

Universidade Federal do Rio Grande - FURG

Instituto de Matemática, Estatística e Física - IMEF

Grupo de Astrofísica Teórica e Computacional - GATC

Determinação de Períodos de Pulsação Estelar Através da Entropia de Shannon Condicional

Gabriel Laufer Ramos

Rio Grande - RS

13 de abril de 2016

Determinação de Períodos de Pulsação Estelar Através da Entropia de Shannon Condicional

Trabalho de Conclusão de Curso apresentado ao curso de Física Bacharelado da Universidade Federal do Rio Grande como requisito parcial para obtenção do título de bacharel em Física.

Orientador: Prof. Dr. Fabricio Ferrari

Prof. Dr. Fabricio Ferrari

Instituto de Matemática, Estatística e Física - IMEF

Orientador

Profa. Dra. Dinalva A. Sales

Instituto de Matemática, Estatística e Física - IMEF

Banca Examinadora

Prof. Dr. Christian Giovanny...

Instituto de Matemática, Estatística e Física - IMEF

Banca Examinadora

Rio Grande, ____ de _____ de _____.

“*colocar alguma citacao aqui.*”

alguem

Determinação de Períodos de Pulsação Estelar Através da Entropia de Shannon Condicional

por Gabriel Lauffer Ramos

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

**Conditional Shannon Entropy Method to Detect Periods on Variable
Stars**

by Gabriel Lauffer Ramos

abstract in english

Agradecimientos

The acknowledgements and the people to thank go here, don't forget to include your project advisor...

Sumário

| | |
|-------------------------------------|-------------|
| Resumo | v |
| Abstract | vii |
| Agradecimentos | ix |
| Lista de Figuras | xiii |
| Lista de Tabelas | xv |
| 1 Introdução | 1 |
| 1.1 A Musica das Esferas | 2 |
| 2 Justificativa e Objetivos | 3 |
| 3 Metodologia | 5 |
| 3.1 Amostragem | 5 |
| 3.2 Analise de Fourier | 5 |
| 3.2.1 Lomb-Scargle | 5 |
| 3.3 Espaço de fase | 5 |
| 3.3.1 Entropia de Shannon | 6 |
| 3.4 Algoritmo | 7 |
| 4 Resultados e Discussão | 11 |
| 4.1 resultado parciais | 11 |
| 4.1.1 Dados Sintéticos | 12 |
| Bibliografia | 17 |

Lista de Figuras

| | | |
|-----|---|----|
| 3.1 | Exemplos de espaço de fase | 6 |
| 3.2 | Fluxograma do algoritmo | 9 |
| 4.1 | Exemplos curva de luz sintética | 13 |
| 4.2 | Resultados obtidos em escala de cinza | 14 |

Lista de Tabelas

| | | |
|-----|--|----|
| 3.1 | Exemplo de dados | 7 |
| 4.1 | Quantidade de dados analisados e resultados corretos | 11 |

*Dedico este trabalho aos meus pais, por todo o incentivo e
esforço deles para me garantirem uma educação de
qualidade.*

Capítulo 1

Introdução

Na astronomia, especialmente no campo das estrelas variáveis, geralmente é necessário analisar dados com períodos desconhecidos. Existem métodos desenvolvidos para lidar com dados que possuem intervalos espaciais uniformes, porém, as observações geralmente são limitadas para o período da noite e possuem limitações devido ao clima e disponibilidade do telescópio, o que faz com que os dados sejam espaçados por uma ordem de horas, dias ou até mesmo meses (Ferraz-Mello, [1981](#)). Assim, os dados obtidos raramente são igualmente espaçados.

A obtenção do período de pulsação de uma estrela variável é fundamental para descrever a estrela. Através do seu período podemos estimar os valores de luminosidade, massa, distância, densidade, etc.

Existem diversos algoritmos para a determinação de períodos em dados astronômicos. Cada um possui um método diferente ou alguma pequena modificação em relação aos demais. Mesmo com uma grande quantidade de métodos, nenhum deles parece se sobressair de uma forma geral (Graham et al., [2013a](#)). Alguns métodos são melhores para lidar com dados que sejam igualmente espaçados, enquanto que outros métodos lidam melhor com períodos senoidais, etc.

Os modelos mais utilizados para determinação de períodos em séries temporais astronômicas fazem um ajuste de curva utilizando o método dos mínimos quadrados (Lomb, [1976](#)) ou utilizam análise de Fourier (Ferraz-Mello, [1981](#)). Outros métodos tentam minimizar alguma grandeza na dispersão da série temporal no espaço de

fase, como é o caso da análise de variância (Schwarzenberg-Czerny, [1989](#)) e da entropia (Cincotta, Mendez e Nunez, [1995](#)).

A ideia deste projeto é testar um método para determinar múltiplos períodos de pulsação de estrelas variáveis, tendo em vista que nenhum destes métodos são aplicados diretamente para esta função. Este método, chamado de Entropia Condicional, busca a minimização da entropia de Shannon condicional na dispersão da série temporal no espaço de fase.

1.1 A Musica das Esferas

Capítulo 2

Justificativa e Objetivos

por quê???

Capítulo 3

Metodologia

descrição das técnicas em detalhes

3.1 Amostragem

falar sobre a amostragem das Lyraes e Nyquist

3.2 Analise de Fourier

3.2.1 Lomb-Scargle

3.3 Espaço de fase

Quando uma estrela possui um comportamento periódico, a variação em sua magnitude é representada em ciclos iguais. Cada ciclo é uma fase. Se os ciclos são iguais, não importa qual ciclo nos estamos observando, apenas onde nos estamos no ciclo. Assim, o espaço de fase é uma representação de todos os ciclos observados em apenas uma fase, ou em apenas um ciclo. Assim, os pontos de sobreposição e formam uma oscilação geral da estrela. Este espaço de fase é calculado pela seguinte expressão,

$$\phi_i = \frac{t_i}{P} - \left[\frac{t_i}{P} \right] \quad (3.1)$$

em que t_i é o i -ésimo dado do tempo, P é o período de oscilação da magnitude e a quantidade entre colchetes representa apenas o numero inteiro da divisão.

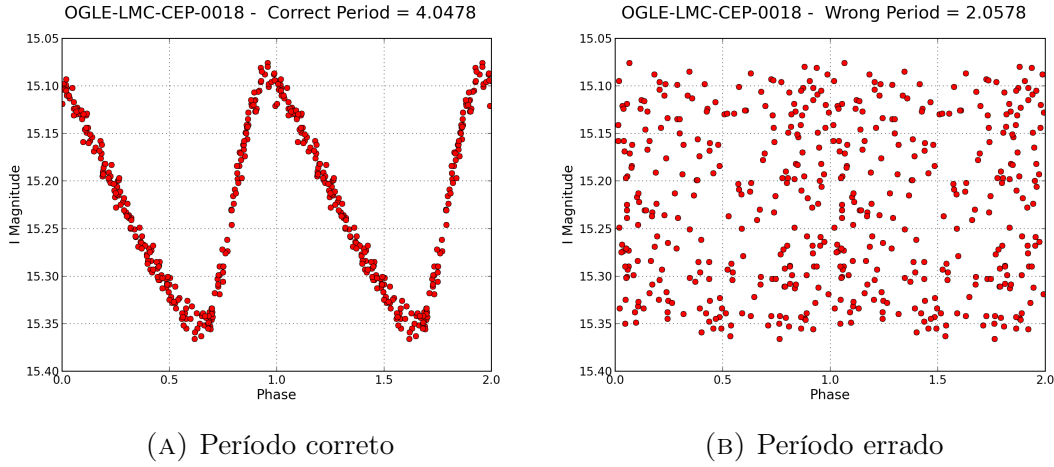


FIGURA 3.1: Exemplos de espaço de fase

Quando uma série temporal é dividida pelo período correto, será gerado uma dispersão com característica oscilante, como é o caso da figura 3.1a. Se o período utilizado na transformação não for o correto, será gerado uma dispersão aleatória, sem forma definida, como mostra a figura 3.1b.

3.3.1 Entropia de Shannon

Na teoria de informação, a entropia, ou entropia de Shannon, é a medida de incerteza de uma variável. A entropia de Shannon mede a falta de informação do nosso sistema, ou seja, quanto maior o seu valor mais incorreto a variável que estamos medindo. Desta forma, vamos procurar pela minimização da entropia no nosso espaço de fase.

Podemos observar que, no caso da figura 3.1a, os pontos se sobrepõem e formam uma curva. Assim, fazendo repartições no dimensão da fase e da magnitude, podemos calcular a probabilidade dos pontos estarem localizados em cada um dos quadrados formados por estas repartições em relação a coluna em que eles estão

e somá-los para obter uma grandeza. Esta grandeza é a entropia condicional, que é calculada pela seguinte formula (Graham et al., 2013b),

$$H_c = \sum_{i,j} p(m_i, \phi_j) \ln \left(\frac{p(\phi_j)}{p(m_i, \phi_j)} \right) \quad (3.2)$$

onde $p(m_i, \phi_j)$ é a probabilidade de ocupação na i -ésima repartição da magnitude e na j -ésima repartição da fase e $p(\phi_j)$ é a probabilidade de ocupação na j -ésima repartição da fase. No caso de repartições retangulares,

$$p(\phi_j) = \sum_i p(m_i, \phi_j) \quad (3.3)$$

A entropia de Shannon mede a falta de informação do sistema, ou seja, quanto maior o seu valor, mais incorreto o período. Por isso que buscamos a minimização da entropia. Considerando estes dois exemplos, a probabilidade de de ocupação das repartições é menor na figura 3.1a do que na figura 3.1b. O menor valor de entropia condicional é associado ao período mais provável da estrela (Graham et al., 2013b).

3.4 Algoritmo

Foi desenvolvido um algoritmo em Python3 para calcular a entropia condicional de dados pertencentes ao [Catálogo OGLE-III de estrelas variáveis](#). Os dados são obtidos no formato .dat e possuem três colunas que significam tempo, magnitude e erro. Um exemplo de arquivo pode ser visto na tabela 3.1.

TABELA 3.1: Exemplo de dados

| Tempo | Magnitude | Erro |
|------------|-----------|-------|
| 2165,85271 | 15,130 | 0,007 |
| 2183,83450 | 15,326 | 0,008 |
| 2238,62899 | 15,102 | 0,007 |
| ⋮ | ⋮ | ⋮ |

Um loop é iniciado e para o primeiro valor do vetor período, o tempo é transformado em fase. São feitas as repartições para este espaço de fase e são contabilizados a quantidade de pontos em cada repartição. Então a entropia condicional é calculada e este valor é armazenado num vetor entropia. O mesmo é feito para o próximo período do vetor período até que sejam calculados a entropia para todos os dados deste vetor. No fim, o algoritmo indica o menor valor do vetor entropia e qual período esta relacionado com este valor. A figura 3.2 apresenta um fluxograma do algoritmo.

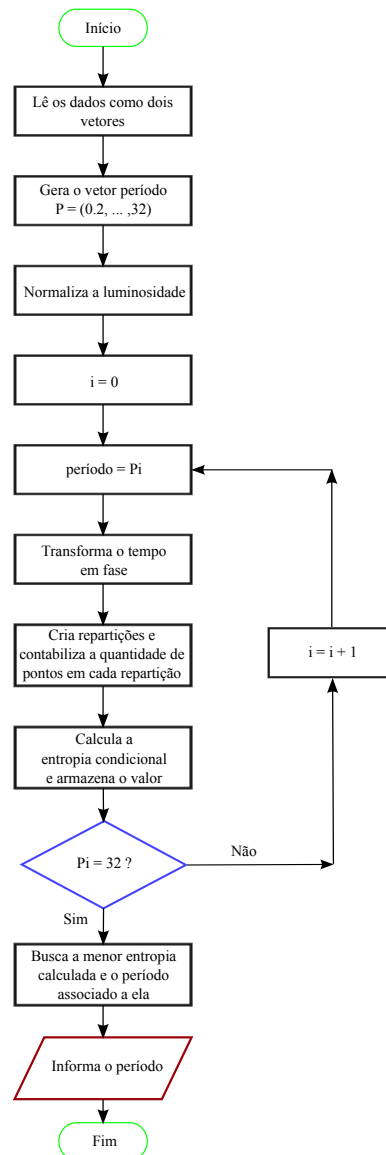


FIGURA 3.2: Fluxograma do algoritmo

Capítulo 4

Resultados e Discussão

Como mostrar os resultados de maneira eficiente

4.1 resultado parciais

Afim de validar o método, foram calculados os períodos de um total de 25707 estrelas variáveis da Grande Nuvem de Magalhães, das quais 3056 eram Cefeidas clássicas tipo FO e FU, e 22651 eram RRLyraes tipo AB e C. Os resultado obtidos foram comparados com os resultados do catálogo e o percentual de acertos pode ser visto na tabela [4.1](#).

Podemos perceber que para as Cefeidas (estrelas de períodos mais longos) o método apresenta um resultado um pouco melhor se comparado com as RRLyraes (estrelas de período mais curto).

Com estes resultados, podemos confiar no método de entropia condicional, porem, ainda queremos entender melhor o comportamento deste método para dados com diferentes níveis de ruído e com diferentes quantidade de pontos de observação.

TABELA 4.1: Quantidade de dados analisados e resultados corretos

| Estrelas | Quantidade | Acertos | Porcentagem |
|--------------|--------------|--------------|----------------|
| Cefeidas FU | 1818 | 1817 | 99,94% |
| Cefeidas FO | 1238 | 1231 | 99,43% |
| RRLyraes AB | 17693 | 17540 | 99,14% |
| RRLyraes C | 4958 | 4535 | 91,47% |
| Total | 25707 | 25123 | 97,73 % |

4.1.1 Dados Sintéticos

Dados sintéticos foram criados a fim de explorar o método e entender até onde podemos utiliza-lo. De acordo com a tabela 4.1, as RRLyraes apresentaram uma taxa menor de acerto então elas foram utilizadas como referencia para construir os dados sintéticos. Assim, foi determinado a partir dos dados qual a variação média entre os pontos de observação para assim calcular a amostragem média dos dados. A amostragem representa a frequência de pontos de observação. Desta forma, foi calculado a amostragem,

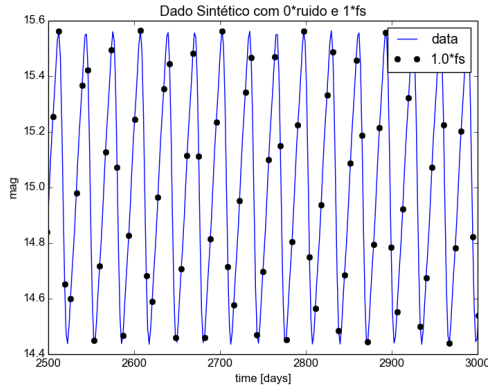
$$f_s = \frac{1}{dt} = 0.1473. \quad (4.1)$$

Obtendo a amostragem, podemos construir dados sintéticos variando a amostragem e o nível de ruído afim de estudar o comportamento do método. De acordo com Graham et al., 2013b e Cincotta, Mendez e Nunez, 1995, para construir dados sintéticos semelhantes com os dados observacionais da maioria dos Surveys de estrelas variáveis, podemos utilizar a seguinte expressão,

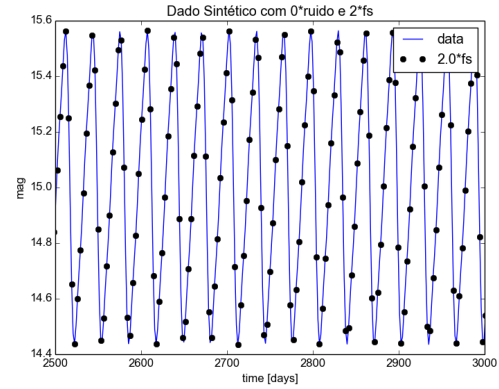
$$m(t) = A_0 + \sum_i^3 A_n \sin\left(\frac{2n\pi t}{P}\right) + \varepsilon\eta \quad (4.2)$$

em que ε é um fator de escala para o ruído entre 0.0 e 1.0, η é uma distribuição gaussiana com média zero e desvio unitário e P é o período médio das RRLyraes que, segundo **lyraes** é de 0.576 dias.

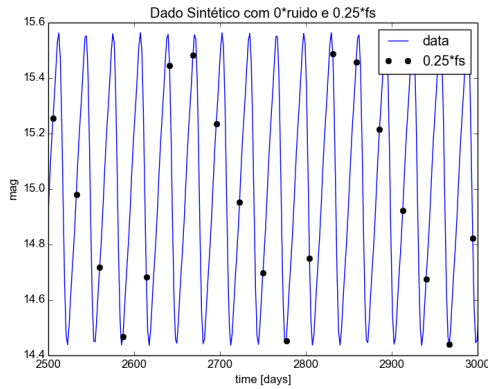
A influencia da amostragem está no vetor t que é construindo a fim de representar de forma mais fiel possível os dados do Catálogo OGLE-III. Sendo assim, o vetor tempo é construído com os seguinte parâmetros: tempo inicial de 2152.5019 HJD, tempo final de 4539.4593 HJD e espaçamento entre os pontos $dt = 1/f$ em que $f = n \times f_s$ e n é um parâmetro de escala para a amostragem. Os tempos iniciais e finais foram escolhidos desta forma por serem os valores de maior frequência entre os dados das RR Lyraes. Quatro exemplos de curva de luz sintética gerada pelo método acima podem ser vistas na figura 4.1.



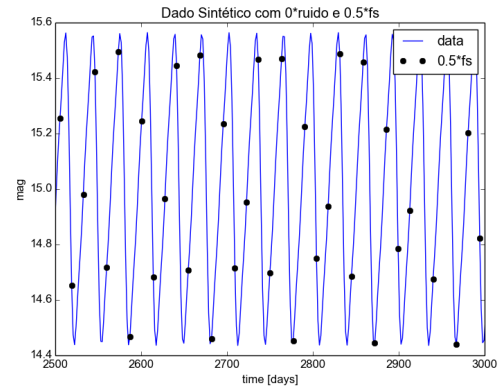
(A) Dado sem ruído e com amostragem padrão



(B) Dado sem ruído com $n = 2$



(C) Dado sem ruído e com $n = 1/4$



(D) Dado sem ruído com $n = 1/2$

FIGURA 4.1: Exemplos curva de luz sintética

Na figura 4.1, os pontos pretos são os pontos de observação e a linha contínua é o dado original com $n = 1$. Podemos perceber que quanto maior a amostragem, maior a quantidade de pontos, assim o método aplicado a um dado com

uma grande amostragem deve retornar um período com maior precisão do que comparado à um dado com pequena amostragem.

Então, para estudar a influencia da amostragem nos dados, foram gerados dados sintéticos variando o parâmetro n da amostragem de 0.25 a 4 com intervalo de 0.25 e variando o fator de escala ε de 0.0 até 1.0 com intervalo de 0.05 assim obtendo 300 curvas de luz. No momento, estamos pensando em como demonstrar os resultados obtidos. Uma forma para demonstrar os dados é fazendo um mapa de cor entre ruído e amostragem onde a cor representa o valor $|(P - P_0)/P_0|$, ou seja, quanto que o período calculado está variando em relação ao período original. O mapa de cor é feito em escala de cinza, em que a cor mais escura representa o valor 0 (período calculado = período real) e quanto mais clara a cor, maior o desvio do período.

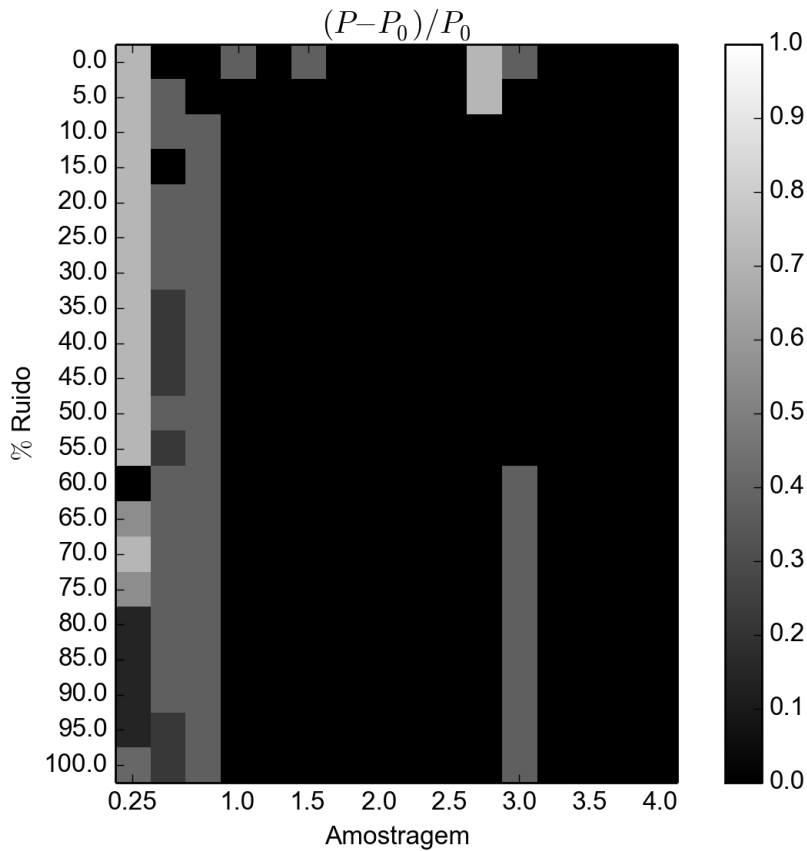


FIGURA 4.2: Resultados obtidos em escala de cinza

Os valores iguais a zero (cor preta) representam as configurações em que o método de entropia condicional calculou o período corretamente.

Com esta análise, é possível construir uma ferramenta que nos indica como os dados influenciam no resultado do método, ou ainda, partindo do resultado que se espera obter, é possível escolher como a observação deve ser feita.

Bibliografia

- Cincotta, P. M., M. Mendez e J. A. Nunez (1995). “Astronomical Time Series Analysis. I. A Search for Periodicity Using Information Entropy”. Em: *The Astrophysical Journal* 449, p. 231. DOI: [10.1086/176050](#).
- Ferraz-Mello, S. (1981). “Estimation of Periods from Unequally Spaced Observations”. Em: *The Astronomical Journal* 86, p. 619. DOI: [10.1086/112924](#).
- Graham, M. J. et al. (2013a). “A comparison of period finding algorithms”. Em: *Monthly Notices of the RAS* 434, pp. 3423–3444. DOI: [10.1093/mnras/stt1264](#). arXiv: [1307.2209 \[astro-ph.IM\]](#).
- Graham, M. J. et al. (2013b). “Using conditional entropy to identify periodicity”. Em: *Monthly Notices of RAS* 434, pp. 2629–2635. DOI: [10.1093/mnras/stt1206](#). arXiv: [1306.6664 \[astro-ph.IM\]](#).
- Lomb, N. R. (1976). “Least-squares frequency analysis of unequally spaced data”. Em: *Astrophysics and Space Science* 39, pp. 447–462. DOI: [10.1007/BF00648343](#).
- Schwarzenberg-Czerny, A. (1989). “On the advantage of using analysis of variance for period search”. Em: *Monthly Notices of the RAS* 241, pp. 153–165.