$x_0 \simeq 2 \log_e(\frac{n}{2\alpha}), \qquad (5.4.2)$$

para $\alpha$ pequeno.

Mas, se X(t) contém uma componente harmônica na frequência $\frac{2\pi\ell}{n}$, então $I_{X,p}$ $(p\neq\ell)$ não são afetadas pelo termo harmônico e têm a mesma distribuição que sob a hipótese nula.

Entretanto, $I_{X,\ell}$ tem uma distribuição $\chi^2$ não central, com fator de não centralidade $\lambda$, dado por:

$$\lambda = \frac{nA_\ell^2}{2\sigma_Y^2}, \qquad (5.4.3)$$

onde $A_\ell^2$ é a amplitude da $\ell$-ésima frequência.

Aproximando essa distribuição por uma $\rho\chi_\nu^2$, onde $\rho$ é um fator escala e $\chi_\nu^2$ é uma $\chi^2$ central, com $\nu$ graus de liberdade (Hartley, 1949), dados por:

$$\rho = \frac{2+2\lambda}{2+\lambda} \quad e \quad \nu = \frac{(2+\lambda)^2}{2+2\lambda}, \tag{5.4.4}$$

e o poder do teste é dado assintóticamente por:

$$p(x_0) = P(g \geqslant x_0) = 1-(1-e^{-x_0/2})^{\frac{n}{2}-1} p_1(x_0,\lambda) \tag{5.4.5}$$

onde

$$p_1(x_0,\lambda) = 2^{\frac{\nu}{2}-1} \Gamma(\alpha/2) \int_0^{x_0/\rho} e^{-\frac{\tau}{2}} \tau^{\frac{1}{2}\nu-1} d\tau .$$

O teste de Whittle, que emprega a estatística dada por (4.2.13), tem poder assintótico (ver Priestley, 1962 II):

$$P_W(x_0) = P[g \geqslant x_0] = 1-(1-e^{-\phi\frac{x_0}{2}})^{\frac{n}{2}-1} p_1(\phi x_0,\lambda') \tag{5.4.6}$$

onde $\phi = \frac{n}{2} \frac{A^2}{4\pi f(\omega_\ell)}$ e $\lambda'$ é o fator de não centralidade da $\chi^2$.

O teste do peridograma agrupado, baseado na estatística $g_k^{(B)}$ dada por (4.3.2), tem o poder assintótico igual a (ver Priestley, 1962 II):

$$P_B(x_0) = P(g_k^{(B)} \geqslant x_0) = 1-(1-e^{-\beta\frac{x_0}{2}})^{\frac{n}{2}-1} p_1(\beta x_0,\lambda') \tag{5.4.7}$$

onde $\beta = \frac{M'}{m'}$ e $\lambda'$ é o fator de não centralidade da $\chi^2$.

A fim de comparar o poder assintótico do teste de Whittle e do teste do periodograma agrupado, notamos que para n e k grandes, e $\alpha$ pequeno, (5.4.6) e (5.4.7), respectivamente, se reduzem a:

$$P_W(x_0) \sim 1-(1-\alpha)p_1(\phi x_0,\lambda'), \tag{5.4.8}$$

e

$$P_B(x_0) \sim 1-(1-\alpha)p_1(\beta x_0,\lambda'). \tag{5.4.9}$$

Como $p_1(.,\lambda')$ cresce monotonicamente, temos que:

$$p_1(\phi x_0,\lambda') > p_1(\beta x_0,\lambda'), \quad \text{se } \phi>\beta$$

Então, $P_B(x_0)>P_W(x_0)$, se $\frac{n}{2} A^2>8\pi f(\omega_\ell)$.

O teste de Priestley se baseia nas quantidades normalizadas $J_q$, dadas por (5.2.10).

De (5.2.12) temos que:

$$p = \lim_{n\to\infty} P\left[0 \leqslant \frac{\max\limits_{0\leqslant \frac{2\pi q}{m'} \leqslant \pi} J_q}{[\frac{1}{2\pi} G(\pi)]^{1/2}} \leqslant \alpha_0\right] =$$

$$= \int_{-\infty}^{\alpha_0\sqrt{\gamma(\pi)}} p_2(x,\frac{2\pi\ell}{n})\,\Phi\left[\frac{(\alpha_0\sqrt{\gamma\pi}-x)/(n/m\Lambda_{n',m})^{\frac{1}{2}}-m_\ell}{\sigma_\ell}\right] dx \tag{5.4.10}$$

onde

$$\gamma(t) = \frac{1}{2\pi} G(t) = \frac{1}{2\pi}\int_0^t f^2(s)ds$$

e

$$p_2(x,t) = [2\pi\gamma(t)]^{-\frac{1}{2}}[\exp\{\frac{-x^2}{2\gamma(t)}\}-\exp\{-\frac{(x-2\alpha_0)^2}{2\gamma(t)}\}]$$

e $\Phi$ é a função de distribuição da normal padrão.

Seja $\phi(\mu)$ a função densidade da normal padrão, então:

$$p = \alpha_0\mu \int_{-\infty}^{1} [\phi(\alpha_0\mu y) - \phi[\alpha_0\mu(y-2)]]\Phi(ay+b)dy$$

onde

$$\mu = \left[\frac{\gamma(\pi)}{\gamma(\frac{2\pi\ell}{n})}\right]^{1/2}$$

$$a = \frac{-\alpha_0[\gamma(\pi)m'\Lambda_{n',m}]^{1/2}}{n^{1/2}\sigma_\ell}$$

sendo $m_\ell$ e $\sigma_\ell$ respectivamente a média e o desvio padrão da distribuição assintótica de $P(\omega_\ell)$.

Expandindo $\Phi(ay+b)$ em série de Taylor e ignorando a variação de $\hat{G}(\pi)$, temos:

$$\hat{G}(\pi) \sim 2\pi\gamma(\pi).$$

Então, o poder do teste $P(\lambda)$ é dado assintóticamente por (ver Priestley, 1962 II):

$$P_0 \sim 1-\{\Phi(b) \int_{-\infty}^{\alpha_0\sqrt{\gamma(\pi)}} p^2(x,\frac{2\pi\ell}{n})dx - \frac{2ae^{-\frac{b^2}{2}}}{\sqrt{2\pi}} \Phi(-\alpha_0\mu) \qquad (5.4.11)$$

Como $e^{-\frac{b^2}{2}} = O(-e^{-n})$, podemos escrever (5.4.11) como:

$$P_0 \sim 1-(1-\alpha)\Phi(b)$$

Observando $P_W$, $P_B$ e $P_0$, podemos comparar a eficiência $V_W$, $V_B$ e $V_0$ dos três testes, respectivamente, em relação ao teste clássico de Fisher.

De (5.4.1) temos:

$$x_0 = 2 \log_e(\frac{n}{2\alpha}) = O(\log n)$$

Então:

$$V_W = O(\frac{m \log n}{n^{1/2}})$$

$$V_B = O(\frac{\log n}{n^{1/2}})$$

$$V_0 = O(\sqrt{\frac{m}{n'}})$$

Então, em geral $V_B=o(V_0)$, e o teste do periodograma agrupado é mais poderoso.

Entretanto, se $n'=n$ e $m=\log n$, então $V_0=o(V_B)$ e o teste $P(\lambda)$ é mais poderoso.

## 5.5 - Aplicações do Teste $P(\lambda)$

Faremos agora uma aplicação do método de Priestley às séries de temperaturas, "Lynx Data" e precipitações de chuvas de Fortaleza, Ceará, listadas no Apêndice.

Lembramos que ao aplicar o teste nós selecionamos o primeiro pico (em ordem de frequência) de $P(\lambda)$, digamos $\lambda=\lambda_0$, e selecionamos a ordenada de $P(\lambda)$ onde ela ocorre, e que os valores críticos para max $J_q$ são obtidos da variável normal padronizada.

Para remover a contribuição do termo harmônico significante da função de covariância, foi usada a fórmula:

$$C'_X(s) = C_X(s) - \frac{1}{2} \hat{A}_i^2 \cos(s\hat{\omega}_0), \qquad (5.5.1)$$

onde $\hat{A}_i^2$ é dado por (5.3.7).

Verificamos que ela não é suficientemente precisa devido ao fato da função de correlação ser instável em sua cauda.

Uma explicação para este resultado (ver Bhansali, 1979) está no fato de que o uso da correção dada por (5.5.1) é válida somente para n grande. Para séries pequenas, como é o caso das que apresentaremos, a função de covariância $C_X$ após remover o efeito desta componente harmônica em $\hat{\omega}_0$ seria melhor estimada por:

$$C'_X(s) = \frac{1}{n} \sum_{i=1}^{n-|s|} X'_i X'_{i+|s|}, \quad s=0,\pm1,\ldots,\pm n-1 \qquad (5.5.2)$$

onde

$$X'_i = Z_i - \hat{A}\cos(\hat{\omega}_0 i) - \hat{B}\,\mathrm{sen}(\hat{\omega}_0 i)$$

$$\hat{A} = \frac{2}{n} \sum_{i=1}^{n} Z_i \cos(\hat{\omega}_0 i)$$

$$\hat{B} = \frac{2}{n} \sum_{i=1}^{n} Z_i\,\mathrm{sen}(\hat{\omega}_0 i)$$

sendo $Z_t$ a série de tempo observada corrigida pela média.

Neste caso, a amplitude $\hat{A}_i^2$ da componente harmônica seria estimada por:

$$\hat{A}_i = \sqrt{(\hat{A})^2 + (\hat{B})^2} \qquad (5.5.3)$$

Série de Temperaturas

Nesta aplicação tomamos m=15 e n'=35. A escolha foi devido ao fato de ter sido o ponto de truncamento com o qual obtivemos melhores resultados (também testamos com m=10 e 20) e para n' obedeceu-se à relação: n'>2m.

Observamos que $P(\lambda)$ tem picos ocorrendo em: $\frac{8\pi}{96}$, $\frac{16\pi}{96}$ e $\frac{24\pi}{96}$.

Para $\lambda = \frac{8\pi}{96}$, o valor de (5.2.10) será dado por:

$$J_2 = \frac{1,6 \times 2,139}{1,871} = 1,83$$

Tomando-se $\alpha = 1\%$, obtemos o valor de max $J_q$ como $\alpha_0 = 2,33$.

Como $J_2 < 2,33$, o pico observado em $\frac{8\pi}{96}$ não acarreta a existência de um período.

Entretanto, para $\lambda = \frac{16\pi}{96}$ o valor de (5.2.10) é dado por:

$$J_2 = \frac{1,6 \times 7,413}{1,871} = 6,339$$

Tomando-se $\frac{\alpha}{2} = 0,5\%$, obtém-se $\alpha_0 = 2,58$, o que nos leva a rejeitar a hipótese nula de que o espectro é puramente contínuo, devido à existência de uma componente harmônica de período doze meses.

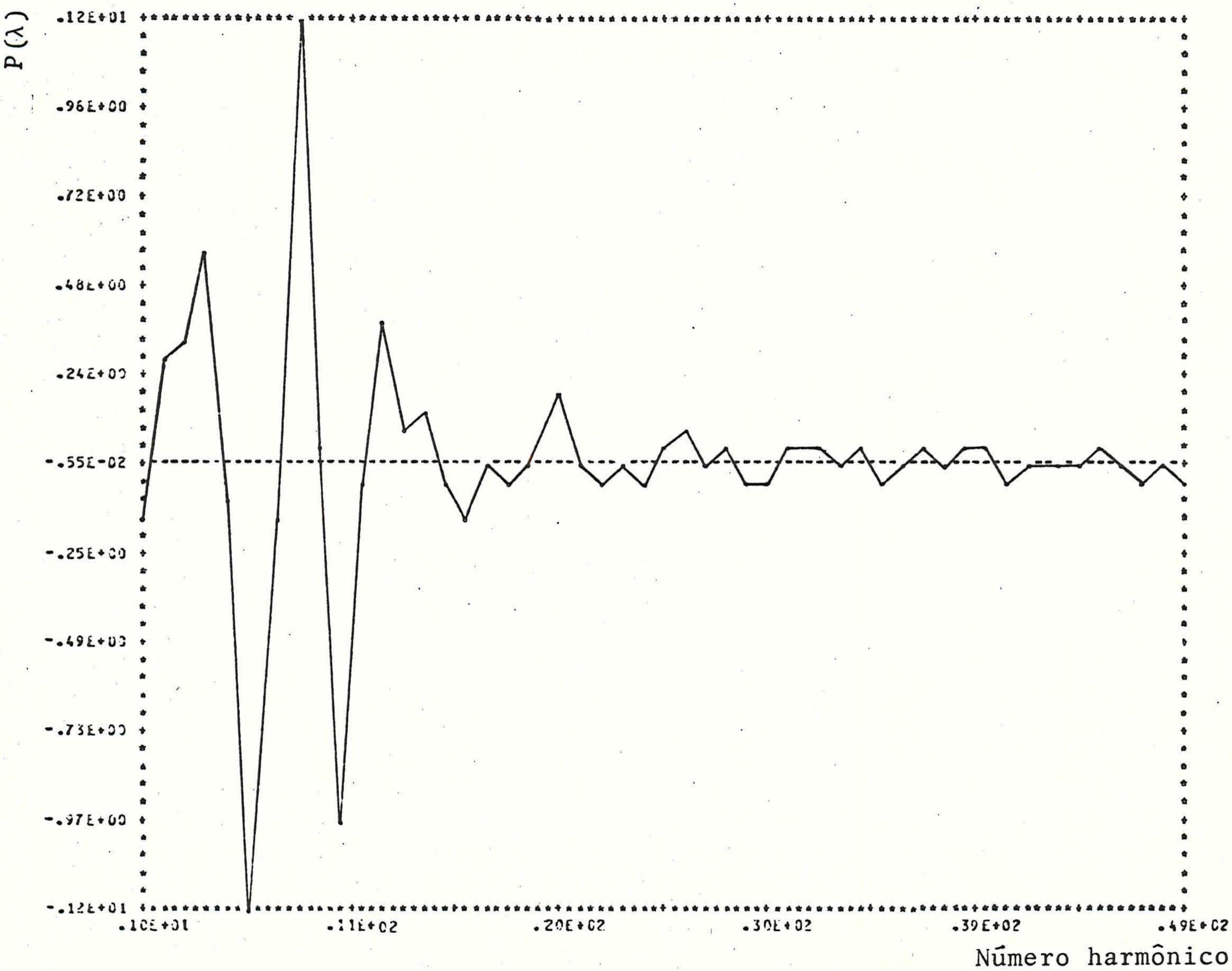
Os valores de $\hat{A} = -3,019$ e $\hat{B} = -0,394$, nos permitem estimar a amplitude desta componente. Então de (5.5.3) temos:

$$\hat{A}_1 = 3,04.$$

Recalculando $C_X$ após eliminar o efeito desta componente harmônica, observamos na Figura 5.2 o novo gráfico de $P(\lambda)$, onde observamos um pico em $\frac{32\pi}{96}$, logo:

$$J_2 = \frac{1,6 \times 1,373}{1,045} = 2,103$$

Ao nível $\frac{\alpha}{3} = 0,33\%$, temos $\alpha_0 = 2,71$ o que nos leva a rejeitar a existência de outro período.




<u>Figura 5.1</u> - Gráfico de P(λ) para a série de temperaturas

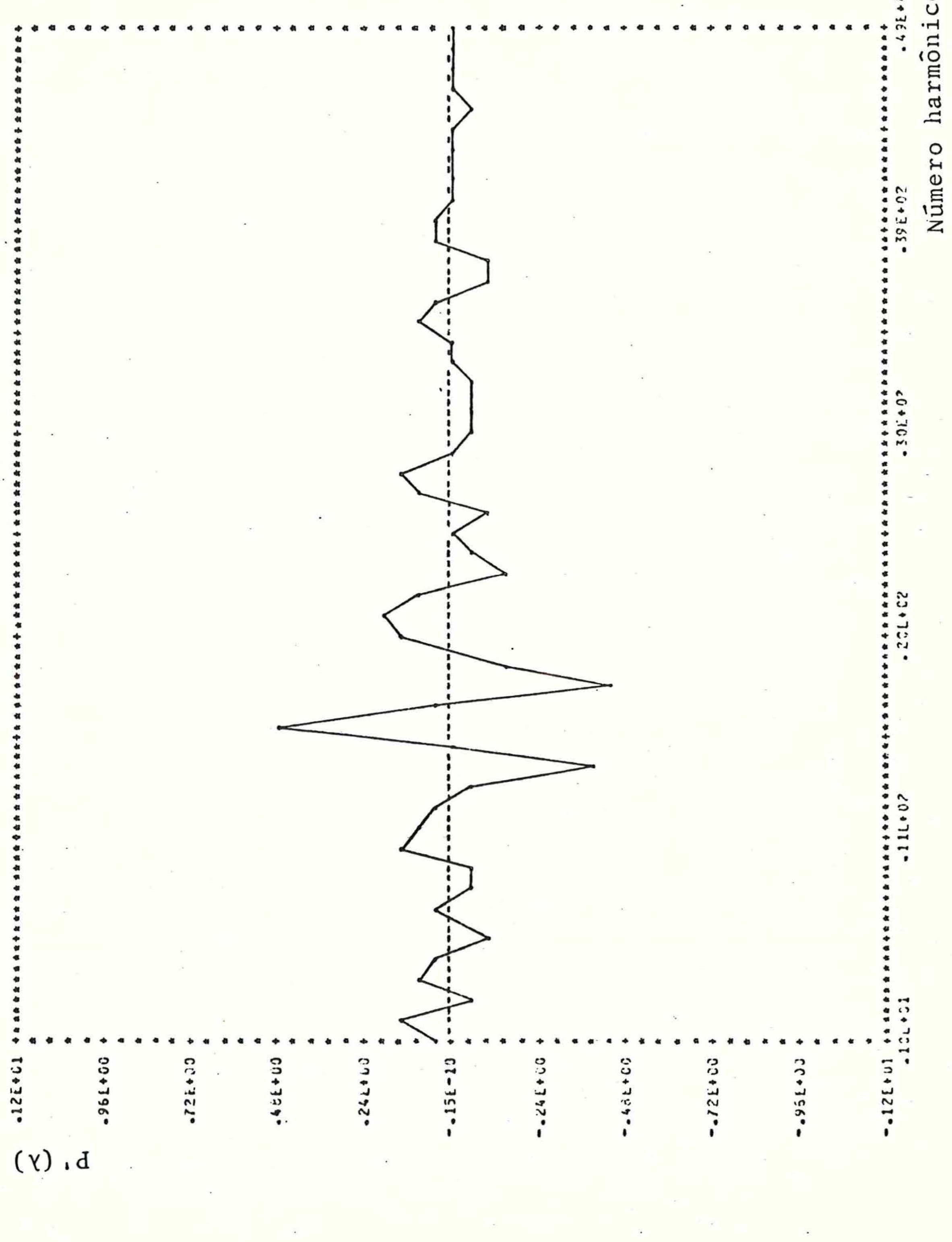



Figura 5.2 - Gráfico de P'(λ) para a série de temperaturas

<u>Série do "Lynx Data"</u>

Nesta aplicação aplicaremos o teste $P(\lambda)$ ao logaritmo na base 10 dos dados observados.

Utilizamos m=10 e n'=25 (também foi testado m=15 e n'=35, e m=20 e n'=45), o qual apresentou mais estabilidade na função de auto-covariância.

Observando a Figura 5.3 verificamos que $P(\lambda)$ apresenta picos em: $\frac{6\pi}{114}$, $\frac{24\pi}{114}$ e $\frac{44\pi}{114}$.

No caso de $\lambda = \frac{6\pi}{114}$, temos:

$$J_2 = \frac{2,387 \times 0,902}{1,199} = 1,796$$

Ao nível de $\alpha$=1%, temos $\alpha_0$=2,33, o que nos leva a rejeitar a existência desta componente harmônica.

No caso $\lambda = \frac{24\pi}{114}$, temos:

$$J_2 = \frac{2,387 \times 8,014}{1,199} = 15,954$$

Fixando $\frac{\alpha}{2} = 0,5\%$, temos $\alpha_0$=2,58 logo aceitamos a existência de uma componente harmônica de período 9,5 anos.

Os valores de $\hat{A}$=0,3427 e $\hat{B}$=-0,3003, nos permitem estimar a amplitude desta componente:

De (5.5.3) temos:

$$\hat{A}_1 = 0,456$$

Eliminando o efeito desta componente e recalculando $C_X$, observamos na Figura 5.4, o gráfico de $P'(\lambda)$, o qual apresenta um pico em $\lambda = \frac{44\pi}{114}$, logo:

$$J_2 = \frac{2,387 \times 1,863}{1,462} = 3,041$$

Ao nível $\frac{\alpha}{3} = 0,33\%$, temos $\alpha_0 = 2,71$ o que nos levaria a aceitar o pico existente em $\lambda = \frac{44\pi}{114}$ como significante.

Entretanto, observando o periodograma da série do "Lynx Data" (ver Figura 3.4) verificamos que ele não apresenta pico na frequência $\lambda = \frac{44\pi}{114}$, o que nos leva a concluir que a existência do pico não se deve à uma componente harmônica.

A sugestão de Priestley de se testar os picos de $P(\lambda)$ em ordem de frequência é feita de maneira a evitar este tipo de problema, essencialmente pela interação entre dois picos diferentes ocorrendo em frequências $\lambda_0$ e $\lambda_1$, digamos, onde $0<\lambda_0<\lambda_1<\pi$.

Entretanto, a análise apresentada sugere que mesmo quando os picos de $P(\lambda)$ são testados em ordem de frequência, o teste pode ser afetado pela interação entre o pico em $\lambda_0$, digamos, e, por simetria, o correspondente pico em $-\lambda_0$.

Para maiores detalhes ver Bhansali (1979).

<u>Série de Precipitações</u>

Foram utilizados m=25 e n'=55 (também testamos com m=20 e n'=45).

Observando o gráfico de $P(\lambda)$ (ver Figura 5.5) verificamos a presença de picos em: $\lambda = \frac{4\pi}{131}, \frac{10\pi}{131}, \frac{20\pi}{131}, \frac{30\pi}{131}$ e $\frac{40\pi}{131}$.

Para $\lambda = \frac{4\pi}{131}$, temos:

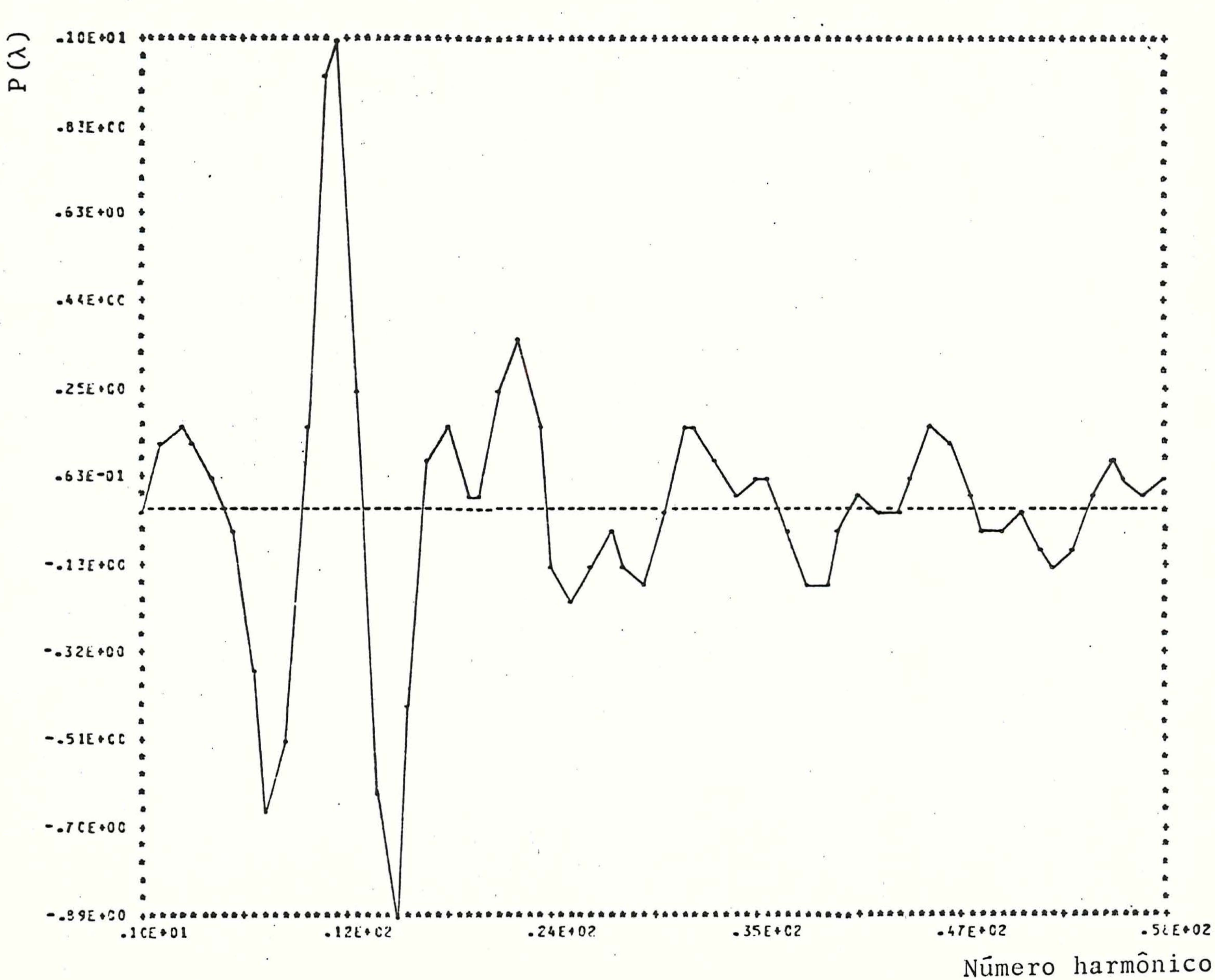



Figura 5.3 - Gráfico de P(λ) para a série do "Lynx Data"

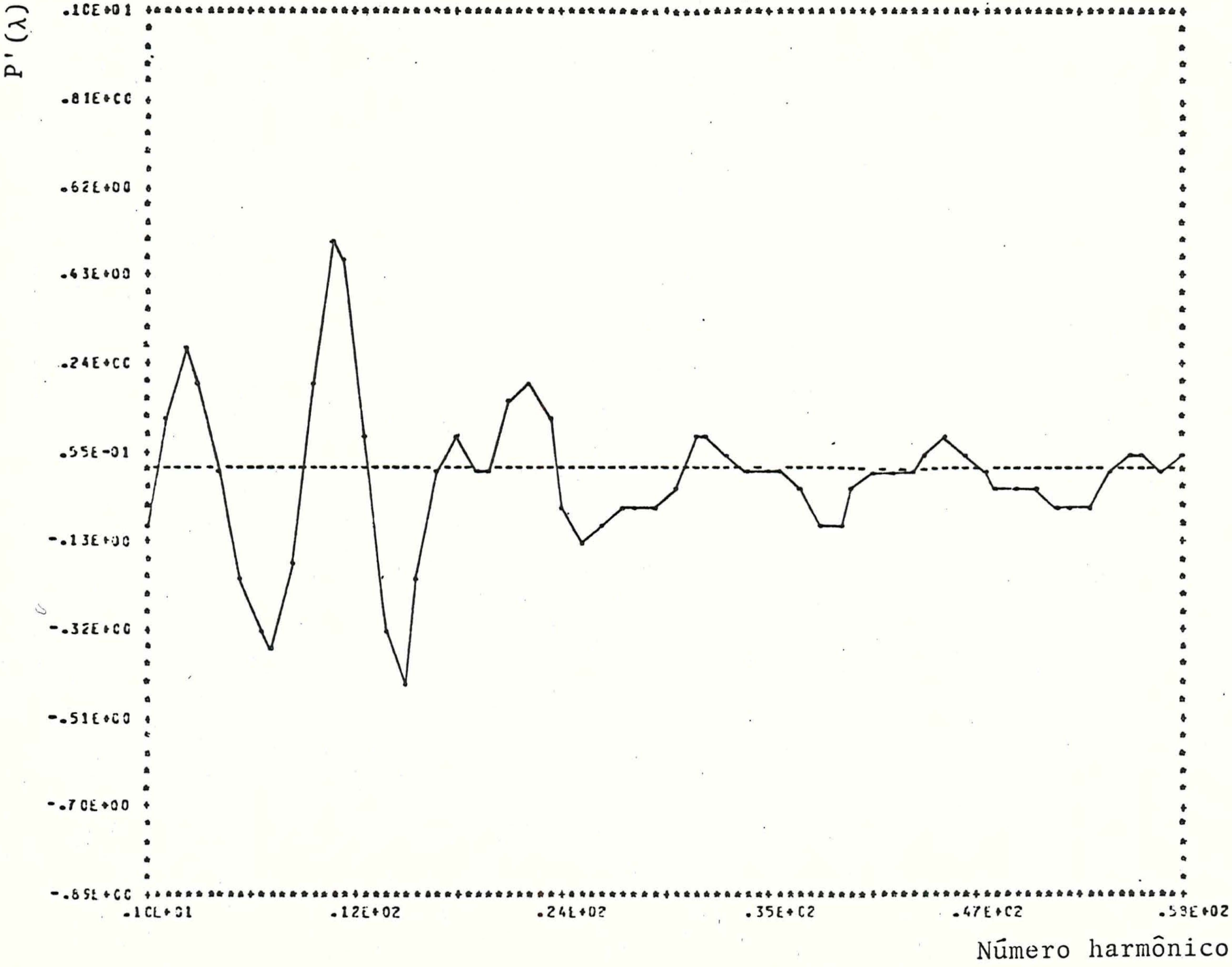



Figura 5.4 - Gráfico de P'(λ) para a série do "Lynx Data"

$$J_2 = \frac{1,144 \times 1,059}{0,989} = 1,225$$

Realizando o teste ao nível $\alpha$=1%, temos $\alpha_0$=2,33 e como $J_2$<2,33 concluimos que o pico observado em $\frac{4\pi}{131}$ não é significante.

Testando o pico observado em $\lambda = \frac{10\pi}{131}$, temos:

$$J_2 = \frac{1,144 \times 2,226}{0,989} = 2,603$$

Ao nível $\frac{\alpha}{2}$ = 0,5% temos $\alpha_0$=2,58, o que nos leva a aceitar a existência de um período de 26,2 anos.

Dos valores $\hat{A}$=-83,089 e $\hat{B}$=-184,747, podemos estimar $\hat{A}_1$,

De (5.5.3) temos:

$$\hat{A}_1 = 202,57$$

Recalculando $C_X$ após eliminar o efeito desta componente, podemos observar que o gráfico de $P'(\lambda)$ apresenta picos em: $\lambda = \frac{20\pi}{131}$ e $\lambda = \frac{72\pi}{131}$.

Para $\lambda = \frac{20\pi}{131}$, temos:

$$J_2 = \frac{1,144 \times 2,6099}{0,987} = 3,026$$

Como $\frac{\alpha}{3}$ = 0,33%, temos $\alpha_0$=2,71, o que nos leva a aceitar a existência de outra componente harmônica de período 13,1 anos.

Como $\hat{A}$=-151,98 e $\hat{B}$=-205,57 podemos estimar a amplitude desta componente.

De (5.5.3) temos:

$$\hat{A}_2 = 255,65$$

Recalculando $C_X$ após remover o efeito desta componente, observamos na Figura 5.7 o gráfico de $P''(\lambda)$, onde verificamos picos em $\lambda = \frac{46\pi}{131}, \frac{54\pi}{131}, \frac{66\pi}{131}$ e $\frac{72\pi}{131}$.

Para $\lambda = \frac{46\pi}{131}$, temos:

$$J_2 = \frac{1,144 \times 0,735}{1,038} = 0,810$$

Ao nível $\frac{\alpha}{4} = 0,25\%$ temos $\alpha_0 = 2,81$, o que nos leva a rejeitar a existência deste período.

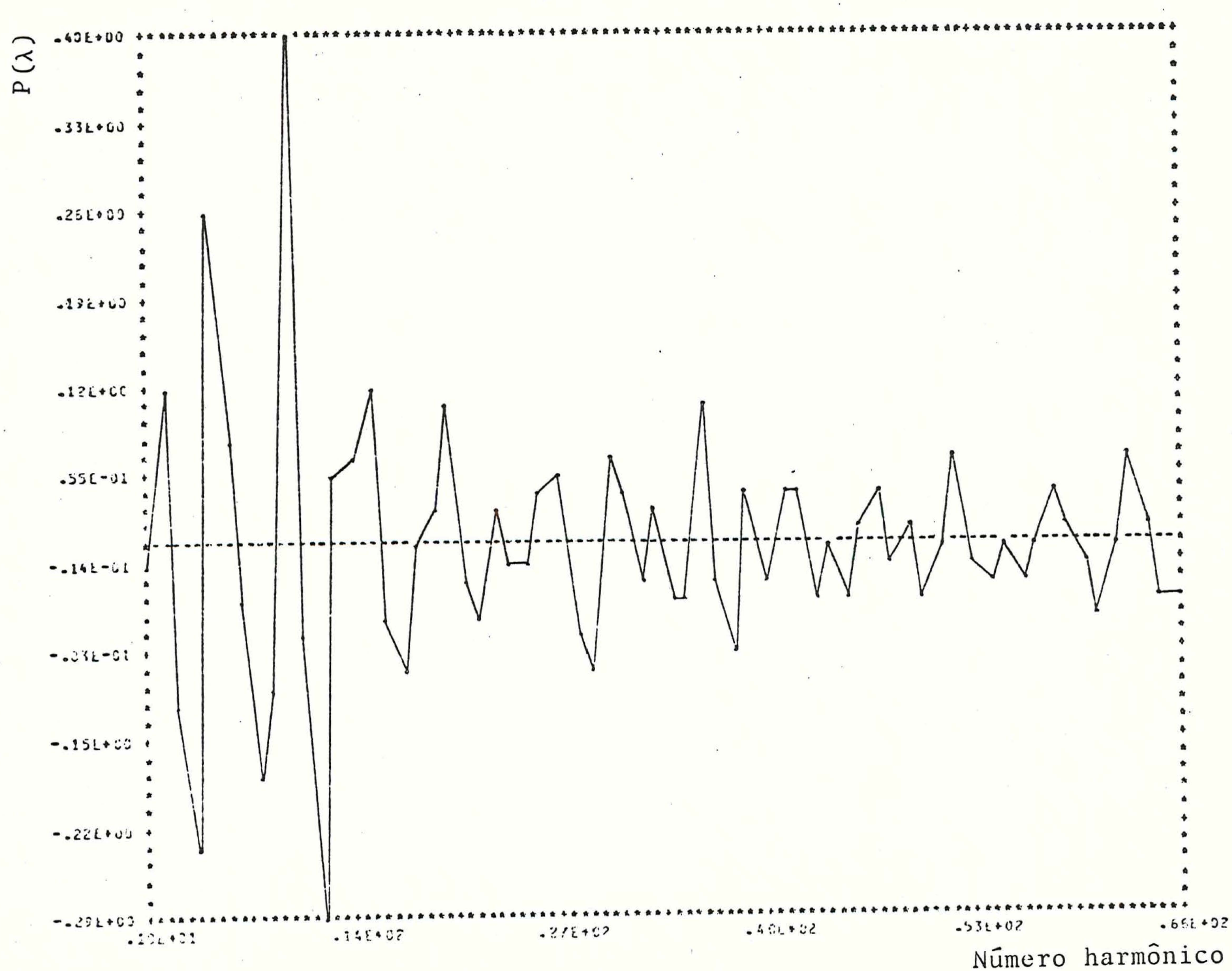



Figura 5.5 - Gráfico de P(λ) para a série de precipitações

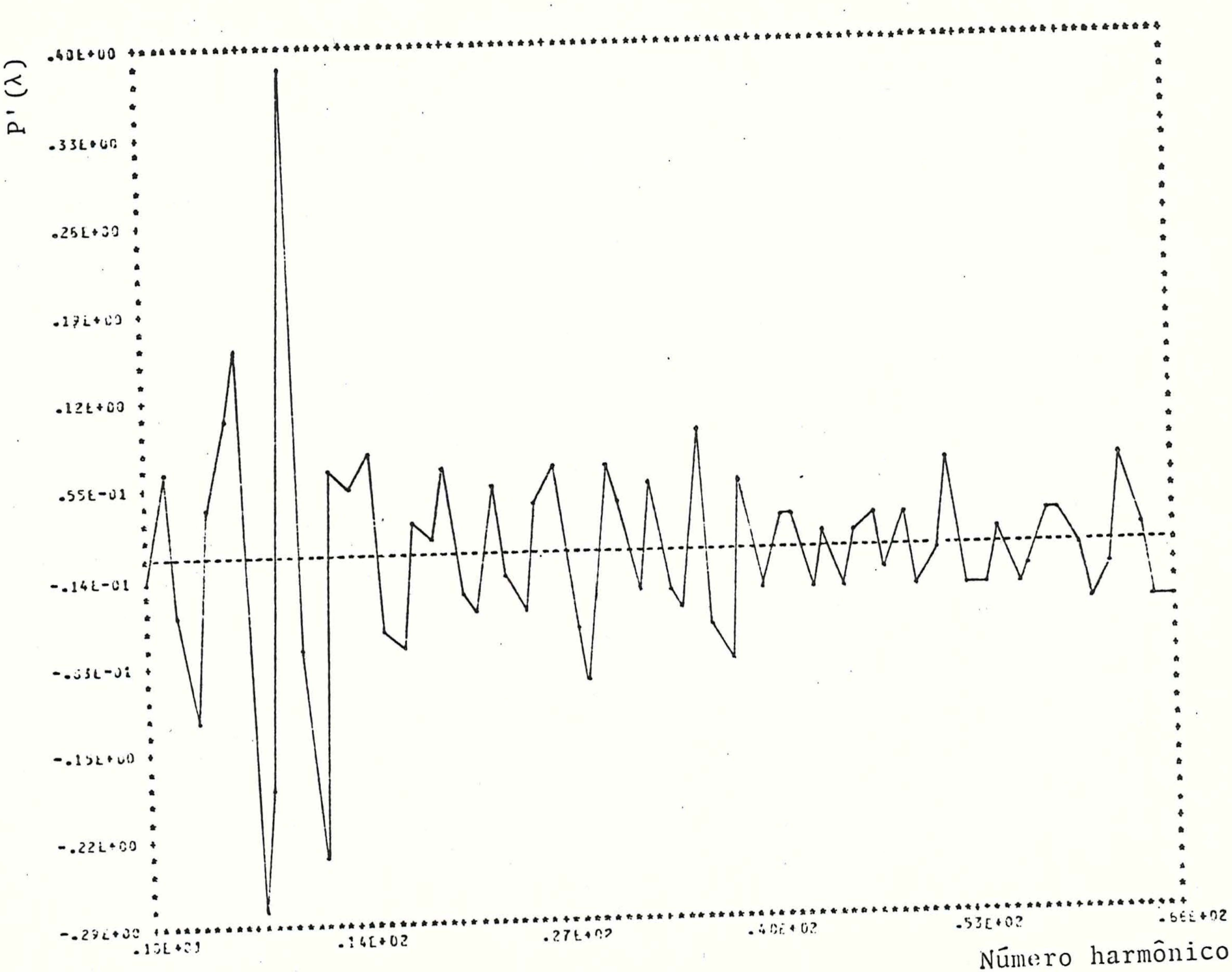



Figura 5.6 - Gráfico de P'(λ) para a série de precipitações

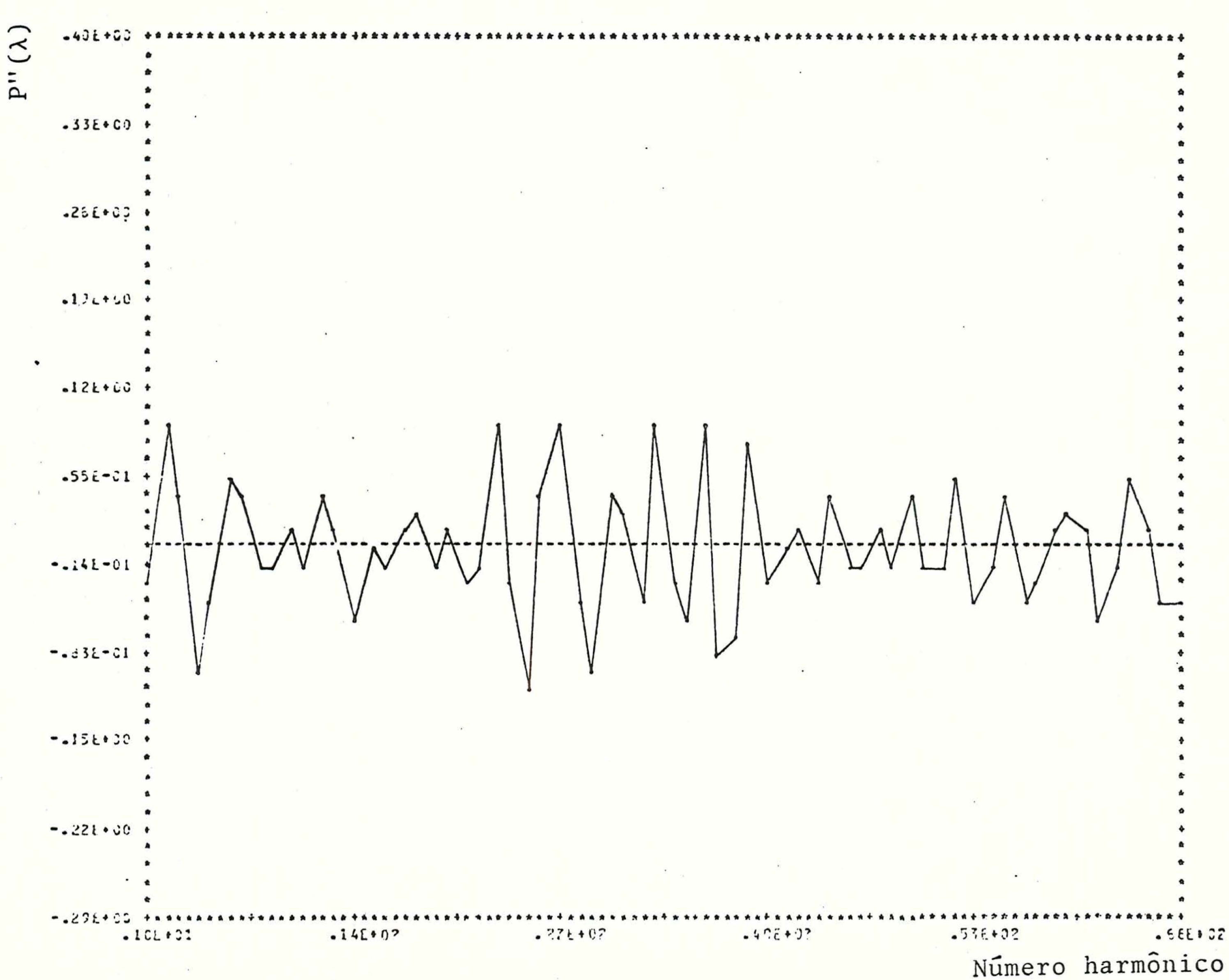



Figura 5.7 - Gráfico de P''(λ) para a série de precipitações

# CAPÍTULO 6

# SIMULAÇÃO

## 6.1 - Introdução

Neste capítulo verificaremos a validade dos métodos descritos nos Capítulos 4 e 5 quando aplicados a amostras finitas, simulando séries de estrutura conhecida e observando o comportamento de cada teste para dois comprimentos diferentes de simulação.

Apresentamos a seguir os principais resultados.

## 6.2 - O Processo $Y_t$ é Ruído Branco

O processo simulado foi:

$$X_t = Y_t + A\cos(\omega t), \quad t=1,\ldots,n \qquad (6.2.1)$$

onde $Y_t=\xi_t$, sendo $\xi_t$ um processo normal independente e identicamente distribuído com $E[\xi(t)]=0$ e $E[\xi^2(t)]=1$, $A=60$, $\omega=2\pi/12{,}5$ e $n=100$ e $500$.

### O Teste $P(\lambda)$

Aplicamos o teste $P(\lambda)$ com a escolha usual das funções peso como sendo o núcleo de Féjer e Dirichlet.

Os valores de m e n' foram 20 e 45, respectivamente.

O gráfico da função $P(\lambda)$ se encontra na Figura 6.1 e nele podemos observar a existência de um pico na frequência $\omega_8 = \frac{16\pi}{100}$

Então:

$$J_2 = \frac{[2\pi P(16\pi/100)+2\pi P(54\pi/100)]\times 1,2}{1,816} = 9,068$$

Ao nível de 0,1% temos max $J_q$=3,3 e portanto o resultado é altamente significante a este nível.

A amplitude da componente harmônica detectada pode ser estimada por:

$$\hat{A} = 65,277$$

Para n=500, obtemos de maneira análoga:

$$J_2 = 308,065$$

Novamente ao nível 0,1% o resultado é altamente significante e neste caso a amplitude estimada é dada por:

$$\hat{A} = 65,435$$

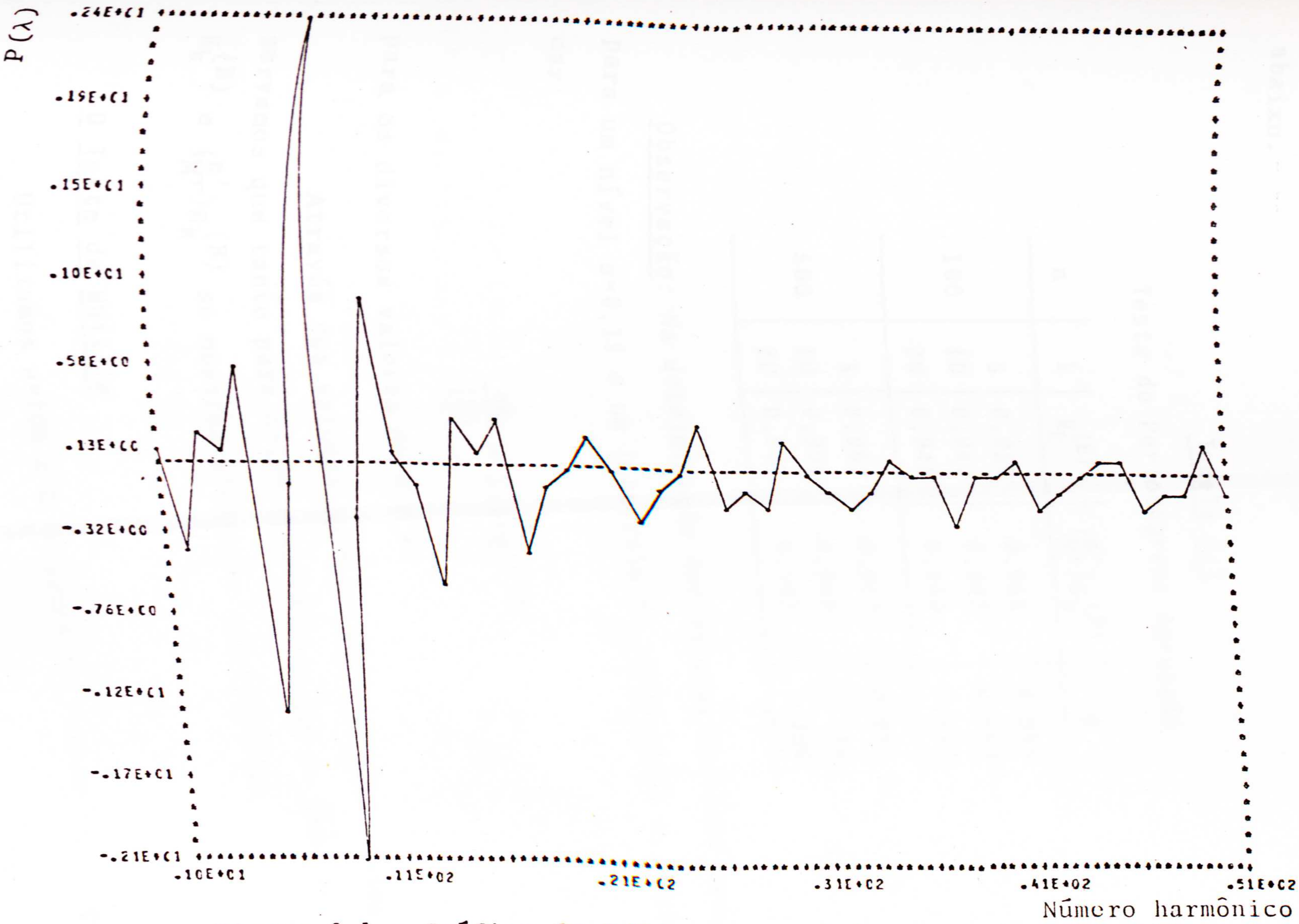
<u>O Teste do Periodograma Agrupado</u>

Aplicaremos agora o teste do periodograma agrupado às mesmas séries. Para cada uma aplicaremos o teste para três valores diferentes de k, o intervalo de agrupamento.

Para o cálculo do fator de correção $(\frac{m'}{M'})$ foi utilizada a função densidade espectral do processo ruído branco, dada por:

$$f(j) = \frac{1}{2\pi}, \qquad 0 \leqslant j \leqslant \frac{1}{2},$$

onde $j = \frac{i}{n}$, sendo i o número harmônico de interesse.




<u>Figura</u> <u>6.1</u> - Gráfico de P(λ) para a série simulada com n=100

Os valores de $g_k^{(B)}$ e $(\frac{m'}{M'})g_k^{(B)}$ são dados na Tabela 6.1, abaixo.

Tabela 6.1

Teste do Periodograma Agrupado

| n | k | $g_k^{(B)}$ | $(\frac{m'}{M'})g_k^{(B)}$ | g |
|---|---|---|---|---|
| 100 | 5 | 0,958 | 0,958 | 0,933 |
| | 10 | 0,942 | 0,942 | 0,699 |
| | 20 | 0,940 | 0,940 | 0,434 |
| 500 | 5 | 0,967 | 0,967 | 0,955 |
| | 10 | 0,966 | 0,966 | 0,749 |
| | 20 | 0,965 | 0,965 | 0,480 |

Observação: Na determinação dos valores críticos para $g_k^{(B)}$ para um nível $\alpha$=0,1% e um intervalo de agrupamento k, basta tomar:

$$\frac{\alpha k}{[\frac{n}{2}]} = k(1-g)^{k-1},$$

para os diversos valores de k e n.

Através dos valores apresentados na tabela acima, observamos que tanto para n=100 como n=500 todos os valores de $g_k^{(B)}$ e $(\frac{m'}{M'})g_k^{(B)}$ se mostram altamente significantes.

O Teste de Whittle

Utilizamos n=100 e $\ell$, o ponto de truncamento, igual a 20.

A maior ordenada do periodograma normalizado ocorre em $\omega_8=16\pi/100$, com:

$$I_8 = 14.293,751 \quad \text{e} \quad 2\pi\hat{f}(\omega_8) = 11.289,841$$

logo:

$$\max_p I_p / 2\pi\hat{f}(\omega_p) = 1,266.$$

Podemos então calcular g e nos referir à distribuição de Fisher com 50 graus de liberdade ou, alternativamente, realizar o teste através da distribuição assintótica apresentada no capítulo anterior, dada por (5.4.2).

Para n=100 e $\alpha$=0,1% temos que o valor crítico é dado aproximadamente por 21,639. Então, o resultado não é significante ao nível de 0,1%.

Para n=500, temos analogamente:

$$\max_p I_p / 2\pi\hat{f}(\omega_p) = 6,330.$$

Neste caso, o valor crítico é 24,858, donde concluímos que o resultado não é significativo ao nível de 0,1%.

<u>O Teste de Hannan</u>

Utilizamos n=100 e $\ell$, o ponto de truncamento, igual a 20.

A maior ordenada do periodograma ocorre em $\omega_8$=16$\pi$/100 e nos fornece g=0,67.

Calculando o valor crítico pela distribuição de Fisher com 50 graus de liberdade e $\alpha$=0,1% temos g=0,198, o que nos leva a aceitar a hipótese do espectro misto.

Para n=500, temos g=0,683 e o valor crítico g=0,049, para $\alpha=0,1\%$ e 250 graus de liberdade, que confirma o fato do espectro ser misto.

## 6.3 - O Processo $Y_t$ é Auto-regressivo de Primeira Ordem

As séries geradas foram da forma:

$$X_t = Y_t + A \cos(\omega t), \quad t=1,\dots,n, \qquad (6.3.1)$$

onde $A=60$, $\omega=2\pi/12,5$ e $Y_t - 0,7Y_{t-1} = \varepsilon_t$, $\varepsilon_t$ como definido em (6.2.1).

### O Teste $P(\lambda)$

Os valores de m e n foram 20 e 45, respectivamente.

Observando o gráfico de $P(\lambda)$ dado pela Figura 6.2, verificamos a existência de um pico na frequência $\lambda=16\pi/100$.

Logo, $J_2=9,536$.

Ao nível de 0,1% o valor crítico é 3,3 e o resultado se confirma altamente significante.

A amplitude da componente detectada pode ser estimada por:

$$\hat{A} = 65,25$$

Para n=500, analogamente temos: $J_2=806,57$, que é altamente significante ao nível $\alpha=0,1\%$.

### O Teste do Periodograma Agrupado

Aplicando o teste do periodograma agrupado às mesmas séries e com diferentes valores de k obtemos os resultados apresentados na Tabela 6.2.

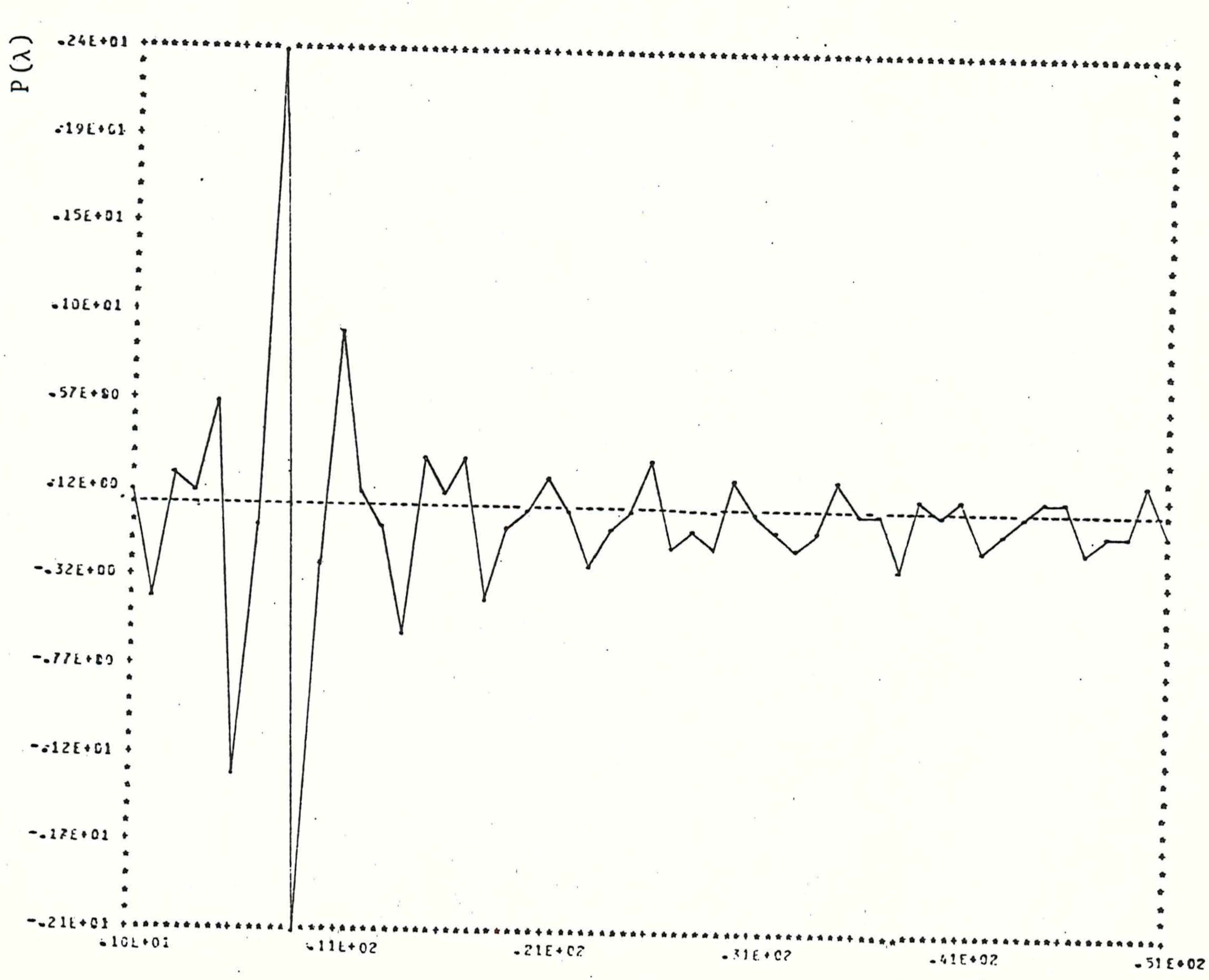



Figura 6.2 - Gráfico de P(λ) para a série simulada com n=100

Para o cálculo do fator de correção $(\frac{m'}{M'})$ foi utilizada a fórmula da função densidade espectral do processo auto-regressivo de primeira ordem, dada por:

$$f(j) = \frac{2}{1+\phi_1^2-2\phi_1 \cos(2\pi j)}, \quad 0 \leqslant j \leqslant \frac{1}{2},$$

onde $j = \frac{i}{n}$, sendo i o número harmônico de interesse.

Os valores críticos apresentados na última coluna foram obtidos através de: $\frac{\alpha k}{[\frac{n}{2}]} = k(1-g)^{k-1}$, onde $\alpha$=0,1%.

Tabela 6.2

Teste do Periodograma Agrupado

| n | k | $g_k^{(B)}$ | $(\frac{m'}{M'})g_k^{(B)}$ | g |
|---|---|---|---|---|
| | 5 | 0,957 | 0,519 | 0,933 |
| 100 | 10 | 0,941 | 0,233 | 0,699 |
| | 20 | 0,939 | 0,083 | 0,434 |
| | 5 | 0,972 | 0,855 | 0,955 |
| 500 | 10 | 0,941 | 0,714 | 0,749 |
| | 20 | 0,928 | 0,490 | 0,480 |

Através da tabela acima observamos que para n=100 todos os valores não corrigidos de $g_k^{(B)}$ são significantes, mas quando o fator de correção é aplicado nenhum valor permanece como significativo.

Entretanto, para n=500 todos os valores não corrigidos de $g_k^{(B)}$ são significantes e quando o fator de correção é aplicado um deles ainda é significante (k=20), nos levando a hipótese de que o espectro é misto.

<u>O Teste de Whittle</u>

Nesta aplicação utilizamos n=100 e $\ell$=20.

A maior ordenada do periodograma normalizado ocorre em $\omega_8=16\pi/100$, com:

$$I_8 = 14.315,760 \quad \text{e} \quad 2\pi\hat{f}(\omega_8) = 11.310,001$$

logo:

$$\max_p I_p/2\pi\hat{f}(\omega_p) = 1,265.$$

Calculando g através da distribuição assintótica, dada por (5.4.2), podemos realizar o teste, que para n=100 e $\alpha$=0,1% apresenta o valor crítico g como sendo aproximadamente 21,639.

Concluímos então que o resultado não é significativo ao nível de 0,1%, rejeitando a hipótese de que o espectro é misto.

Para n=500, temos analogamente

$$\max_p I_p/2\pi\hat{f}(\omega_p) = 6,327.$$

De (5.4.2) temos que o valor crítico é dado por 24,858, o que nos leva a concluir que o resultado não é significante ao nível de 0,1% e portanto rejeitamos a hipótese de que o espectro é misto.

<u>O Teste de Hannan</u>

Tomando-se $\ell$=20, observamos a maior ordenada do periodograma ocorrendo no harmônico 8, que nos fornece g=0,999.

Ao nível de 0,1% e com 50 graus de liberdade obtemos da Tabela de Fisher o valor crítico 0,198, que nos confirma o fato do espectro ser misto.

Para n=500 ainda temos g=0,999 e ao mesmo nível mas com 250 graus de liberdade obtemos o valor crítico 0,049 que confirma o fato anterior do espectro misto.

## 6.4 - O Processo $Y_t$ é Auto-regressivo de Segunda Ordem

As séries geradas foram da forma:

$$X_t = Y_t + A\cos(\omega t), \quad t=1,\ldots,n,$$

onde A=60, $\omega=2\pi/12,5$, n=100 e 500 e $Y_t - 0,7Y_{t-1} + 0,49Y_{t-2} = \varepsilon(t)$, sendo $\varepsilon(t)$ como definido em (6.1.1).

### O Teste $P(\lambda)$

Observando o gráfico de $P(\lambda)$ dado pela Figura 6.3, observamos um pico na frequência $\omega_8 = 16\pi/100$.

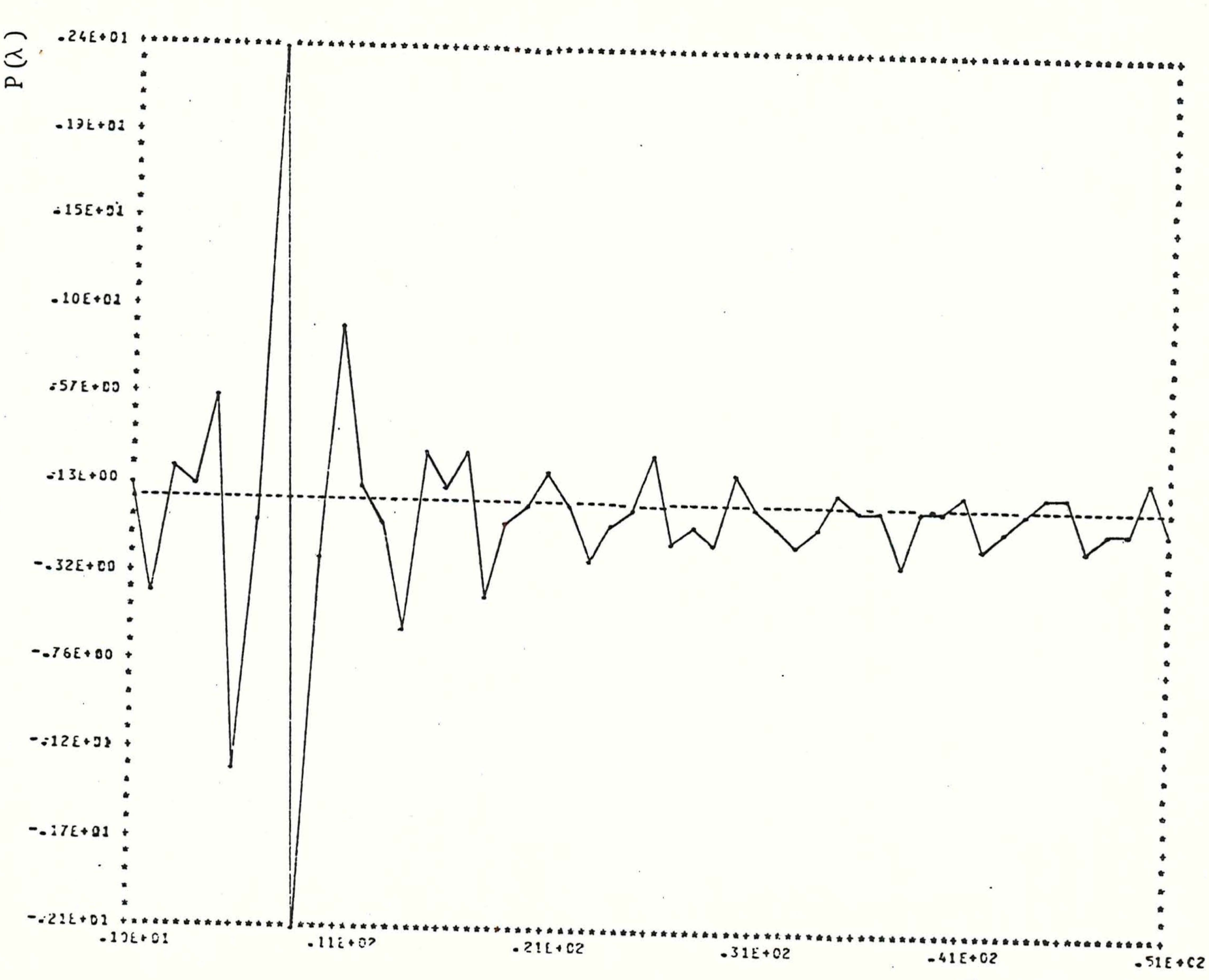
Logo: $J_2 = 56,32$.

Ao nível de 0,1% o valor crítico é 3,3 e o resultado é altamente significante.

A amplitude da componente harmônica pode ser estimada por:

$$\hat{A} = 65,19.$$

Para n=500, analogamente: $J_2 = 806,570$, que também é altamente significante ao nível de 0,1%.




<u>Figura</u> <u>6.3</u> - Gráfico de P(λ) para a série simulada com n=100

O Teste do Periodograma Agrupado

Aplicando o teste acima às mesmas séries e para k=5, 10,20, obtém-se os resultados apresentados na Tabela 6.3 abaixo.

Para o cálculo do fator de correção $(\frac{m'}{M'})$ foi utilizada a fórmula da função densidade espectral de um processo auto-regressivo de segunda ordem, dada por:

$$f(j) = \frac{2}{1+\phi_1^2+\phi_2^2-2\phi_1(1-\phi_2)\cos 2\pi j-2\phi_2 \cos 2\pi j}, \quad 0\leq j\leq \frac{1}{2},$$

onde $j=\frac{i}{n}$, sendo i o número harmônico do interesse.

Para o cálculo do valor crítico g, utilizamos: $\frac{\alpha k}{[\frac{n}{2}]}=k(1-g)^{k-1}$, onde $\alpha=0,1\%$.

Tabela 6.3

Teste do Periodograma Agrupado

| n | k | $g_k^{(B)}$ | $(\frac{m'}{M'})g_k^{(B)}$ | g |
|---|---|---|---|---|
| 100 | 5 | 0,957 | 0,797 | 0,933 |
| | 10 | 0,941 | 0,681 | 0,695 |
| | 20 | 0,931 | 0,429 | 0,434 |
| 500 | 5 | 0,967 | 0,934 | 0,955 |
| | 10 | 0,958 | 0,889 | 0,749 |
| | 20 | 0,954 | 0,827 | 0,480 |

Através dos valores apresentados na Tabela 6.2, observamos que para n=100 todos os valores não corrigidos de $g_k^{(B)}$ são significantes, mas quando o fator de correção é aplicado nenhum valor se revela significante ao nível de 0,1%.

Para n=500, observamosque todos os valores não corrigi dos de $g_k^{(B)}$ são significantes, e mesmo quando o fator de correção é aplicado todos continuam significantes, nos levando a aceitar a hipótese do espectro ser misto.

O Teste de Whittle

Nesta aplicação utilizamos n=100 e $\ell$=20.

A maior ordenada do periodograma normalizado ocorre em $\omega_8=16\pi/100$, com:

$$I_8 = 14.278{,}225 \quad \text{e} \quad 2\pi\hat{f}(\omega_8) = 51.628{,}714$$

logo:

$$\max_p I_p/2\pi\hat{f}(\omega_p) = 0{,}276.$$

Calculando o valor crítico por (5.4.2) para n=100 e $\alpha$=0,1% temos 21,639, o que nos leva a hipótese de que o espectro não é misto.

Para n=500, temos analogamente:

$$\max_p I_p/2\pi\hat{f}(\omega_p) = 6{,}327$$

Neste caso temos que o valor crítico é dado por 24,858, o que nos leva a rejeitar a hipótese do espectro ser misto.

O Teste de Hannan

Utilizando $\ell$=20, observamos que a maior ordenada do periodograma ocorre em $\omega_8=16\pi/100$ e obtemos g=1,0.

Calculamos o valor crítico pela distribuição de Fisher com 50 graus de liberdade e $\alpha$=1%, e obtem-se $g_{crítico}$=0,198, o que nos leva ao fato do espectro ser misto.

Analogamente para n=500, temos g=0,999 e o valor crítico para $\alpha$=1% e 250 graus de liberdade, donde $g_{crítico}$=0,049 que confirma o fato do espectro ser misto.

## 6.5 - Conclusões

Analisando os resultados para as séries simuladas concluimos que o teste $P(\lambda)$ e o teste do periodograma agrupado são razoavelmente eficientes, enquanto que o teste de Whittle falha ao não detectar a existência de uma componente periódica muito forte.

Comparando o teste $P(\lambda)$ com o teste do periodograma agrupado observamos que o teste $P(\lambda)$ é ligeiramente mais eficiente, pois todos os valores não corrigidos de $g_k^{(B)}$ são significantes ao nível de 0,1%, enquanto que ao utilizar os valores corrigidos $(\frac{m'}{M'})g_k^{(B)}$ alguns deles falham ao detectar a existência de tais componentes.

Surge também uma dúvida ao constatar que $g_5^{(B)}$ e $g_{10}^{(B)}$ não são significantes, enquanto $g_{20}^{(B)}$ é significante no caso do processo auto-regressivo de primeira ordem, pois ao tomar k=20 estamos admitindo que a função densidade espectral seja constante numa região grande, e quando dimimuimos esta região (k=5, 10) o teste não mais acusa a existência da componente harmônica.

Um ponto que fica claro nesta análise é a importância do fator de correção $\frac{m'}{M'}$.

Embora o verdadeiro valor da estatística do teste seja, em geral, maior que $(\frac{m'}{M'}\, g_k^{(B)})$, nos parece perigoso basear o teste puramente no valor de $g_k^{(B)}$.

## DISCUSSÃO E COMENTÁRIOS

Dos testes apresentados concluimos que o teste $P(\lambda)$ e o teste do periodograma agrupado são os que obtém melhor desempenho na detectação de termos periódicos ocultos.

Ao aplicar o teste do periodograma agrupado devemos observar com cautela a suposição de que a função densidade espectral seja constante numa região de largura k, pois em função do valor escolhido para k esta hipótese pode ser muito restritiva.

Também, ao rejeitarmos a hipótese nula para $k=k_1$ e aceitarmos para $k=k_2$, $k_1<k_2$, devemos levar o fato acima em consideração.

Outro detalhe importante é o fator de correção $(\frac{m'}{M'})$ da estatística $g_k^{(B)}$; apesar de $(\frac{m'}{M'})g_k^{(B)}$ superar o valor da estatística $T_k$ do teste, seria perigoso basear o teste apenas em $g_k^{(B)}$, pois aceitaríamos valores que nem sempre corresponderiam a componentes periódicas.

Ao realizar o teste de Priestley, um aspecto a ser analisado é a escolha de m e n', os pontos de truncamento das janelas escolhidas (geralmente m=15% de n e n'=30% de n), a fim de se obter estabilidade na função de covariância e, consequentemen te, em $P(\lambda)$.

Um outro aspecto a ser considerado é a utilização do método de mínimos quadrados para recalcular a covariância após remover o efeito de uma componente harmônica detectada pelo teste $P(\lambda)$, para séries pequenas a fim de evitar instabilidades em $P(\lambda)$.

O teste $P(\lambda)$, apesar de mais poderoso é mais trabalhoso para ser aplicado, pois devemos remover o efeito da componente observada antes de realizar o teste para um novo pico.

## BIBLIOGRAFIA


ANDERSON, T.W. 1971. *The statistical analysis of time series.* New York, John Wiley. 704p. (Wiley Series in Probability and Mathematical Statistics)

BARTLETT, M.S. 1955. *An introduction to the stochastic process with special reference to methods and applications.* Cambridge, University Press. 312p.

BHANSALI, R.J. 1977. An application of Priestley's P(λ) test to the southend tide heights data. In: BARRA, J.R., ed. *Recent developments in statistics.* Amsterdam, North-Holland. p.351-359.

BHANSALI, R.J. 1979. A mixed spectrum analysis of the Lynx data. *Journal of the Royal Statistical Society, Series A,* *142*(2):199-209.

BHANSALI, R.J. & SARHAN, S.H. 1982. Effects of the presence of a harmonic term on the spectral factorisation procedure. *Mathematische Operationsforschung und Statistik,* *13*(2):317-349.

BLOOMFIELD, P. 1976. *Fourier analysis of time series:* an introduction. New York, John Wiley. 258p. (Wiley Series in Probability and Mathematical Statistics)

BRILLINGER, D.R. c1975. *Time series, data analysis and theory.* New York, Holt. 500p.

CAMPBELL, M.J. & WALKER, A.M. 1977. A survey of statistical work on the Mackenzie river series of annual Canadian Lynx trappings for the years 1821-1934 and a new analysis. *Journal of the Royal Statistical Society, Series A, 140*(4): 411-431.

FIGUEIREDO, D.G. de. c1977. *Analise de Fourier e equações diferenciais parciais.* Rio de Janeiro, IMPA. 274p. (Projeto Euclides)

FISHER, R.A. 1929. Tests of significance in harmonic analysis. *Proceedings of the Royal Society of London, Series A, 130*:54-59.

GRENANDER, U. & ROSEMBLATT, M. c1957. *Statistical analysis of stationary time series.* New York, John Wiley. 300p.

HANNAN, E.J. 1960. *Time series analysis.* London, Methuen. 152p.

HANNAN, E.J. 1961. Testing for a jump in the spectral function. *Journal of the Royal Statistical Society, Series B, 23*(2):394-404.

HANNAN, E.J. 1970. *Multiple time series.* New York, John Wiley. 536p. (Wiley Series in Probability and Mathematical Statistics)

JENKINS, G.M. & PRIESTLEY, M.B. 1957. The spectral analysis of time series. *Journal of the Royal Statistical Society, Series B, 19*(1):1-12.

JENKINS, G.M. & WATTS, D.G. c1968. *Spectral analysis and its applications.* San Francisco, Holden-Day. 525p.

KANTOR, I.I. 1982. *Previsibilidade da série de precipitação de chuvas de Fortaleza pelo método de máxima entropia de Burg.* São José dos Campos, INPE. (Publicação RPE/420)

KOOPMANS, L.H. 1974. *The spectral analysis of time series.* New York, Academic Press. 366p. (Probability and Mathematical Statistics, 22)

MACNEILL, I.B. 1974. Test for periodic components in multiple time series. *Biometrika, 61*(1):57-70.

MACNEILL, I.B. 1977. A test of whether several time series share common periodicities. *Biometrika, 64*(3):495-508.

MORETTIN, P.A. 1979. *Analise harmônica de processos estocasticos.* Rio de Janeiro, IMPA. 176p.

NOWROOZI, A.A. 1967. Table for Fisher's test of significance in harmonic analysis. *Geophysical Journal of the Royal Astronomical Society, 12*:517-520.

PRIESTLEY, M.B. 1957. Analysis of geophysical time series: discussion. *Journal of Royal Statistical Society, Series A, 120*(4):432-435.

PRIESTLEY, M.B. 1962. The analysis of stationary processes with mixed spectra. I, II. *Journal of the Royal Statistical Society, Series B, 24*(1):215-233; *24*(2):511-529.

PRIESTLEY, M.B. 1964. Estimation of the spectral density function in the presence of harmonic components. *Journal of the Royal Statistical Society, Series B*, 26(1):123-132.

PRIESTLEY, M.B. 1981. *Spectral analysis and time series*. New York, Academic Press. v.1

SHIMSHONI, M. 1971. On Fisher's test of significance in harmonic analysis. *Geophysical Journal of the Royal Astronomical Society*, 23:373-377.

WHITTLE, P. 1952. The simultaneous estimation of a time series harmonic components and covariance structure. *Trabajos de Estadistica*, 3:43-57.

WHITTLE, P. 1954. The statistical analysis of a seiche record. *Journal of Marine Research*, 13(1):76-100.

## BIBLIOGRAFIA ADICIONAL


BARTLETT, M.S. 1950. Periodogram analysis and continuous spectra. *Biometrika*, 37:1-16.

DANIELL, P.J. 1946. Symposium on autocorrelation in time series: discussion. *Journal of the Royal Statistical Society, Supplement*, 8:88-90.

HARTLEY, H.O. 1949. Tests of significance in harmonic analysis. *Biometrika*, 36:194-201.

JEFFREYS, H. 1929. *The earth.* Cambridge, University Press. 526p.

MORAN, P.A.P. 1953a. The statistical analysis of the canadian Lynx cicle I. *Australian Journal of Zoology*, 1:163-173.

MORAN, P.A.P. 1953b. The statistical analysis of the canadian Lynx cicle II. *Australian Journal of Zoology*, 1:291-298.

PARZEN, E. 1957a. On consistent estimates of the spectrum of a stationary time series. *The Annals of Mathematical Statistics*, 28:329-348.

PARZEN, E. 1957b. On choosing an estimate of the spectral density function of a stationary time series. *The Annals of Mathematical Statistics*, *28*:921-932.

PARZEN, E. 1958. On asymptotically efficient consistent estimates of the spectral density function of a stationary time series. *Journal of the Royal Statistical Society, Series B*, *20*:303-322.

SCHUSTER, A. 1898. On the investigation of hidden periodicities with application to the supposed 26 day period of metereological phenomena. *Terrestrial Magnetism and Atmospheric Electricity*, 3:13-41.

Tabela A.1

Temperatura em São Paulo
(Média de 24 leituras diárias)

| MESES \ ANOS | 1949 | 1950 | 1951 | 1952 | 1953 | 1954 | 1955 | 1956 | 1957 |
|---|---|---|---|---|---|---|---|---|---|
| Janeiro | - | 20,6 | 20,5 | 20,8 | 21,9 | 22,1 | 21,1 | 23,4 | 21,3 |
| Fevereiro | - | 20,9 | 20,7 | 20,1 | 20,7 | 21,5 | 21,9 | 21,2 | 20,4 |
| Março | - | 20,6 | 20,0 | 20,5 | 20,9 | 21,0 | 20,7 | 20,3 | 20,4 |
| Abril | - | 18,8 | 15,9 | 17,1 | 18,1 | 18,3 | 18,6 | 18,1 | 18,1 |
| Maio | - | 17,5 | 15,7 | 16,7 | 16,9 | 16,5 | 15,7 | 15,1 | 15,9 |
| Junho | - | 16,1 | 14,6 | 15,2 | 15,5 | 16,3 | 14,9 | 13,5 | 15,0 |
| Julho | 14,5 | 14,7 | 13,2 | 14,8 | 13,3 | 15,7 | 14,7 | 14,0 | - |
| Agosto | 15,6 | 17,0 | 14,0 | 16,8 | 15,8 | 16,2 | 15,5 | 13,6 | - |
| Setembro | 16,0 | 17,0 | 16,3 | 17,9 | 17,8 | 17,7 | 16,6 | 17,6 | - |
| Outubro | 17,7 | 16,9 | 17,5 | 18,0 | 18,4 | 17,7 | 16,4 | 17,0 | - |
| Novembro | 17,6 | 17,8 | 19,0 | 19,1 | 18,6 | 18,5 | 16,9 | 16,7 | - |
| Dezembro | 19,6 | 19,9 | 18,4 | 19,8 | 19,4 | 19,6 | 20,2 | 19,1 | - |

Fonte: IAG-USP

Tabela A.2

Ano (Coluna 1); Número de linces mortos = y (Coluna 2); $X=\log_{10}y$ (Coluna 3)

| 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1821 | 269 | 2.430 | 1850 | 361 | 2.557 | 1879 | 201 | 2.303 | 1907 | 1836 | 3.264 |
| 1822 | 321 | 2.506 | 1851 | 377 | 2.576 | 1880 | 229 | 2.360 | 1908 | 345 | 2.538 |
| 1823 | 585 | 2.767 | 1852 | 225 | 2.352 | 1881 | 469 | 2.671 | 1909 | 382 | 2.582 |
| 1824 | 871 | 2.940 | 1853 | 360 | 2.556 | 1882 | 736 | 2.867 | 1910 | 808 | 2.907 |
| 1825 | 1475 | 3.169 | 1854 | 731 | 2.864 | 1883 | 2042 | 3.310 | 1911 | 1388 | 3.142 |
| 1826 | 2821 | 3.450 | 1855 | 1638 | 3.214 | 1884 | 2811 | 3.449 | 1912 | 2713 | 3.433 |
| 1827 | 3928 | 3.594 | 1856 | 2725 | 3.435 | 1885 | 4431 | 3.646 | 1913 | 3800 | 3.580 |
| 1828 | 5943 | 3.774 | 1857 | 2871 | 3.458 | 1886 | 2511 | 3.400 | 1914 | 3091 | 3.490 |
| 1829 | 4950 | 3.695 | 1858 | 2119 | 3.326 | 1887 | 389 | 2.590 | 1915 | 2985 | 3.475 |
| 1830 | 2577 | 3.411 | 1859 | 684 | 2.835 | 1888 | 73 | 1.863 | 1916 | 3790 | 3.579 |
| 1831 | 523 | 2.718 | 1860 | 299 | 2.476 | 1889 | 39 | 1.591 | 1917 | 674 | 2.829 |
| 1832 | 98 | 1.991 | 1861 | 236 | 2.373 | 1890 | 49 | 1.690 | 1918 | 81 | 1.909 |
| 1833 | 184 | 2.265 | 1862 | 245 | 2.389 | 1891 | 59 | 1.771 | 1919 | 80 | 1.903 |
| 1834 | 279 | 2.446 | 1863 | 552 | 2.742 | 1892 | 188 | 2.274 | 1920 | 108 | 2.033 |
| 1835 | 409 | 2.612 | 1864 | 1623 | 3.210 | 1893 | 377 | 2.576 | 1921 | 229 | 2.360 |
| 1836 | 2285 | 3.359 | 1865 | 3311 | 3.520 | 1894 | 1292 | 3.111 | 1922 | 399 | 2.601 |
| 1837 | 2685 | 3.429 | 1866 | 6721 | 3.828 | 1895 | 4031 | 3.605 | 1923 | 1132 | 3.054 |
| 1838 | 3409 | 3.533 | 1867 | 4254 | 3.628 | 1896 | 3495 | 3.543 | 1924 | 2432 | 3.386 |
| 1839 | 1824 | 3.261 | 1868 | 687 | 2.837 | 1897 | 587 | 2.769 | 1925 | 3574 | 3.553 |
| 1840 | 409 | 2.612 | 1869 | 255 | 2.406 | 1898 | 105 | 2.021 | 1926 | 2935 | 3.468 |
| 1841 | 151 | 2.179 | 1870 | 473 | 2.675 | 1899 | 153 | 2.185 | 1927 | 1537 | 3.187 |
| 1842 | 45 | 1.653 | 1871 | 358 | 2.554 | 1900 | 387 | 2.588 | 1928 | 529 | 2.723 |
| 1843 | 68 | 1.832 | 1872 | 784 | 2.894 | 1901 | 758 | 2.880 | 1929 | 485 | 2.686 |
| 1844 | 213 | 2.328 | 1873 | 1594 | 3.202 | 1902 | 1307 | 3.115 | 1930 | 662 | 2.821 |
| 1845 | 546 | 2.737 | 1874 | 1676 | 3.224 | 1903 | 3465 | 3.540 | 1931 | 1000 | 3.000 |
| 1846 | 1033 | 3.014 | 1875 | 2251 | 3.352 | 1904 | 6991 | 3.845 | 1932 | 1590 | 3.201 |
| 1847 | 2129 | 3.328 | 1876 | 1426 | 3.154 | 1905 | 6313 | 3.800 | 1933 | 2657 | 3.424 |
| 1848 | 2536 | 3.404 | 1877 | 756 | 2.878 | 1906 | 3794 | 3.579 | 1934 | 3396 | 3.531 |
| 1849 | 957 | 2.981 | 1878 | 299 | 2.476 | | | | | | |

Fonte: JRSS A, 140, 411-431.

# Tabela A.3

## Série de Precipitação Anual de Chuvas em Fortaleza, Ceará

| Ano | Precipitação | Ano | Precipitação | Ano | Precipitação | Ano | Precipitação | Ano | Precipitação |
|---|---|---|---|---|---|---|---|---|---|
| 1849 | 2001 | 1876 | 1569 | 1902 | 878 | 1928 | 995 | 1954 | 1032 |
| 1850 | 852 | 1877 | 468 | 1903 | 789 | 1929 | 1230 | 1955 | 1152 |
| 1851 | 1806 | 1878 | 503 | 1904 | 1136 | 1930 | 1107 | 1956 | 806 |
| 1852 | 1356 | 1879 | 597 | 1905 | 1189 | 1931 | 1133 | 1957 | 1225 |
| 1853 | 1233 | 1880 | 1539 | 1906 | 1430 | 1932 | 879 | 1958 | 504 |
| 1854 | 1590 | 1881 | 1423 | 1907 | 697 | 1933 | 937 | 1959 | 1493 |
| 1855 | 1273 | 1882 | 1246 | 1908 | 834 | 1934 | 1888 | 1960 | 1011 |
| 1856 | 1770 | 1883 | 1508 | 1909 | 1015 | 1935 | 1661 | 1961 | 1737 |
| 1857 | 1734 | 1884 | 1047 | 1910 | 2051 | 1936 | 820 | 1962 | 1258 |
| 1858 | 1457 | 1885 | 1307 | 1911 | 1373 | 1937 | 1313 | 1963 | 2102 |
| 1859 | 1357 | 1886 | 1399 | 1912 | 2446 | 1938 | 1586 | 1964 | 2428 |
| 1860 | 1716 | 1887 | 1320 | 1913 | 1905 | 1939 | 1911 | 1965 | 1630 |
| 1861 | 1445 | 1888 | 736 | 1914 | 1512 | 1940 | 1447 | 1966 | 1288 |
| 1862 | 1468 | 1889 | 784 | 1915 | 530 | 1941 | 916 | 1967 | 1839 |
| 1863 | 1452 | 1890 | 1534 | 1916 | 1328 | 1942 | 780 | 1968 | 1385 |
| 1864 | 1098 | 1891 | 1077 | 1917 | 2077 | 1943 | 1042 | 1969 | 1805 |
| 1865 | 1238 | 1892 | 1211 | 1918 | 1319 | 1944 | 1090 | 1970 | 1192 |
| 1866 | 2478 | 1893 | 1430 | 1919 | 656 | 1945 | 1750 | 1971 | 2093 |
| 1867 | 832 | 1894 | 2505 | 1920 | 1847 | 1946 | 1724 | 1972 | 1299 |
| 1868 | 1289 | 1895 | 2491 | 1921 | 2496 | 1947 | 1726 | 1973 | 2331 |
| 1869 | 1470 | 1896 | 2144 | 1922 | 1595 | 1948 | 1384 | 1974 | 2512 |
| 1870 | 1628 | 1897 | 1839 | 1923 | 1513 | 1949 | 1881 | 1975 | 1778 |
| 1871 | 1459 | 1898 | 863 | 1924 | 1847 | 1950 | 1114 | 1976 | 1417 |
| 1872 | 2256 | 1899 | 2414 | 1925 | 1137 | 1951 | 747 | 1977 | 1941 |
| 1873 | 2058 | 1900 | 940 | 1926 | 1571 | 1952 | 1378 | 1978 | 1752 |
| 1874 | 1487 | 1901 | 1545 | 1927 | 1195 | 1953 | 1068 | 1979 | 996 |
| 1875 | 1581 | | | | | | | | |

Fonte: Kantor, I.J. (1982)

TESTEF ON DISK
===============

```
C     ESTE PROGRAMA CALCULA O PERIODOGRAMA,UTILIZANDO O ALGORITMO DE****
C     TUKEY (FFT) E OBTEM A ESTATISTICA DE FISHER PARA DETECTAR ********
C     PERIODICIDADES OCULTAS.*******************************************
C     X - SERIE ORIGINAL
C     N - COMPRIMENTO DA SERIE ORIGINAL

      DIMENSION X(1024),Y(500),IWK(513),WK(513),ICHAR(10),ITITLE(144)
      DIMENSION IMAG4(5151),RANGE(4),Z(513)
      COMPLEX XT(513)
      DATA PI/3.14159/,ICHAR(1)/1H./,RANGE/4*0.0/,P/0.20/
      READ(5,1) (ITITLE(J),J=1,144)

      READ(5,/) I1,I2
      002:0017:1  IS THE LOCATION FOR EXCEPTIONAL ACTION ON THE I/O STATEMENT
      READ(5,/) N
      002:0020:1  IS THE LOCATION FOR EXCEPTIONAL ACTION ON THE I/O STATEMENT
      READ(5,/) (X(I),I=1,N)
      002:002F:1  IS THE LOCATION FOR EXCEPTIONAL ACTION ON THE I/O STATEMENT
      DO 5 I=1,N
    5 Z(I)=I
      CALL USPLT(Z,X,N,N,1,1,ITITLE,RANGE,ICHAR,1,IMAG4,IER)
C     SUAVIZA A SERIE ATRAVES DA JANELA COSENO
      CALL DETRND(X,N,0)
      CALL TAPER(PI,X,N,P)
C     CALCULA O PERIODOGRAMA INTRODUZINDO O ALGORITMO FFT
      CALL FFTRC(X,N,XT,IWK,WK)
      NPGM=N/2+1
      CON=2*PI*N
      DO 10 I=1,NPGM
      XT(I)=CONJG(XT(I))
      X(I)=(REAL(XT(I))**2+AIMAG(XT(I))**2)/CON
      Y(I)=I
   10 PRINT*/,I,X(I)

      CALL USPLT(Y,X,NPGM,NPGM,1,1,ITITLE,RANGE,ICHAR,1,IMAG4,IER)
      SOM=0.0
      DO 15 I=I1,I2
   15 SOM=SOM+X(I)
      DO 20 I=I1,I2
      WK(I)=X(I)/SOM
   20 PRINT*/,I,WK(I)
C     CALCULA A ESTATISTICA DE FISHER
      SOMA=0.0
      DO 30 I=1,NPGM
   30 SOMA=SOMA+X(I)
      DO 40 I=1,NPGM
      X(I)=X(I)/SOMA
      IO=I-1
   40 WRITE(6,3) IO,X(I)
      PRINT*/,SOMA
    1 FORMAT(72A1)
    3 FORMAT(1X,"G(",I3,")=",F7.3)
      STOP
      END
```

==========================================================================

```
      SUBROUTINE DETRND(X,N,K)
      DIMENSION X(1024)
      SUMX=0.0
      DO 200 I=1,N
  200 SUMX=SUMX+X(I)
      XBAR=SUMX/FLOAT(N)
      DO 300 I=1,N
```

```
300 X(I)=X(I)-XBAR
    IF(K.LE.0) RETURN
    TBAR=FLOAT(N+1)/2.0
    SUMTT=FLOAT(N*(N*N-1))/12.0
    SUMTX=0
    DO 400 I=1,N
400 SUMTX=SUMTX+X(I)*(FLOAT(I)-TBAR)
    BETA=SUMTX/SUMTT
    DO 500 I=1,N
500 X(I)=X(I)-BETA*(FLOAT(I)-TBAR)
    RETURN
    END

=============================================================================

    SUBROUTINE TAPER(PI,X,N,P)
    DIMENSION X(1024)
    IF((P.LE.0).OR.(P.GT.1.0)) RETURN
    M=INT(P*FLOAT(N)+0.5)/2
    DO 600 I=1,M
    WEIGHT=0.5-0.5*COS(PI*(FLOAT(I)-0.5)/FLOAT(M))
    X(I)=X(I)*WEIGHT
    X(N+1-I)=X(N+1-I)*WEIGHT
600 CONTINUE
    RETURN
    END

=============================================================================

                                              W H I T T L E   O N   D I S K
                                              = = = = = = = = = = = = = = =

    DIMENSION X(600),CX(300),CXX(300),F(300),IWK(500),WK(500)
    COMPLEX XT(513)
    DATA PI/3.14159/
    READ(5,/) N,L

    002:0003:1  IS THE LOCATION FOR EXCEPTIONAL ACTION ON THE I/O STATEMENT
    READ(5,/) (X(I),I=1,N)
    002:001A:1  IS THE LOCATION FOR EXCEPTIONAL ACTION ON THE I/O STATEMENT
    CALL FISHM(CX,X,N,L)
    CALL FFTRC(X,N,XT,IWK,WK)
    NPCM=N/2+1
    CON=2*PI*N
    DO 10 I=1,NPCM
    XT(I)=CONJG(XT(I))
 10 X(I)=(REAL(XT(I))**2+AIMAG(XT(I))**2)/CON
    DO 20 LL=1,L-1
 20 CXX(LL)=CX(L+1)
    SOMO=0.0
    DO 40 I=2,NPCM
    SOMA=0.0
    DO 30 K=1,LL
 30 SOMA=SOMA+CXX(K)*COS((2*PI*I*K)/N)
    F(I)=CX(1)+2*SOMA
    F(I)=X(I)/F(I)
 40 SOMO=SOMO+F(I)
    DO 50 I=2,NPCM
    F(I)=F(I)/SOMO
    II=I-1
 50 WRITE(6,1) II,F(I)

  1 FORMAT(4X,"G(",I2,")=",F7.3)
    STOP
    END
```

==========================================================================

```
      SUBROUTINE FISHM(CX,X,N,L)
      DIMENSION CX(300),X(600)
      XM=0.0
      DO 100 II=1,N
  100 XM=XM+X(II)
      XM=XM/N
      DO 300 II=1,L
      CX(II)=0.0
      DO 200 KK=1,N-II+1
      KI=KK+II-1
  200 CX(II)=CX(II)+(X(KK)-XM)*(X(KI)-XM)
  300 CX(II)=CX(II)/N
      RETURN
      END
```

H A N N A N N   O N   D I S K
= = = = = = = = = = = = = = =

```
      DIMENSION X(600),CX(300),CXX(300),F(300),IWK(500),WK(500)
      COMPLEX XT(513)
      DATA PI/3.14159/
      READ(5,/) N,L

      002:000B:1  IS THE LOCATION FOR EXCEPTIONAL ACTION ON THE I/O STATEMENT
      READ(5,/) (X(I),I=1,N)
      002:001A:1  IS THE LOCATION FOR EXCEPTIONAL ACTION ON THE I/O STATEMENT
      CALL FISHM(CX,X,N,L)
      CALL FFTRC(X,N,XT,IWK,WK)
      NPGM=N/2+1
      CON=2*PI*N
      DO 10 I=1,NPGM
      XT(I)=CONJG(XT(I))
   10 X(I)=(REAL(XT(I))**2+AIMAG(XT(I))**2)/CON
      DO 20 LL=1,L-1
   20 CXX(LL)=CX(L+1)
      SOMO=0.0
      DO 40 I=2,NPGM
      SOMA=0.0
      DO 30 K=1,LL
      PESO=(1-K/N)*(1+COS(PI*FLOAT(K)/FLOAT(L)))
   30 SOMA=SOMA+CXX(K)*PESO*COS((2*PI*K*I)/N)
      F(I)=CX(1)+SOMA
      F(I)=F(I)-(L/(2*N))*X(I)
      F(I)=F(I)/(1-L/N)
      F(I)=X(I)/F(I)
   40 SOMO=SOMO+F(I)
      DO 50 I=2,NPGM
      F(I)=F(I)/SOMO
      II=I-1
   50 WRITE(6,1) II,F(I)

    1 FORMAT(4X,"G(",I2,")=",F7.3)
      STOP
      END
```

==========================================================================

```
      SUBROUTINE FISHM(CX,X,N,L)
      DIMENSION CX(300),X(600)
      XM=0.0
```

```
      DO 100 II=1,N
100.  XM=XM+X(II)
      XM=XM/N
      DO 300 II=1,L
      CX(II)=0.0
      DO 200 KK=1,N-II+1
      KI=KK+II-1
200   CX(II)=CX(II)+(X(KK)-XM)*(X(KI)-XM)
300   CX(II)=CX(II)/N
      RETURN
      END
```

==========================================================================

P R I L S T   O N   D I S K
= = = = = = = = = = = = = = =

```
      DIMENSION X(900),CX(300),CXX(300),F(300),ICHAR(10),ITITLE(144)
      DIMENSION IMAG4(5252),RANGE(4),XC(600)
      DATA PI/3.14159/,ICHAR(1)/1H./,RANGE/4*0.0/
      READ(5,60) (ITITLE(I),I=1,144)

      READ(5,/) L1,L2
      002:0017:1  IS THE LOCATION FOR EXCEPTIONAL ACTION ON THE I/O STATEMENT
      READ(5,/) N
      002:0020:1  IS THE LOCATION FOR EXCEPTIONAL ACTION ON THE I/O STATEMENT
      READ(5,/) (X(I),I=1,N)
      002:002F:1  IS THE LOCATION FOR EXCEPTIONAL ACTION ON THE I/O STATEMENT
      CALL FISHM(CX,X,N,L1)
      NP=N/2+1
      PRINT*/,N,L1,L2

      WRITE(6,70)
      DO 30 I=1,NP
      SOM1=0.0
      DO 10 K=2,L1
      ARG=PI*FLOAT(2*I*(K-1))/FLOAT(N)
      CXX(K)=CX(K)*COS(ARG)
 10   SOM1=SOM1+(1-FLOAT(K-1)/FLOAT(N))*CXX(K)
      SOM2=0.0
      DO 20 K=2,L2
 20   SOM2=SOM2+CXX(K)
      XC(I)=I
      SOM1=CX(1)/(2*PI)+SOM1/PI
      SOM2=CX(1)/(2*PI)+SOM2/PI
      F(I)=(SOM1-SOM2)/CX(1)
 30   PRINT*/,I,F(I)
      CALL USPLT(XC,F,NP,NP,1,1,ITITLE,RANGE,ICHAR,1,IMAG4,IER)
      SOM2=0.0
      DO 40 K=2,2*L2
      CX(K)=CX(K)/CX(1)
 40   SOM2=SOM2+CX(K)**2
      SOM1=0.0
      DO 50 K=2,L2
 50   SOM1=SOM1+CX(K)**2
      GPI=1+4*SOM1-2*SOM2
      PRINT*/,GPI
      GPI=GPI/2
      SQGPI=SQRT(GPI)
      PRINT*/,SQGPI
      DELTA=(2*L1)/3-2*L2+(2*(L2**2))/L1
      PRINT*/,DELTA
 60   FORMAT(72A1)
 70   FORMAT(10X,"VALORES DE F LAMBDA")
      STOP
      END
```

==========================================================================

```
      SUBROUTINE FISHM(CX,X,N,L)
      DIMENSION CX(300),X(600)
      XM=0.0
      DO 100 II=1,N
  100 XM=XM+X(II)
      XM=XM/N
      DO 300 II=1,L
      CX(II)=0.0
      DO 200 KK=1,N-II+1
      KI=KK+II-1
  200 CX(II)=CX(II)+(X(KK)-XM)*(X(KI)-XM)
  300 CX(II)=CX(II)/N
      RETURN
      END
```

================================================================================

```
      SUBROUTINE FATCR(N,L1,W,PI,X,CX)
      DIMENSION X(600),CX(300),XC(600)
      RN=FLOAT(N)
      W=FLOAT(2*PI*W)/RN
      SOM1=0.0
      DO 10 I=1,N
   10 SOM1=SOM1+X(I)
      SOM1=SOM1/RN
      DO 20 I=1,N
   20 X(I)=X(I)-SOM1
   25 CONTINUE
      SOM1=0.0
      SOM2=0.0
      DO 30 I=1,N
      ARG=FLOAT(I-1)*W
      SOM1=SOM1+COS(ARG)*X(I)
   30 SOM2=SOM2+SIN(ARG)*X(I)
      A=2.0*SOM1/RN
      B=2.0*SOM2/RN
      PRINT*,A,B
      DO 50 J=1,L1
      CX(J)=0.0
      DO 40 I=1,N-J+1
      KJ=I+J-1
      ARG1=FLOAT(I-1)*W
      ARG2=FLOAT(KJ-1)*W
      XC(I)=X(I)-A*COS(ARG1)-B*SIN(ARG1)
      YC=X(KJ)-A*COS(ARG2)-B*SIN(ARG2)
   40 CX(J)=CX(J)+YC*XC(I)
   50 CX(J)=CX(J)/RN
      DO 60 I=1,N
   60 X(I)=XC(I)
      RETURN
      END
```

================================================================================

P R I E S T   O N   D I S K

```
DIMENSION Y(900),CX(300),CXX(300),F(300),ICHAR(10),ITITLE(144)
DIMENSION IMAG4(5151),RANGE(4)
DIMENSION ARPS(2),START(2),WA(910),PMAS(2)
DOUBLE PRECISION DSEED
DATA PI/3.14159/,ICHAR(1)/1H./,RANGE/4*0.0/
READ(5,60) (ITITLE(I),I=1,144)

READ(5,/) L1,L2.
Q02:0017:1  IS THE LOCATION FOR EXCEPTIONAL ACTION ON THE I/O STATEMENT
PMAS(1)=0.0
PMAS(2)=0.0
ARPS(1)=0.7
ARPS(2)=-0.49
IP=1
N=100
A=60
INP=0
S=12.5
S=(2*PI)/S
PRINT*/,ARPS

DO 450 II=1,4
INP=INP+1
DSEED=343925.D0
START(1)=10                          7
START(2)=4
NN=N+200
CALL FTGEN(ARPS,PMAS,0.0,START,1.0,DSEED,IP,2,NN,X,WA)
DO 5 I=1,N
5 X(I)=X(I+200)+A*COS(S*I)
CALL FISHM(CX,X,N,L1)
NP=N/2+1
PRINT*/,N,L1,L2
WRITE(6,70)
DO 30 I=1,NP
SOM1=0.0
DO 10 K=2,L1
CXX(K)=CX(K)*COS((2*PI*I*K)/N)
10 SOM1=SOM1+(1-K/N)*CXX(K)
SOM2=0.0
DO 20 K=2,L2
20 SOM2=SOM2+CXX(K)
X(1)=I
SOM1=CX(1)/(2*PI)+SOM1/PI
SOM2=CX(1)/(2*PI)+SOM2/PI
F(I)=(SOM1-SOM2)/CX(1)
30 PRINT*/,I,F(I)
CALL USPLT(X,F,NP,NP,1,1,ITITLE,RANGE,ICHAR,1,IMAG4,IER)
SOM2=0.0
DO 40 K=2,2*L2
CX(K)=CX(K)/CX(1)
40 SOM2=SOM2+CX(K)**2
SOM1=0.0
DO 50 K=2,L2

50 SOM1=SOM1+CX(K)**2
GPI=1+4*SOM1-2*SOM2
PRINT*/,GPI
GPI=GPI/2
SOGPI=SQRT(GPI)
PRINT*/,SOGPI
DELTA=(2*L1)/3-2*L2+(2*(L2**2))/L1
PRINT*/,DELTA
IF(INP-2) 400,350,400
```

```
350 IP=2
    N=100
    L1=45
    L2=20
    GO TO 450
400 N=500
    PW=F(3)
    CALL FATOR(L1,L2,PI,PW,CX)
    L1=225
    L2=100
450 CONTINUE
    PW=F(40)
    CALL FATOR(L1,L2,PI,PW,CX)
 60 FORMAT(72A1)
 70 FORMAT(10X,"VALORES DE P LAMBDA")
    STOP
    END
```

==========================================================================

```
    SUBROUTINE FISHM(CX,X,N,L)
    DIMENSION CX(300),X(600)
    XM=0.0
    DO 100 II=1,N
100 XM=XM+X(II)
    XM=XM/N
    DO 300 II=1,L
    CX(II)=0.0
    DO 200 KK=1,N-II+1
    KI=KK+II-1
200 CX(II)=CX(II)+(X(KK)-XM)*(X(KI)-XM)
300 CX(II)=CX(II)/N
    RETURN
    END
```

==========================================================================

```
    SUBROUTINE FATOR(L1,L2,PI,PW,CX)
    DIMENSION CX(300)
    PW=PW*CX(1)
    A=(8*PI*PW)/(L1-L2)
    SQA=SQRT(A)
    PRINT*/,A
    PRINT*/,SQA
    RETURN
    END
```