Bayesian Adaptive Exploration

Thomas J. Loredo
David F. Chernoff

ABSTRACT We describe a framework for adaptive astronomical exploration based on iterating an *Observation–Inference–Design* cycle that allows adjustment of hypotheses and observing protocols in response to the results of observation on-the-fly, as data are gathered. The framework uses a unified Bayesian methodology for the inference and design stages: Bayesian inference to quantify what we have learned from the available data; and Bayesian decision theory to identify which new observations would teach us the most. In the design stage, the utility of possible future observations is determined by how much information they are expected to add to current inferences as measured by the (negative) entropies of the probability distributions involved. Such a Bayesian approach to experimental design dates back to the 1970s, but most existing work focuses on linear models. We use a simple *nonlinear* problem—planning observations to best determine the orbit of an extrasolar planet—to illustrate the approach and demonstrate that it can significantly improve observing efficiency (i.e., reduce uncertainties at a rate faster than the familiar "root-N" rule) in some situations. We highlight open issues requiring further research, including dependence on model specification, generalizing the utility of an observation (e.g., to include observing "costs"), and computational issues.

1 Introduction

Incremental learning from experience, where one proceeds step by step to a desired goal, making decisions and asking questions on the basis of available information, is a basic aspect of human behavior. The classical paradigm for the scientific method, with its rigid sequence of hypothesis formation, followed by experiment and then analysis, bears little resemblance to this adaptive, self-adjusting learning behavior. The classical paradigm has served science well but its limitations are apparent in settings where data collection and analysis may proceed in concert, where learning proceeds on-the-fly and what has been learned from past data may be profitably used to alter the collection of future data.

We describe here an adaptive extension of the scientific method built on a model for scientific exploration where, after an initial setup phase, exploration proceeds by iterating a three-stage cycle: *Observation–Inference–Design*. Figure 1 depicts the flow of information through one such cycle.

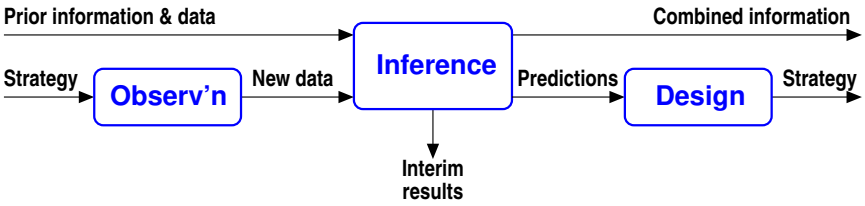

FIGURE 1. Information flow through one cycle of the adaptive exploration process. Information (e.g., data) and an observing strategy are input from a previous cycle on the left; combined information and a new observing strategy are output to the next cycle on the right.

In the observation stage, new data are obtained based on an observing strategy produced by the previous cycle of exploration. The inference stage synthesizes the information provided by previous and new observations to assess hypotheses of interest. This synthesis produces interim results such as signal detections, parameter estimates, or object classifications. Finally, in the design stage the results of inference are used to predict future data for a variety of possible observing strategies; the strategy that offers the greatest predicted improvement in inferences (subject to any resource constraints) is passed on to the next Observation–Inference–Design cycle.

The Bayesian approach to statistics provides ideal tools for developing a unified framework for adaptive exploration: Bayesian inference for the inference stage, and Bayesian experimental design for the design stage. Bayesian inference—using probability theory to combine prior information and data to produce posterior probabilities for hypotheses of interest—is a formal description of learning perfectly suited for the tasks of the inference stage of the exploration cycle. It is now widely used in several astronomical disciplines and its basic features will be familiar to many astronomers. In contrast, formal methods for experimental design (Bayesian or otherwise) will likely be new to most astronomers. Bayesian design—an application of Bayesian decision theory—identifies an optimal experimental or observational design by first specifying the purpose for a study, and then comparing how well candidate designs achieve that purpose by using the techniques of Bayesian inference to predict and analyze future data. A main goal of this brief paper is to introduce astronomers to Bayesian design, in the context of adaptive exploration.

In 1956, Lindley described how one could use tools from information theory and Bayesian statistics to compare experimental designs when one's purpose is simply to gain knowledge about a phenomenon [Lin56]. He later incorporated these ideas into a more general theory of Bayesian experimental design, described in his influential 1972 review of Bayesian statistics [Lin72]. Although non-Bayesian methods for optimal design predate Lindley's work (standard references are [Fed72, Che72, AF97]), the Bayesian approach provides a more fundamental rationale for many earlier methods,

and unifies and generalizes them (see [CV95] for discussion of the relationships between Bayesian and non-Bayesian design). In the three decades since Lindley's review, the theory of design has matured significantly. But as noted in Toman's recent review, "unfortunately much of the work in this area remains purely theoretical" [Tom99]. This is largely due to the computational complexity of Bayesian design, an obstacle noted already in Lindley's foundational work. In experimental design, one must account for both uncertainty regarding the hypotheses under consideration, and uncertainty about the values of future data. For the former, one must perform the difficult parameter space integrals that are characteristic of Bayesian inference [Lor99]; for the latter, one must additionally integrate in the sample space as is typically done in frequentist calculations. In a sense, experimental design is the arena in which the Bayesian and frequentist outlooks meet, producing problems with the combined complexity of both approaches.

As a result of this complexity, the vast majority of research in optimal design (Bayesian or non-Bayesian) has focused on simple models for which the required integrals can be evaluated analytically, such as linear models with additive Gaussian errors. Existing work on nonlinear design typically linearizes about a best-fit model [Mac92, SS98]. But the last decade has seen enormous strides in Bayesian computation due largely to the development of sampling-based methods for evaluating parameter space integrals, particularly Markov Chain Monte Carlo (MCMC) methods. Such methods not only facilitate rigorous calculations with complicated models; they also provide results in a form that can be readily interpreted and processed by end-users, even when the hypothesis space is of large dimension. We describe them further below.

Only recently have sampling-based algorithms that combine parameter and data sampling been brought to bear on Bayesian design [MP95, CMP95, MP96, Mul99]. Here we use simple sampling algorithms to implement the adaptive exploration strategy outlined above in the context of a simple but realistic *non*linear astronomical design problem. The sampling approach not only allows us to evaluate integrals without approximating the integrands, but also allows straightforward graphical display of all elements of the calculation. We hope this example provides an accessible introduction to Bayesian experimental design for astronomers, as well as a demonstration of the potential of adaptive exploration.

The following section describes the motivation for our interest in adaptive exploration—optimal allocation of observing resources for the Space Interferometry Misssion—and then introduces adaptive exploration by example. We follow the strategy through one full cycle and through the observation and inference stages of a second cycle, using as an example radial velocity observations of a star with the goal of determining the orbital parameters of an unseen planetary companion. The final section discusses several directions for future research.

2 Example: Measuring an Exoplanet Orbit

Our work on adaptive exploration is motivated by the Space Interferometry Mission, the first main mission of NASA's *Origins* program.[1] SIM is designed to measure the directions to astronomical sources with unprecedented accuracy. In its highest precision mode it is expected to achieve 1 microarcsecond astrometric accuracy. This will allow detection of the reflex motion "wobble" of a star with an Earth-like planet at a distance of several parsecs, or with a Jupiter-like planet at kiloparsecs. But SIM's high-accuracy measurements are time consuming, seriously restricting the number of stars that can be examined in a search for extrasolar planets. SIM observations are thus a precious resource that must be optimally allocated (not only for planet searches, but also for other diverse science SIM will undertake). During the mission, targets with no planets must be quickly weeded out, and observations of targets with companions must be scheduled to optimally determine the number of planets and their orbital parameters so that SIM can characterize as many systems as possible. In addition, before the launch of the SIM spacecraft in 2009, the SIM project will undertake extensive preparatory observations in order to carefully select both science target stars and reference stars against which the motions of the science targets will be measured. Reference stars must be free of planetary companions that would complicate their motion. The SIM Extrasolar Planet Interferometric Survey (EPIcS) key project is considering using binary stars with eccentric orbits as reference stars, since planets will have been swept from such systems. The preparatory observing campaign must identify hundreds of such stars and measure their orbits with high precision. This will require a huge expenditure of observational resources that must be optimized.

As a simple example of the kind of problem that must be addressed for optimizing SIM mission and preparatory observing, we consider here the problem of making radial velocity (RV) measurements of a star in order to best determine the parameters of the orbit of an unseen Jupiter-mass companion. Observations of this type will comprise much of SIM preparatory observing; similar ideas will apply to analysis of astrometric data. We consider observations of a 1 $M_\odot$ star known to have a single planetary companion; our goal is to choose future observations to best improve our estimates of the planet's orbital parameters. The function giving the radial velocity vs. time for a star exhibiting Keplerian reflex motion has six parameters. To simplify the calculations, we focus here on the three most important parameters—the orbital period, τ, the eccentricity, e, and the velocity amplitude, K—and we presume the remaining geometric parame-

[1] For detailed information about SIM, see the SIM web site:
`http://sim.jpl.nasa.gov/`

ters are known a priori (these include the time of periastron crossing, the longitude of periastron, and the orbital inclination). We model the value of each datum d_i as having additive noise, so that

$$d_i = v(t_i; \tau, e, K) + e_i, \tag{1.1}$$

where $v(t; \tau, e, K)$ gives the velocity at time t as a function of the parameters, and e_i represents the unknown noise contribution for datum i. We take the noise to have independent Gaussian distributions with standard deviation $\sigma = 8$ m s^{-1} (typical of current RV surveys).

The first cycle of exploration requires a "setup" strategy specifying the initial observations. Ideally, such a strategy would be developed using design theory and predictions based solely on prior information about the possible orbits (e.g., an assumed period distribution for orbits). For simplicity, the setup strategy here specifies 10 equally-spaced velocity measurements.

2.1 Cycle 1: Observation

Figure 2a shows the results of the observation stage of the first Observation-Inference-Design cycle. The points with error bars show the results of 10 simulated observations. For reference, the dashed curve shows the true velocity curve, with $\tau = 800$ d, $e = 0.5$, and $K = 50$ m s^{-1} (typical parameters for current observations of Jupiter-like extrasolar planets). The observations span somewhat less than two periods.

2.2 Cycle 1: Inference

For the inference stage, we calculate the posterior probability density for the parameters given the available data. Bayes's theorem gives this as

$$p(\tau, e, K|D, I) \propto p(\tau, e, K|I)\, \mathcal{L}(\tau, e, K), \tag{1.2}$$

where $p(\tau, e, K|I)$ is the prior probability density for the orbital parameters, $\mathcal{L}(\tau, e, K)$ is the likelihood function (the probability for the data presuming τ, e, and K are known), and I denotes the modeling assumptions (Keplerian orbit, noise properties, etc.). We assume we have no significant prior knowledge of the parameters, and take the prior to be a constant. Our assumption of Gaussian noise probabilities leads to a likelihood proportional to $\exp[-\chi^2(\tau, e, K)/2]$, where $\chi^2(\tau, e, K)$ is the familiar goodness-of-fit statistic given by a weighted sum of squared residuals. Thus,

$$p(\tau, e, K|D, I) \propto \exp[-\chi^2(\tau, e, K)/2]. \tag{1.3}$$

To find best-fit parameters, we could maximize the posterior density (corresponding to minimizing χ^2). To constrain the parameters, we could

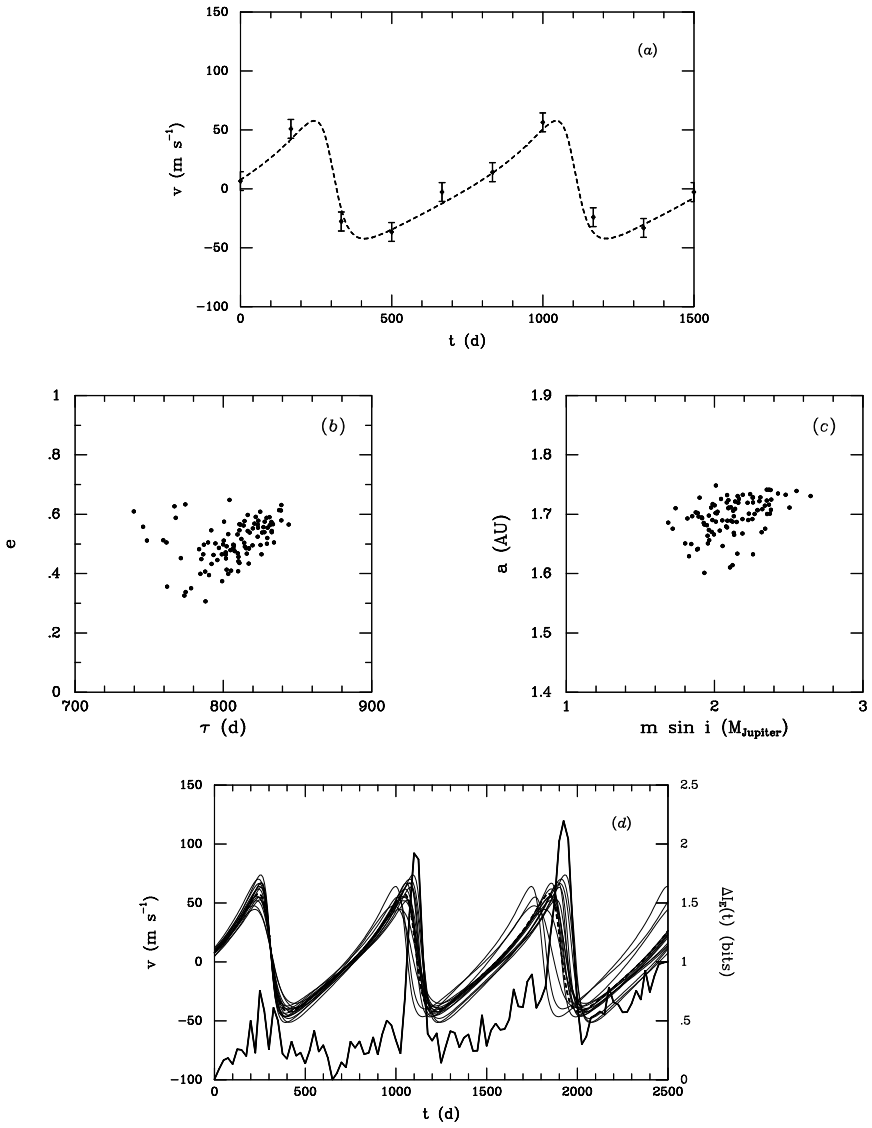

FIGURE 2. One cycle of the exploration process for simulated planet search data. (a) Observation stage, showing 10 simulated observations and true velocity curve (dashed). (b,c) Inference stage, showing samples from the posterior distribution for two velocity curve parameters (b) and two derived orbital parameters (c). (d) Design stage, showing predicted velocity curves (thin solid curves), true velocity curve (dashed curve), and the expected information gain for a sample at each time (thick solid curve, right axis).

locate the constant-χ^2 surface that encloses, say, 90% of the posterior probability for all three parameters; such a region is called a 90% (joint) credible region. If we were primarily interested in just the period, we could separately focus on it by calculating the marginal distribution for τ, given by integrating out the other parameters;

$$p(\tau|D,I) \propto \int de \int dK \ \exp[-\chi^2(\tau,e,K)/2]. \tag{1.4}$$

A 90% credible region for τ alone would be a region of the τ axis containing 90% of this marginal density.

All of these summaries of the posterior distribution could be calculated with common numerical methods (optimization and quadrature). But for problems with more dimensions, such calculations can be challenging. A more flexible approach is to use *posterior sampling* (see [Lor99] for a brief introduction and references). In this approach one constructs a random number generator that samples from the parameter space according to the posterior distribution (in contrast to more common Monte Carlo methods that sample from the data space). In this case, each sample would be a triplet (τ, e, K) drawn from $p(\tau, e, K|D, I)$; repeated sampling will produce a set of values, $\{\tau_j, e_j, K_j\}$. Once a set of such samples is available, many quantities of interest can be found by simple manipulations of the samples. In addition, posterior samples can be used directly to report results in a way that is easy to interpret and easy to use in future calculations.

Figures 2b and 2c are examples of interim results from the inference stage of the exploration cycle based on the observations shown in Figure 2a. We used a simple rejection method [PTVF92] to sample the posterior distribution; Figure 2b shows the τ and e coordinates of 100 such samples, displaying the marginal distribution $p(\tau, e|D, I)$. In a more careful calculation, we would use more samples and smoothing to find contours of credible regions; here it suffices to note that the displayed cloud of points should conservatively bound a 90% credible region. We see that the period and eccentricity are usefully constrained by the 10 data points, although significant uncertainty remains. Also, the posterior distribution is obviously not well-approximated by a Gaussian. Figure 2c shows how easily a complicated marginal distribution can be found using the samples; it displays the marginal distribution for the planet's semimajor axis, a, and $m \sin i$, the product of its mass and the sine of its orbital inclination. These are each nonlinear functions of the three model parameters. To produce Figure 2c we simply evaluated these functions for each of the 100 samples of (τ, e, K) already produced; this is much simpler than numerically evaluating the multiple integral defining the marginal distribution over a $(m \sin i, a)$ grid. By reporting the actual sample values, other investigators could use the results of these observations in their own calculations and fully account for the uncertainties simply by evaluating any quantities of interest over the set of samples.

2.3 Cycle 1: Design

For the design stage, we locate the time at which to make the next observation so that we have the best chance of significantly reducing our uncertainty in the parameters. We accomplish this in three steps: predict future data at various times, find the effect of the predicted data on inferences, and then identify the time for which the expected improvement in precision is greatest. We discuss each step in turn.

To predict the value, d, of a future datum at time t, we calculate the *predictive distribution*. To find it, we first predict d assuming we know the true parameter values, and then account for parameter parameter uncertainty by averaging over the parameter space. For given values of (τ, e, K), the predictive probability density for d is just the likelihood for d (a Gaussian centered at $v(t;\tau,e,K)$). The averaging weight we must use to account for parameter uncertainty is the posterior distribution from the inference stage. The predictive distribution is thus the convolution of the Gaussian likelihood for d and the posterior from the inference stage;

$$\begin{aligned} p(d|t,D,I) &= \int d\tau \int de \int dK\; p(\tau,e,K|D,I) \\ &\quad \times \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{[d-v(t;\tau,e,K)]^2}{2\sigma^2}\right) \\ &\approx \frac{1}{N} \sum_{\{\tau_j,e_j,K_j\}} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{[d-v(t;\tau_j,e_j,K_j)]^2}{2\sigma^2}\right) \end{aligned} \quad (1.5)$$

where the last line gives a Monte Carlo integration estimate of the predictive distribution using N posterior samples from the inference stage. To give some sense of what the predictive distribution looks like for various values of time, Figure 2c shows the $v(t)$ curves for the first 15 sampled parameter points as thin solid lines; the true curve is again displayed as a thick dashed curve. The ensemble of thin curves depicts our uncertainty in $v(t)$. The predicted data values at each time are additionally uncertain due to the noise which "blurs" the curves by 8 m s^{-1}. The ensemble of blurred curves represents the predictive distribution as a function of time. The uncertainty is greatest near times of periastron crossing when the velocity is changing most quickly (it is minimal at 300 d, the initial time of periastron crossing we assumed was known). Also, the uncertainty in the period makes the velocity uncertainty at periastron crossing grow with time as predictions with different periods fall increasingly out of synchronization.

Next we must measure how future data would affect our inferences. If datum d at time t were available, we could update our inferences simply by multiplying the posterior distribution from the previous stage by the likelihood function based on the single new datum (the Gaussian factor in equation (1.5)), and renormalizing. (This is equivalent to doing a new χ^2 calculation considering all 11 data points at once.) The new posterior,

$p(\tau, e, K|d, t, D, I)$, will hopefully be more informative about the parameters than the current one. The information in the posterior is given by the negative Shannon entropy of the posterior distribution,[2]

$$\mathcal{I}(d,t) = \int d\tau \int de \int dK\; p(\tau, e, K|d, D, I) \log[p(\tau, e, K|d, t, D, I)]. \quad (1.6)$$

This is the information gain for a particular datum at time t; to account for prediction uncertainty, we must calculate the *expected* information gain, averaging over d using the predictive distribution of equation (1.5):

$$\mathcal{EI}(t) = \int dd\, \mathcal{I}(d,t) p(d|t, D, I). \quad (1.7)$$

The best sampling time is the one that maximizes the information gain, so we must evaluate $\mathcal{EI}(t)$ as a function of time. For problems such as this where the width of the noise distribution does not depend on the value of the underlying signal, one can show that the expected information gain is equal to the entropy of the predictive distribution [SW97, SW00],

$$\mathcal{EI}(t) = -\int dd\, p(d|t, D, I) \log[p(d|t, D, I)]. \quad (1.8)$$

Thus the best sampling time is the time at which the entropy (uncertainty) of the predictive distribution is maximized. This is an eminently reasonable criterion: Bayesian design is telling us that we will learn the most by sampling where we know the least.

We use nested Monte Carlo methods to calculate $\mathcal{EI}(t)$ as a function of time. At each time, we sample a datum from the predictive distribution by first drawing a set of parameter values from the posterior, and then drawing a data value from the sampling distribution with those parameters. We then estimate $p(d|t, D, I)$ for that datum using equation (1.5). Repeating this process and averaging the logarithm of the estimates provides a Monte Carlo estimate of equation (1.8). The thick solid curve in Figure 2d shows this estimate of $\mathcal{EI}(t)$, using base-2 logarithms so that the relative information gain is measured in bits (with an offset so the smallest $\mathcal{EI}(t)$ is at 0 bits; the raggedness in the curve reflects the Monte Carlo uncertainties). $\mathcal{EI}(t)$ quantifies the uncertainty that is apparent in the set of thin sampled $v(t)$ curves. It is maximized near the periastron crossing subsequent to the available data, at $t = 1925$ d. Thus the observing strategy produced by this observation–inference–design cycle is: observe at $t = 1925$ d.

[2]For a Gaussian distribution, $\mathcal{I}$ is proportional to $-\log(\sigma)$ and thus increases with decreasing σ as one would expect; but it is a more general measure of spread than the standard deviation. To be formally correct, the argument of the logarithm in equation (1.6) should be divided by a measure on the parameter space so the argument is dimensionless; this has no significant effect on our results. An alternative definition of information is the cross-entropy or Kullback-Leibler divergence between the posterior and prior; it gives the same results as the Shannon entropy for this calculation [Mac92].

2.4 Cycle 2: Observation and Inference

Figure 3 shows the consequences of following this strategy. Figure 3a shows the previous data and a new datum obtained by simulating an observation at $t = 1925$ d. Incorporating this new datum into the posterior yields posterior samples shown in Figure 3b. We also used these samples to produce 15 predicted $v(t)$ curves in Figure 3a to display the velocity curve uncertainty after incorporating the new datum. Finally, Figure 3c shows the updated marginal distribution for the planet's mass and semimajor axis. Comparing to the corresponding panels in Figure 2, we see very significant reduction in uncertainty. In particular, the period uncertainty has decreased by more than a factor of two and the semi-major axis uncertainty is also drastically decreased; this was accomplished by incorporating the information *from a single well-chosen datum*. This is a dramatically larger increase in precision than one might have expected using rule-of-thumb "root-n" arguments based on random sampling. This is typical behavior for this problem; we have not chosen the simulated data set in any special way to obtain this behavior. It continues for subsequent cycles.

3 Challenges

This simple example illustrates the adaptive exploration methodology and demonstrates its potential. Several issues need to be addressed to make adaptive exploration useful in more complicated settings. Befitting a conference on statistical challenges, we close with a list of topics for future research. The field of experimental design has a wide and diverse literature spread across several disciplines, and some of these topics are being addressed in current research under such titles as sequential design, active data selection, and active, adaptive, or incremental learning.

In our example the goal was inference of the parameters of a system known to contain a single planet. In reality, the goals of inference may not be so clear-cut. Observers may not be sure a system has a planetary companion at the start of an exploration, so the goal is initially detection of a planet. Or if a system is chosen because it is known to have a companion, the goals may include detection of possible additional planets. At some point, the goal may shift from detection to estimation. How do design criteria for detection compare to those for estimation? When and how should the adaptive methodology shift its goal from detection to estimation? The work of Toman [Tom96] on Bayesian design for multiple hypothesis testing provides a starting point for addressing these questions.

Our utility function was simply the information provided by new data. In some settings, one may wish to incorporate other elements in the utility function, such as the cost of observing as a function of time or sample size. How can an observer map such costs to an information scale so that

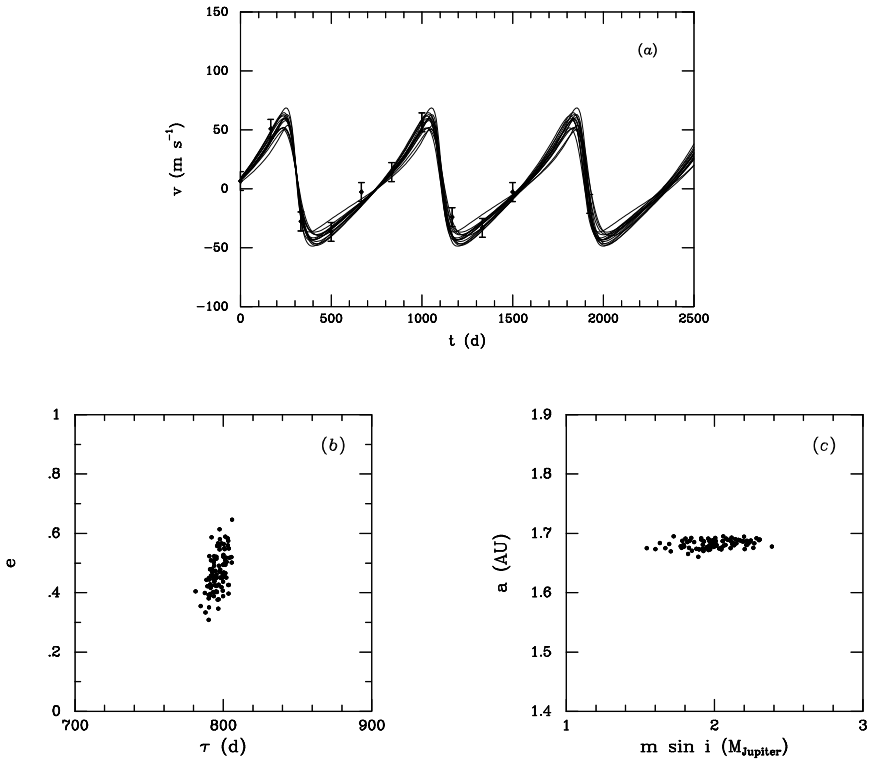

FIGURE 3. The beginning of the next cycle of the exploration process for simulated planet search data. (a) Observation stage, showing original 10 simulated observations, a new datum at 1925 d. Also shown are predicted velocity curves from the inference stage. (b,c) Inference stage, showing samples from the posterior distribution for two velocity curve parameters (b) and two derived orbital parameters (c). The single new datum has greatly increased the precision of inferences due to optimal selection of the observing epoch.

information and other costs or benefits can be combined into a single utility function?

We used a simple rejection method for generating posterior samples in our example. While attractively simple, in our experience such an approach will not be useful for problems with more than five or six parameters (even fairly sophisticated envelope functions will waste too many samples). The obvious tool for addressing this is MCMC, but the Markov chain must ultimately sample over both the parameter space and the sample space (of future observations). Are there MCMC algorithms uniquely suited to adaptive exploration? Müller and Parmigiani and their colleagues [MP95, CMP95, MP96, Mul99] have developed a variety of Monte Carlo approaches to Bayesian design in various settings that should be helpful in this regard. Also, since adaptive exploration offers the hope of quickly

reducing uncertainties, at some point it may make sense to linearize about the best-fit model and use analytic methods. Criteria need to be developed to identify when this is useful.

Finally, in our example, the observing strategy for the first cycle was chosen somewhat arbitrarily. Ideally, it would be chosen using design principles and prior information. This raises many practical and theoretical questions. What should the size of a "setup" sample be? Should adaptive exploration start after a single sample, or are there benefits (perhaps associated with computational complexity) for starting with larger samples? Can the algorithms used for analysis when several samples are available also be used for designing the setup strategy, or are different algorithms required if prior information is very vague? Clearly, there is overlap between these issues and those already raised. This kind of design issue has been addressed informally for planning observations for the Hubble Space Telescope Cepheid key project [FHM+94]. Can a more formal approach improve on such a priori designs?

We hope this brief introduction will encourage astronomers and statisticians to explore these issues together in a variety of astronomical contexts.

4 References

[AF97] A. C. Atkinson and V. V. Fedorov. Optimum design of experiments. In Samuel Kotz, Campbell B. Read, and David L. Banks, editors, *Encyclopedia of statistical sciences*, pages 107–114. Wiley, New York, 1997.

[Che72] H. Chernoff. *Sequential Analysis and Optimal Design.* SIAM, Philadelphia, 1972.

[CMP95] M. Clyde, P. Muller, and G. Parmigiani. Exploring expected utility surfaces by markov chains, 1995.

[CV95] K. Chaloner and I. Verdinelli. Bayesian experimental design: A review. *Stat. Sci.*, 10:273–304, 1995.

[Fed72] V. V. Fedorov. *Theory of Optimal Experiments.* Academic, New York, 1972.

[FHM+94] W. L. Freedman, S. M. Hughes, B. F. Madore, J. R. Mould, M. G. Lee, P. Stetson, R. C. Kennicutt, A. Turner, L. Ferrarese, H. Ford, J. A. Graham, R. Hill, J. G. Hoessel, J. Huchra, and G. D. Illingworth. The hubble space telescope extragalactic distance scale key project. 1: The discovery of cepheids and a new distance to m81. *Ap. J.*, 427:628–655, June 1994.

[Lin56] D. V. Lindley. On the measure of information provided by an experiment. *Ann. Stat.*, 27:986–1005, 1956.

[Lin72] D. V. Lindley. *Bayesian statistics–a review*. SIAM, Philadelphia, 1972.

[Lor99] T. J. Loredo. Computational technology for bayesian inference. In *ASP Conf. Ser. 172: Astronomical Data Analysis Software and Systems VIII*, volume 8, pages 297+, 1999.

[Mac92] D. MacKay. Information-based objective functions for active data selection. *Neural Computation*, 4(4):590–604, 1992.

[MP95] P. Muller and G. Parmigiani. Numerical evaluation of information theoretic measures. In D. A. Berry, K. M Chaloner, and J. F. Geweke, editors, *Bayesian Statistics and Econometrics: Essays in Honor of A. Zellner*, pages 397–406. Wiley, New York, 1995.

[MP96] Peter Muller and Giovanni Parmigiani. Optimal design via curve fitting of monte carlo experiments. *J. Am. Stat. Assoc.*, 90:1322–1330, 1996.

[Mul99] P. Muller. Simulation based optimal design. In J. O. Berger, J. M. Bernardo, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics 6*, 1999.

[PTVF92] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, 1992.

[SS98] P. Sebastiani and R. Settimi. First-order optimal designs for non-linear models. *J. Stat. Plan. Inf.*, 74:177–192, 1998.

[SW97] P. Sebastiani and H. P. Wynn. Bayesian experimental design and shannon information. In *1997 Proceedings of the Section on Bayesian Statistical Science*, pages 176–181. American Statistical Association, 1997.

[SW00] P. Sebastiani and H. P. Wynn. Maximum entropy sampling and optimal bayesian experimental design. *J. Roy. Stat. Soc. B*, 62:145–157, 2000.

[Tom96] B. Toman. Bayesian experimental design for multiple hypothesis testing. *J. Am. Stat. Assoc.*, 91:185–190, 1996.

[Tom99] B. Toman. Bayesian experimental design. In Samuel Kotz and N. L. Johnson, editors, *Encyclopedia of statistical sciences, Update Vol. 3*, pages 35–39. Wiley, New York, 1999.