Astronomical Data Analysis Software and Systems XIV **P3.1.7**
ASP Conference Series, Vol. XXX, 2005
P. L. Shopbell, M. C. Britton, and R. Ebert, eds.

Optimizing Observing Sequence Design for Periodic and Non-periodic Phenomena: A Bayesian Approach

Mark D. Johnston and Russell Knight

Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive, M/S 126-347, Pasadena, CA 91109-8099
email: {firstname.lastname}@jpl.nasa.gov

Abstract. The problem of designing observing sequences to detect and characterize periodic phenomena occurs regularly in astronomical investigations. Examples of current interest include Cepheid variable searches in external galaxies (with Hubble Space Telescope), and future high accuracy astrometric observations of nearby stars with the Space Interferometry Mission (SIM) satellite to search for planetary companions. Various sampling strategies have been proposed to obtain good phase coverage over an interesting range of periods. Recently, Loredo and Chernoff have proposed the use of "Bayesian adaptive exploration", a model-based Bayesian method that exploits observations made to date to determine the best future observation times according to a maximum information criterion. While this method makes the best possible use of any results already obtained, it does not address the "bootstrap" problem of scheduling in advance of any data collection. It also is highly compute-intensive, which is especially problematic when an integrated observing schedule for hundreds of targets is required, taking into account all of the various other constraints and preferences that come into play. In this paper we report on our progress on addressing these issues. We have developed an approximate expression for the uniformity of phase coverage that can be used when scheduling to assess candidate sample times. We describe the results obtained using this estimator, and compare them with detailed simulations. We describe our progress and plans for integrating optimizing criteria for both periodic and non-periodic observations into a single observation sequence.

1. Introduction

Data analysis techniques for periodic phenomema have become very sophisticated (see e.g. Schwarzenberg-Czerny 1999, Bretthorst 2001), providing powerful techniques for the estimation of periods and other source parameters. From the perspective of planning to *acquire* data, the question naturally arises of how best to schedule data collection, in order to optimize post-observation data analysis. Information-theoretic approaches have been proposed, such as described by Sebastiani and Wynn (2000), motivated by the criterion that it is preferable to make observations which provide the greatest additional information, given all that is currently known.

This approach has been elaborated in detail by Loredo and Chernoff (2003) who propose a technique called *Bayesian Adaptive Exploration* (BAE). Their approach is based on a Bayesian model of the phenomena, from which is derived an expression for the expected information gain as a function of (a) results from observations to date and (b) the derived model probability distribution. Such an approach is extremely appealing in that it makes maximum use of all available information, and is specific to the detailed form of the model for the phenomena under investigation. However, there are some practical drawbacks to the method: it is computationally very expensive (multiple Markov Chain Monte Carlo simulations may be required to numerically evaluate the model), and it requires data from past observations to estimates times for future ones. In large scale planning and scheduling applications, where many future observation possibilities must be considered in the context of interacting constraints and preferences, it is infeasible to use this kind of approach.

In the following we describe an approximation, motivated and based on the Bayesian modeling approach, which allows us to estimate the uniformity of phase coverage as a function of a set of observing times, past and future. We first describe the approach, then provide several examples of how our approximation can be used in practice. We conclude with some comments on future research and application areas.

2. Approach

Our approach is based on the results of Bretthorst (2001) who analyzed in detail the case of a noisy sinusoidal signal with non-uniform samples. He derived an expression for the Bayesian posterior probability distribution for the frequency f, given a set of observations $\mathbf{d}_i(t) = \mathbf{A}\exp(-2\pi i f t_i) + \mathbf{n}(t_i)$, where $\mathbf{d}_i$ are the data samples at times t_i, $\mathbf{A}$ is the signal amplitude, and $\mathbf{n}$ the noise level. If we examine the expression for $P(f)$, the posterior probability of frequency f, we note that the denominator of Bretthorst's $\overline{h^2}$ statistic is a function only of t_i (i.e. is independent of the data values $\mathbf{d}_i$) and provides a measure of the variation in the measured probability that arises *solely* from the choice of sampling times t_i. This expression has the form:

$$V(f) = \left(\sum_i \cos^2(2\pi f t_i)\right)\left(\sum_i \sin^2(2\pi f t_i)\right) - \left(\sum_i \cos(2\pi f t_i)\sin(2\pi f t_i)\right)^2$$

We use the normalized mean square deviation of the quantity $V(f)$, over a frequency range of interest $[f_{min}, f_{max}]$, as a measure of *phase coverage nonuniformity* — the quantity to minimize. We have compared this quantity to the results of simulation runs with noisy data analyzed to determine the signal frequency: we find that this approximation tracks quite well the likelihood that the true frequency will be determined.

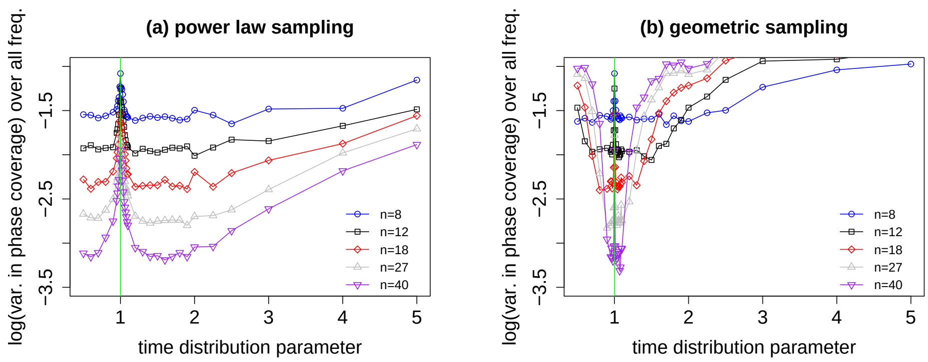

Figure 1. Comparison of (a) power law and (b) geometric series sampling for a range of sample sizes.

3. Examples

1. Sample Time Selection — Consider the case of selecting in advance a set of observation times with the goal of detecting a periodic signal of frequency in the range $[f_{min}, f_{max}]$. This situation arises in Cepheid searches (e.g. Freedman et al. 1994) and other variable star observations. A comparison of power law and geometric series sampling strategies is shown in Fig. 1 for a range of sample sizes (time span of 100d, period range 2-100d). The conclusion that power law sampling is preferred is consistent with previous investigations.

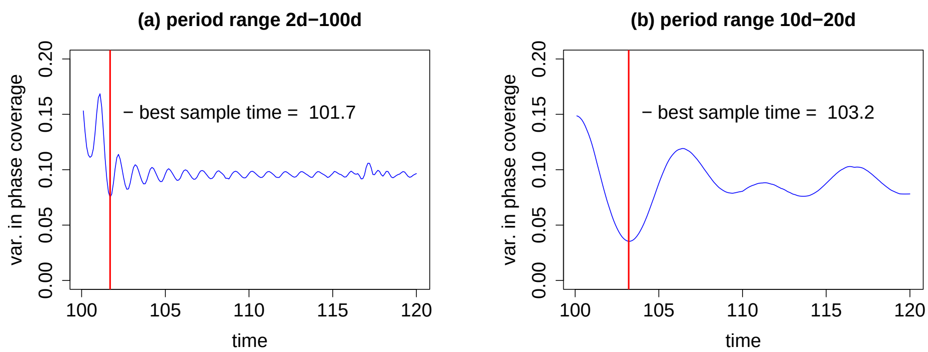

Figure 2. Variation in phase coverage as a function of scheduled time of observation $N+1$. The results depend on the frequency range of interest: in (a) the period range is 2d to 100d, while in (b) the range is 10d-20d.

2. Incremental Scheduling — Suppose we have scheduled N observations and seek the best time to schedule observation $N+1$. This is the situation addressed by BAE when there exists both data and a model. When we have no data, we can use our approximation to find a time of minimal variation in phase coverage given a frequency range of interest. An example where $N = 4$ is shown in Fig. 2, where the best sample time depends on what one expects to be the period range of interest.

3. Eliminating Redundant Schedule Times — In some cases we have given a set of candidate schedule times that is larger than can be actually accomplished – a situation which arises in some satellite mission scheduling problems such as SIM (Schwartz and Wehrle 2004). The problem is to *reduce* the set to one of manageable size. The approximation formula can be used to assess candidate subsets. We have conducted numerical experiments that show that a guided selection of a good subset can improve by factors of several the variation in phase coverage. Such an improvement can make a dramatic difference in the detectability of periodicities at the sensitivity limit.

4. Future Directions

There are several directions for future research and application of this work:

- Incorporate information gained from past observations
- Broaden to non-sinusoidal variability (if analysis shows this useful)
- Automate the analysis of redundant candidate schedule times
- Investigate integration with schedulers such as ASPEN (Chien et al. 2000) as a specialized heuristic

Acknowledgments. The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. The authors wish to recognize and thank the SIM team at JPL, and Arnie Schwartz in particular, for very useful discussions related to this work.

References

Bretthorst, G. L. 2001, “Nonuniform Sampling: Bandwidth and Aliasing”, in Maximum Entropy and Bayesian Methods in Science and Engineering, ed. J. Rychert, G. Erickson and C. R. Smith, AIP: 1-28.

Chien, S., G. Rabideau, et al. 2000, “ASPEN – Automating Space Mission Operations using Automated Planning and Scheduling”, SpaceOps 2000. Toulouse, France.

Freedman, W. L., S. M. Hughes, et al. 1994, “The Hubble Space Telescope Extragalactic Distance Scale Key Project. 1: The discovery of Cepheids and a new distance to M81.” ApJ, 427: 628-655.

Loredo, T. J. and D. F. Chernoff 2003, “Bayesian Adaptive Exploration”, in Statistical Challenges in Modern Astronomy III, ed. E. D. Feigelson and G. J. Babu. New York, Springer-Verlag: 57-69.

Sebastiani, P. and H. P. Wynn 2000, “Maximum entropy sampling and optimal Bayesian experimental design.” J. R. Statist. Soc. B 62: 145-157.

Schwartz, A. A. and A. E. Wehrle 2004, “Formulating an Observation Plan for SIM”, AAS 203rd Meeting.

Schwarzenberg-Czerny, A. 1999, ”Optimum Period Search: Quantitative Analysis.” ApJ, 516: 315-323.