STATISTICS
TEXAS A&M UNIVERSITY

Period Estimation for Variable Sources

James Long

September 14, 2016

Objective Function Approaches

Modeling Approaches

Data and Generic Model

- observe brightness m_{jb} at time t_{jb} with uncertainty σ_{jb} in band b
- data for single star is

$$D = \{(t_{jb}, m_{jb}, \sigma_{jb})\}_{j=1}^{n_b} \quad \text{for } b = 1, \ldots, B$$

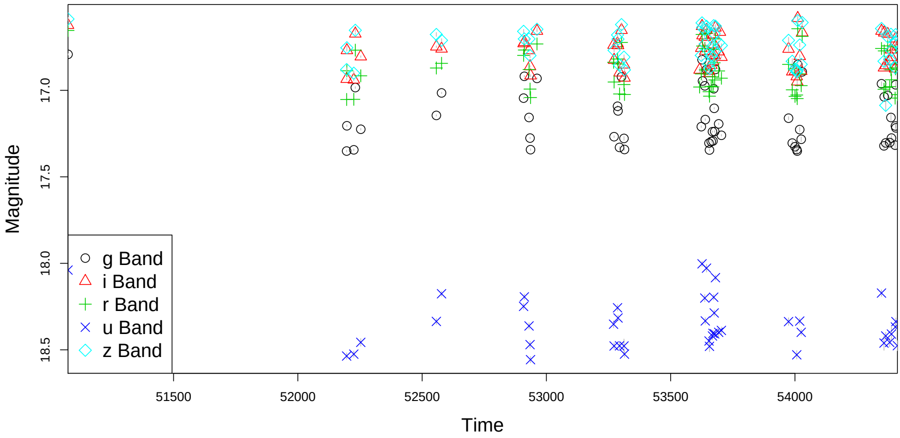
- parameters for source:
  - p is light curve period
  - $\omega = 1/p$ is light curve frequency
  - f_b is periodic function with period 1 ie

    $$f_b(t) = f_b(t+1)$$

    for all t, b.
- $$m_{jb} = f_b(\omega t_{jb}) + \epsilon_{jb}$$

  where $\text{Var}\,(\epsilon_{jb}) = \sigma_{jb}^2$ and ϵ_{jb} indep.

Data



Magnitude
17.0
17.5
18.0
18.5
g Band
i Band
r Band
u Band
z Band
51500
52000
52500
53000
53500
54000
Time

Goals and Challenges

Goals:

- estimate period p
- estimate lightcurve shape f_b
- use p and f_b for classification, eg feature extraction
- characterize population of f_b
- estimate PL relations, Hubble constant

Challenges:

- how much data do I need?
- does multiband data help?
- how much can I assume about f_b?
- chicken and egg: physical models tell me about f_b, but want to learn about physical models from data
- computational speed

Simplification: Single Band Data

- observe brightness m_j at time t_j with uncertainty σ_j
- data

$$D = \{(t_j, m_j, \sigma_j)\}_{j=1}^{n}$$

- model parameters:
  - p is light curve period
  - $\omega = 1/p$ is light curve frequency
  - f is periodic function with period 1 ie

$$f(t) = f(t+1)$$

    for all t.

-

$$m_j = f(\omega t_j) + \epsilon_j$$

  where $\text{Var}\,(\epsilon_j) = \sigma_j^2$ and ϵ_j indep.

Connecting Data and Parameters

Traditional Statistical Approach: Use a model with parameters, eg

$$f(t) = \beta_0 + a \sin(2\pi\omega t + \phi)$$

- bayesian: priors on $(\omega, \beta_0, a, \phi)$, computes posterior, summarizes posterior with MAP or post. mean, credible sets, etc.
- frequentist: computes MLE, confidence interval, etc.

Objective Function Minimization Approach: Estimate is

$$\widehat{p} = \underset{p}{\operatorname{argmin}}\ M(p, D)$$

M chosen intuitively, essentially M–estimators.

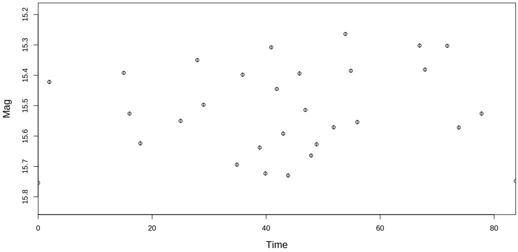
Outline

Objective Function Approaches

Modeling Approaches

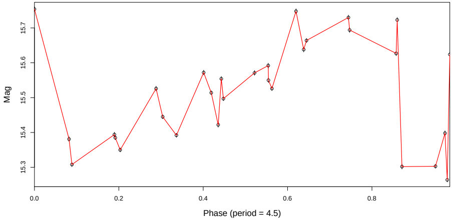
Minimum String Length Method

$$M(p, D) = \text{length of string fit through light curve points folded on period } p$$

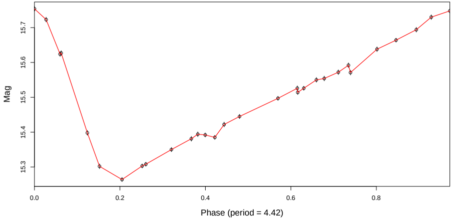


Example of Two Strings

Fold light curve on $p = 4.50$ (incorrect period)



Fold light curve on $p = 4.42$ (correct period)



Idea: String is shorter when you are near correct period.

Formal Definition

- $D \leftarrow$ light curve data
- $p_1, \dots, p_M$ grid of frequencies
- **for** $i = 1, \dots, M$
  - $t'_j \leftarrow (t_j \mod p_i)/p_i$ for $j = 1, \dots, n$
  - $(t'_{(j)} m'_{(j)})$ are ordered time/mag pairs after folding ie

$$t'_{(j)} < t'_{(k)} \text{ for } j < k$$

  - $M(p_i, D) \leftarrow \sum_{i=1}^{n} \sqrt{(t'_{(j)} - t'_{(j-1)})^2 + (m'_{(j)} - m'_{(j-1)})^2}$
- $i \leftarrow \underset{i}{\text{argmin}}\ M(p_i, D)$
- **return** p_i

Define $t'_{(0)} \equiv t'_{(n)}$, $m'_{(0)} \equiv m'_{(n)}$, $t_{(1)} - t_{(0)} \equiv t_{(1)} - t_{(0)} \mod 1$.

Related Methods

- **“An RR Lyrae Star Survey with Ihe Lick 20-INCH Astrograph II. The Calculation of RR Lyrae Periods by Electronic Computer.”** ApJ 1965 Lafler, Kinman
- **“A period-finding method for sparse randomly spaced observations or ’How long is a piece of string?”’** MNRAS 1983 Dworetsky
- **“Period Determination Using Phase Dispersion Minimization”** ApJ 1978 Stellingwerf

Project Idea: Use M–estimators to study asymptotic properties of some of these procedures.

Outline

Objective Function Approaches

Modeling Approaches

Fit a Sinusoid with K Harmonics

Model:

$$m_j = \beta_0 + \sum_{k=1}^{K} a_k \sin(2\pi k \omega t_j + \rho_k) + \epsilon_j$$

$$\epsilon_j \sim N(0, \sigma_j^2) \text{ independent}$$

Maximum Likelihood Frequency Fit:

Let: $\mathbf{a} = (a_1, \ldots, a_K)$, $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_K)$.

$$RSS(\omega, \beta_0, \mathbf{a}, \boldsymbol{\rho}) = \sum_{i=1}^{n} \left(\frac{m_j - \sum_k a_k \sin(2\pi k \omega t_j + \rho_k) - \beta_0}{\sigma_j} \right)^2$$

Frequency estimate is $\widehat{\omega} = \underset{\omega}{\text{argmin}} \ \min_{\mathbf{a},\boldsymbol{\rho},\beta_0} RSS(\omega, \beta_0, \mathbf{a}, \boldsymbol{\rho})$.

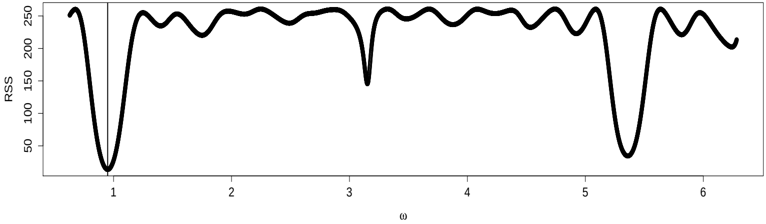
Computing the ML Estimate $\widehat{\omega}$

- Grid search across frequency.
- For fixed frequency ω, weighted least squares (linear model)

$$m_j = \beta_0 + \sum_k a_k \sin(2\pi k\omega t_{bi} + \rho_k) + \epsilon_j$$

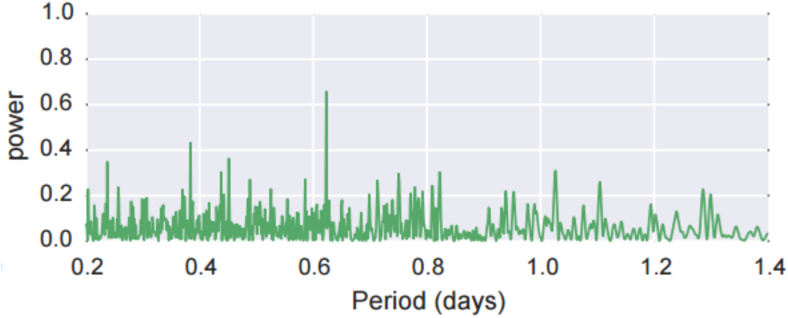
$$= \beta_0 + \sum_k (\underbrace{a_k \cos(\rho_k)}_{\equiv \beta_{k1}} \sin(2\pi k\omega t_j) + \underbrace{a_k \sin(\rho_k)}_{\equiv \beta_{k2}} \cos(2\pi k\omega t_j))$$

$$+ \epsilon_j$$



Periodogram

$$periodogram(\omega) \propto \frac{1}{RSS(\omega)}$$



Source: Figure 2 of `http://arxiv.org/pdf/1502.01344.pdf`

Relevant Literature

- **sinusoid:** lomb-scargle (LS), generalized lomb scargle (GLS)
- **sinusoid with k harmonics:** "Fast and statistically optimal period search in uneven sampled observations" ApJ 1996 Schwarzenberg-Czerny
- **non-parametric methods:** "Nonparametric estimation of a periodic function" Biometrika 2000 Hall et al.
- **bayesian methods:** "A new method for the detection of a periodic signal of unknown shape and period" ApJ Gregory and Loredo
- **comparison studies:** "A Comparison of Period Finding Algorithms" ApJ 2013 Graham et al

Preview Next Week

- multiband period estimation
- uncertainty quantification for period estimators
- building models for light curve classes
    - physical models
    - data based models
- application: finding structure in the Milky Way Halo