Observation Scheduling for Real-Time Lightcurve Classification

David Jones

Harvard University Statistics Department

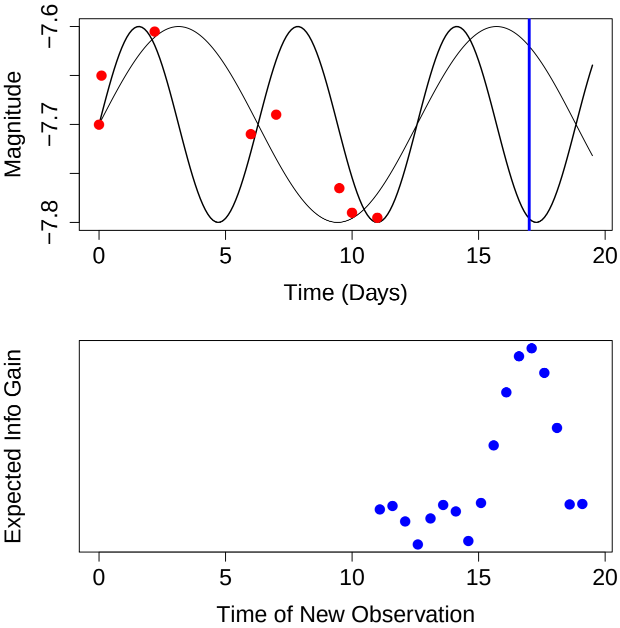
Collaborators: Xiao-Li Meng (Harvard), Aneta Siemiginowska (CfA), Vinay Kashyap (CfA)

International CHASC Astrostatistics Center

SCMA 6

June 8, 2016

One slide version



Challenges in practice

- Model parameter uncertainty
- Prior knowledge
- Observation error bars
- More than two classes – how to measure separation?
- Others I will mention at the end . . .

Two class setup

Data / model:

- Classes: C_1 and C_2, with prior probabilities π_1 and π_2 (sum to one)
- **Task:** choose times to observed a lightcurve, $t = (t_1, \ldots, t_n)$
- Magnitudes $x = (x_1, \ldots, x_n)$ are then observed
- Models: $f(x|C_i, t, \theta_i)$, unknown θ_i, for $i = 1, 2$

Two class setup

Data / model:

- Classes: C_1 and C_2, with prior probabilities π_1 and π_2 (sum to one)
- **Task:** choose times to observed a lightcurve, $t = (t_1, \ldots, t_n)$
- Magnitudes $x = (x_1, \ldots, x_n)$ are then observed
- Models: $f(x|C_i, t, \theta_i)$, unknown θ_i, for $i = 1, 2$

Method:

- Bayesian comparison: BF$(x|C_1, C_2) = \frac{f(x|C_1,t)}{f(x|C_2,t)} = \frac{\int_{\Theta_1} f(x|C_1,t,\theta_1)\pi(\theta_1|C_1)d\theta_1}{\int_{\Theta_2} f(x|C_2,t,\theta_2)\pi(\theta_2|C_2)d\theta_2}$
- Question: how should we choose t?
- Usual design perspective is to maximize some criterion / information measure

Generalized variance of Bayes factor

- $\mathcal{V}$ = *evidence function* (concave)
- $\mathcal{V}$(BF) = evidence for C_1

$$\begin{aligned}\mathcal{I}_{\mathcal{V}}(t; C_1, C_2, \pi) &= \text{Initial evidence for } C_1 - \text{Expected posterior evidence for } C_1 \\ &= \mathcal{V}(1) - E_X[\mathcal{V}(\text{BF}(X|C_1, C_2))|C_2]\end{aligned}$$

Generalized variance of Bayes factor

- $\mathcal{V}$ = *evidence function* (concave)
- $\mathcal{V}$(BF) = evidence for C_1

$$\mathcal{I}_{\mathcal{V}}(t; C_1, C_2, \pi) = \text{Initial evidence for } C_1 - \text{Expected posterior evidence for } C_1$$
$$= \mathcal{V}(1) - E_X[\mathcal{V}(\text{BF}(X|C_1, C_2))|C_2]$$

- Usual variance if $\mathcal{V}(\text{BF}) = -(\text{BF} - 1)^2$
- $\mathcal{V}(\text{BF}) = \log(\text{BF})$ gives $KL(f(\cdot|C_2, t)||f(\cdot|C_1, t))$ (Nicolae et al. (2008))

Statistical information for decision problems, Jones and Meng (2016+)

Sequential version

- Observed magnitudes x_{ob} at times t_{ob}
- Want to schedule new observation X_{new} for time t_{new}

$$\begin{aligned}\mathcal{I}_{\mathcal{V}}(t_{\text{new}}|t_{\text{ob}}, x_{\text{ob}}) &= \text{Observed evidence for } C_1 - \text{Expected complete data evidence for } C_1 \\ &= \mathcal{V}(\text{BF}(x_{\text{ob}}|C_1, C_2)) - E_{X_{\text{new}}}[\mathcal{V}(\text{BF}(x_{\text{ob}}, X_{\text{new}}|C_1, C_2))|C_2, x_{\text{ob}}]\end{aligned}$$

Appealing choice: $\mathcal{V}(\text{BF}) = \text{BF}/(\pi_2 + \pi_1\text{BF}) = \frac{P(C_1|x,t)}{\pi_1}$

Appealing choice: $\mathcal{V}(\text{BF}) = \text{BF}/(\pi_2 + \pi_1\text{BF})$=$\frac{P(C_1|x,t)}{\pi_1}$

- Sequential information:

$$\mathcal{I}_{\mathcal{V}}(t_{\text{new}}|t_{\text{ob}}, x_{\text{ob}}) = \frac{\text{current probability for } C_1 - \text{ expected new probability for } C_1}{\text{prior probability for } C_1}$$

Statistical information for decision problems, Jones and Meng (2016+)

Appealing choice: $\mathcal{V}(\text{BF}) = \text{BF}/(\pi_2 + \pi_1\text{BF})$=$\frac{P(C_1|x,t)}{\pi_1}$

- Sequential information:

$$\mathcal{I}_{\mathcal{V}}(t_{\text{new}}|t_{\text{ob}}, x_{\text{ob}}) = \frac{\text{current probability for } C_1 - \text{ expected new probability for } C_1}{\text{prior probability for } C_1}$$

- Scheduling method: choose t_{new} that maximizes $\mathcal{I}_{\mathcal{V}}(t_{\text{new}}|t_{\text{ob}}, x_{\text{ob}})$

Coherence identity

If $\mathcal{V}$ satisfies

$$\frac{\text{Evidence for } C_1}{\text{Evidence for } C_2} = \frac{\mathcal{V}(\text{BF}; C_1, C_2)}{\mathcal{V}(1/\text{BF}; C_2, C_1)} = \text{BF}$$

then the following coherence identity holds

$$\mathcal{I}_{\mathcal{V}}(t_{\text{new}} | t_{\text{ob}}, x_{\text{ob}}; C_1, C_2) = \text{BF}(x_{\text{ob}}) \mathcal{I}_{\mathcal{V}}(t_{\text{new}} | t_{\text{ob}}, x_{\text{ob}}; C_2, C_1)$$

$\Rightarrow$ **the optimal time to collect new data does not depend on the true class**

m classes

Ideas can be extended in two ways:

1. Compare all pairs (under a hierarchy)

$$\sum_{i=2}^{m} \sum_{j=1}^{i-1} \mathcal{I}_{\mathcal{V}}(t_{\text{new}} | t_{\text{ob}}, x_{\text{ob}}; C_j, C_i)$$

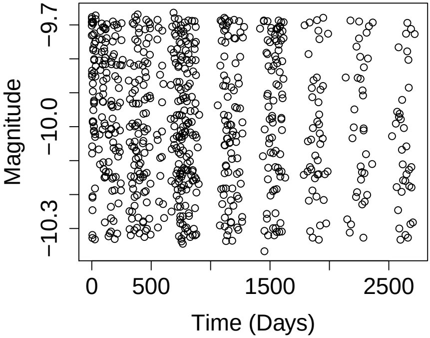
2. Compare each class to a baseline class

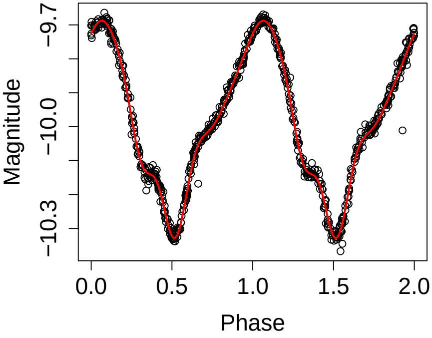
$$\sum_{i=1}^{m} \mathcal{I}_{\mathcal{V}}(t_{\text{new}} | t_{\text{ob}}, x_{\text{ob}}; C_B, C_i) P(C_i | t_{\text{ob}}, x_{\text{ob}})$$

   Box and Hill (1967) is a special case

Data

- MACHO lightcurve catalog subset
- Periodic sources
- 66 Cepheids, 180 eclipsing binaries, 266 RR Lyrae variables





Gaussian Process Model (what we called $f(x|C_i, t, \theta_i)$)

Magnitudes:

$$(X_1, \ldots, X_n) \sim N(\mu \mathbf{1}_n, D + V)$$

where μ = mean magnitude

Gaussian Process Model (what we called $f(x|C_i, t, \theta_i)$)

Magnitudes:

$$(X_1, \ldots, X_n) \sim N(\mu \mathbf{1}_n, D + V)$$

where μ = mean magnitude

Covariance matrix $D + V$:

1. Observation errors: $D = \text{diag}(s_1^2, \ldots, s_n^2)$
2. Periodic kernel: $V_{ij} = \sigma^2 \exp\left(-\beta \sin\left(\frac{\pi(t_i - t_j)}{\tau}\right)^2\right)$ for $i, j \in \{1, \ldots, n\}$
   - τ = period
   - σ = standard deviation around the mean
   - β = inverse length-scale (inverse relaxation time)

Gaussian Process Model (what we called $f(x|C_i, t, \theta_i)$)

Magnitudes:

$$(X_1, \ldots, X_n) \sim N(\mu \mathbf{1}_n, D + V)$$

where μ = mean magnitude

Covariance matrix $D + V$:

1. Observation errors: $D = \text{diag}(s_1^2, \ldots, s_n^2)$
2. Periodic kernel: $V_{ij} = \sigma^2 \exp\left(-\beta \sin\left(\frac{\pi(t_i - t_j)}{\tau}\right)^2\right)$ for $i, j \in \{1, \ldots, n\}$
   - τ = period
   - σ = standard deviation around the mean
   - β = inverse length-scale (inverse relaxation time)

Note:

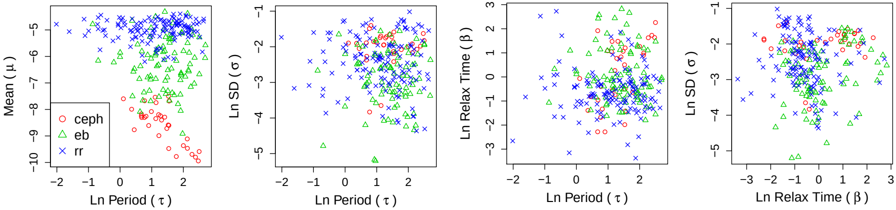
Same model $f(x|C_i, t, \theta)$ for each class C_1, C_2, C_3

Different prior distribution on parameters $\theta = (\mu, \ln \tau, \ln \sigma, \ln \beta)$

$\Rightarrow f(x|C_i, t)$ depends on class

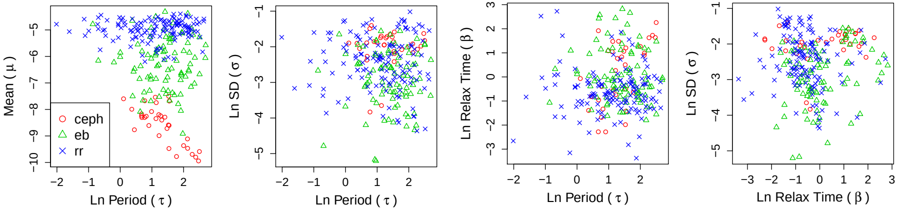
Empirical priors from training data

Training data fits: Maximum likelihood parameter fits for half the lightcurves



Empirical priors from training data

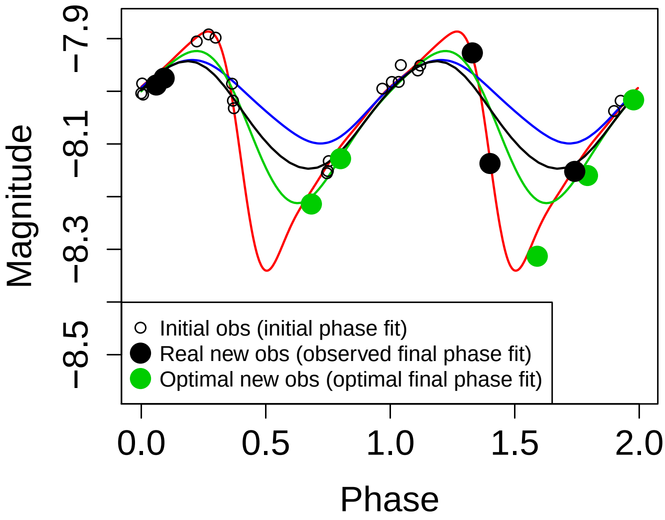
Training data fits: Maximum likelihood parameter fits for half the lightcurves



Simple construction of class specific priors:

1. $\hat{\theta}^{(j)}_{\text{train(C)}}$: training data fit for class C source j
2. Mean$(\hat{\theta}_{\text{train(C)}})$: training data mean fit for class C
3. Cov$(\hat{\theta}_{\text{train(C)}})$: training data estimated covariance matrix for class C
4. Priors

$$\theta|C \sim N(\text{Mean}(\hat{\theta}_{\text{train(C)}}), \text{Cov}(\hat{\theta}_{\text{train(C)}})) \qquad \text{for } C \in \{\text{ceph, eb, rr}\}$$



Magnitude
−7.9
−8.1
−8.3
−8.5
Phase
0.0
0.5
1.0
1.5
2.0
Initial obs (initial phase fit)
Real new obs (observed final phase fit)
Optimal new obs (optimal final phase fit)

Results: posterior probability based classification

After first new obs.: / After all 5 new obs.:

Real obs:

	ceph	eb	rr
ceph	27	6	0
eb	10	55	25
rr	0	34	99

	ceph	eb	rr
ceph	29	2	0
eb	8	60	22
rr	0	31	102

Selected obs:

	ceph	eb	rr
ceph	28	5	0
eb	10	56	24
rr	0	34	99

	ceph	eb	rr
ceph	31	2	0
eb	5	66	19
rr	0	26	107

$\sim 7\%$ improvement after 5 steps

Analysis Extensions and Summary

Analysis extension:

- Additional model flexibility e.g. class specific, changing period / damping
- Include different types of classes e.g. non-periodic, event-based
- Incorporate additional wavelength information e.g. Mandel (2009)
- Penalize longer wait times until next observation
- Improve computational efficiency

Summary

- Astrostatistics loop
- Design for decision problems
- Return to lightcurve problem
- Future: design for estimation **and** decision problems

Thanks!