Bayesian Adaptive Exploration

Tom Loredo & David Chernoff

Department of Astronomy, Cornell University

`loredo@astro.cornell.edu`

`http://www.astro.cornell.edu/staff/loredo/bayes/`

Outline

- *Bayesian* adaptive exploration
  - Inference
  - Decision theory
  - Experimental design
- Proof of concept: Exoplanets
  - Motivation: SIM EPIc Survey
  - Demonstration: A few BAE cycles
- Challenges

Bayesian Inference

Assess hypotheses by calculating their probabilities $p(H_i|\ldots)$ conditional on known and/or presumed information using the rules of probability theory.

Probability Theory

'OR' (sum rule) $$P(H_1 + H_2|I) = P(H_1|I) + P(H_2|I) - P(H_1, H_2|I)$$

'AND' (product rule) $$P(H_1, D|I) = P(H_1|I)\, P(D|H_1, I) = P(D|I)\, P(H_1|D, I)$$

Bayes's Theorem

$$P(H_i|D, I) = P(H_i|I)\,\frac{P(D|H_i, I)}{P(D|I)}$$

posterior $\propto$ prior $\times$ likelihood

Marginalization

Note that for exclusive, exhaustive $\{H_i\}$,

$$\sum_i P(D, H_i|I) = \sum_i P(D|I)P(H_i|D, I) = P(D|I)$$
$$= \sum_i P(H_i|I)P(D|H_i, I)$$

prior predictive for D = Average likelihood for H_i

$\rightarrow$ We can use $\{H_i\}$ as a "basis" to get $P(D|I)$. This is sometimes called "extending the conversation."

Bayesian Decision Theory

Decisions depend on consequences

Might bet on an improbable outcome provided the payoff is large if it occurs and the loss is small if it doesn't.

Utility and loss functions

Compare consequences via *utility* quantifying the benefits of a decision, or via *loss* quantifying costs.

Utility $= U(c, o)$

- c: Choice of action (decide b/t these)
- o: Outcome (what we are uncertain of)

Deciding amidst uncertainty

We are uncertain of what the outcome will be
→ average:

$$EU(c) = \sum_{\text{outcomes}} P(o|I)\, U(c, o)$$

The best choice maximizes the expected utility:

$$\hat{c} = \arg\max_{c} EU(c)$$

Bayesian Experimental Design

Basic principles

Choices = $\{e\}$, possible experiments (sample times, sample sizes...).

Outcomes = $\{d\}$, values of future data.

Utility balances value of d for achieving experiment goals against the cost of the experiment.

Choose the experiment that maximizes

$$EU(e) = \sum_d p(d|e, I)\, U(e, d)$$

To predict d we must know which of several hypothetical "states of nature" H_i is true. → Average over H_i:

$$EU(e) = \sum_{H_i} p(H_i|I) \sum_d p(d|H_i, e, I)\, U(e, d)$$

Average over both hypothesis and data spaces.

Information as Utility

Common goal: discern among the H_i.
→ Utility = information $\mathcal{I}(e, d)$ in $p(H_i|d, e, I)$:

$$\begin{aligned} U(e, d) &= \sum_{H_i} p(H_i|d, e, I) \log\left[p(H_i|d, e, I)\right] \\ &= -\text{Entropy of posterior} \end{aligned}$$

Design to maximize expected information.

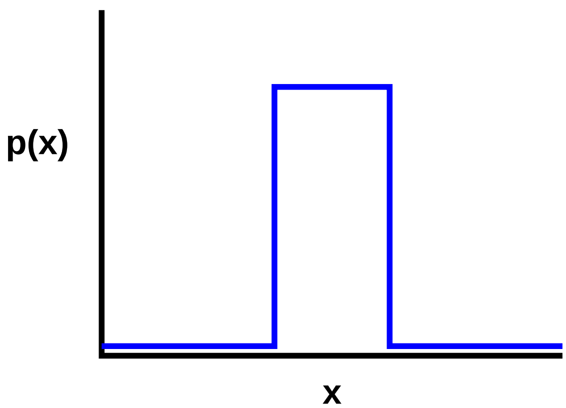
Measuring Information With Entropy

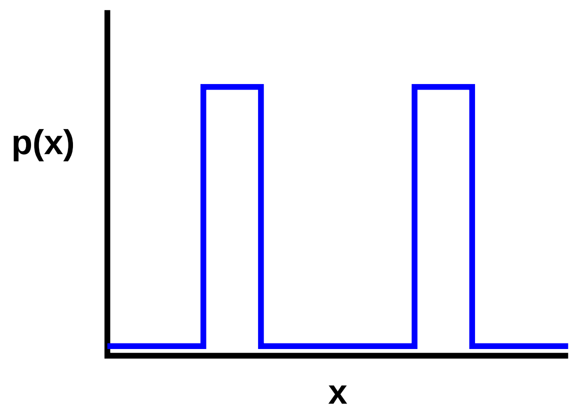
Entropy of a Gaussian

$$p(x) \propto e^{-(x-\mu)^2/2\sigma^2} \quad \rightarrow \quad \mathcal{I} \propto -\log(\sigma)$$

$$p(\vec{x}) \propto \exp\left[-\frac{1}{2}\vec{x} \cdot \mathbf{V}^{-1} \cdot \vec{x}\right] \quad \rightarrow \quad \mathcal{I} \propto -\log(\det \mathbf{V})$$

Entropy measures volume, not width



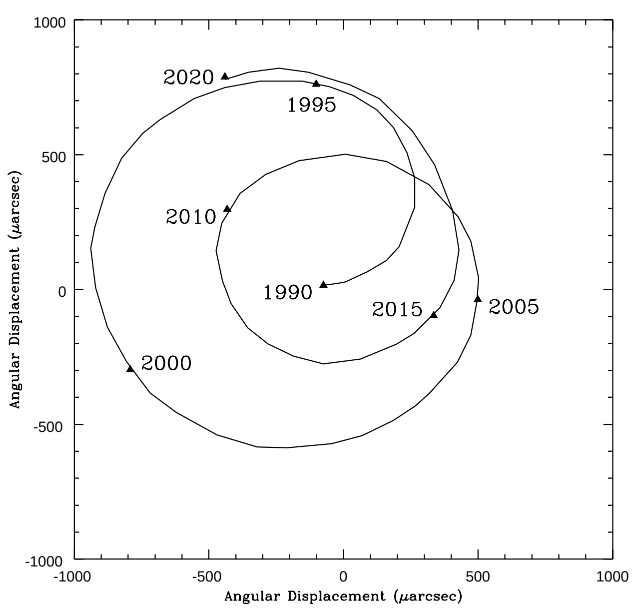


These distributions have the same entropy/amount of information.

Finding Exoplanets: The Space Interferometry Mission

SIM in 2009 (?)

Detecting Planets Without Seeing Them: The Sun's Wobble From 10 pc



EPIcS: Extrasolar Planet Interferometric Survey

Tier 1

- Goal: Identify Earth-like planets in habitable regions around nearby Sun-like stars
- Requires 1 μas astrometry
  - Long integration times
  - Astrometrically stable reference stars
- $\sim$ 75 MS stars within 10 pc, $\sim$ 70 epochs per target

Tier 2

- Goal: Explore the nature and evolution of planetary systems in their full variety
- Requires 4 μas astrometry, short integration times
- $\sim$ 1000 targets, "piggyback" on Tier 1

Preparatory observing

- High precision radial velocity and adaptive optics observing
- Identify science targets
- Identify reference stars (K giants? eccentric binaries?)

Huge resource expenditures

$\rightarrow$ must optimize use of resources

Example: Orbit Estimation With Radial Velocity Observations

Data are Kepler velocity plus noise:

$$d_i = V(t_i; \tau, e, K) + e_i$$

3 remaining geometrical params (t_0, λ, i) are fixed.

Noise probability is Gaussian with known $\sigma = 8$ m s^{-1}.

Simulate data with "typical" Jupiter-like exoplanet parameters:

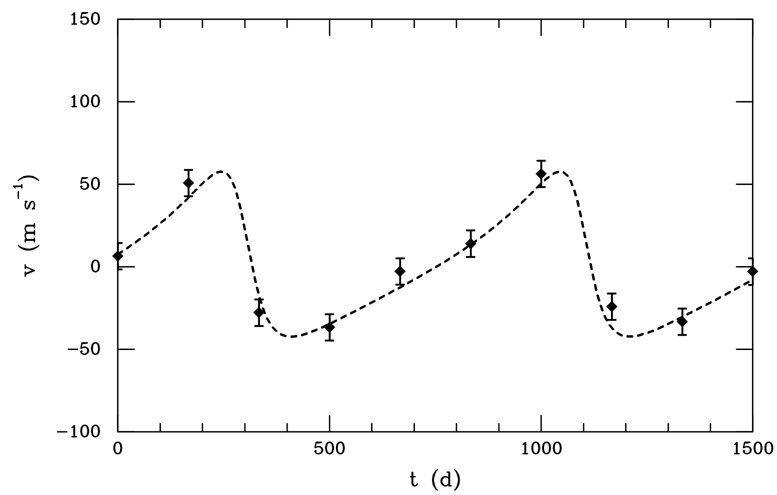
$$\begin{aligned} \tau &= 800 \text{ d} \\ e &= 0.5 \\ K &= 50 \text{ ms}^{-1} \end{aligned}$$

Goal: Estimate parameters τ, e and K.

Cycle 1: Observation and Inference

Initial observations

Prior "setup" stage specifies 10 equispaced observations.



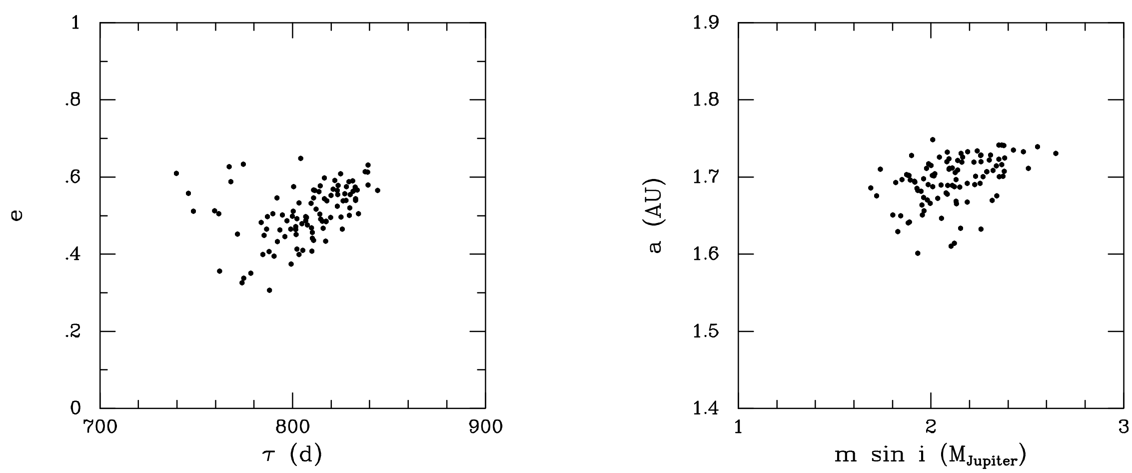
Inferences

Use flat priors,

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

χ^2 = familiar weighted sum of squared residuals.

Generate $\{\tau_j, e_j, K_j\}$ via posterior sampling.



Aside: Kepler Periodograms

Bayesian periodograms (Bretthorst)

Data are superposition of periodic functions + noise:

$$d_i = \sum A_\alpha g_\alpha(t_i;\omega;\theta) + e_i$$

Calculate $\mathcal{L}(\{A\},\omega,\theta)$ using χ^2.

Integrate out A's → least squares + volume factors:

$$p(\omega,\theta|D) \propto p(\omega,\theta) J(\omega,\theta) \exp\left[-\frac{r^2(\omega,\theta)}{2}\right]$$

Integrate out θ numerically → $p(\omega|D)$;

$$S(\omega) \equiv \ln\left[p(\omega|D)\right]$$

Generalizes Schuster periodogram & LSP.

Radial Kepler periodogram

$$V(t) = A_1 + A_2[e + \cos\upsilon(t)] + A_3 \sin\upsilon(t)$$

$$\upsilon(t) = f(t;\tau,e,T) \qquad \text{via Kepler's eqn}$$

Period τ
3 linear amplitudes (COM velocity, orbital velocity, λ)
2 other nonlinear parameters (e, T)

Follow the recipe! For $e = 0$ → LSP.

For astrometry, 2D data require $x(t)$, $y(t)$.
Extra parameters: inclination, parallax, proper motion.

Cycle 1: Design

Predict value of future datum at t

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

Effect of a datum on inferences

Information if we sample at t and get datum d:

$$
\mathcal{I}(d,t) = \int d\tau dedK\, p(\tau,e,K|d,t,D,I) \log[p(\tau,e,K|d,t,D,I)]
$$

Average over unknown datum value

Expected information:

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

Width of noise dist'n is independent of value of the signal$\rightarrow$

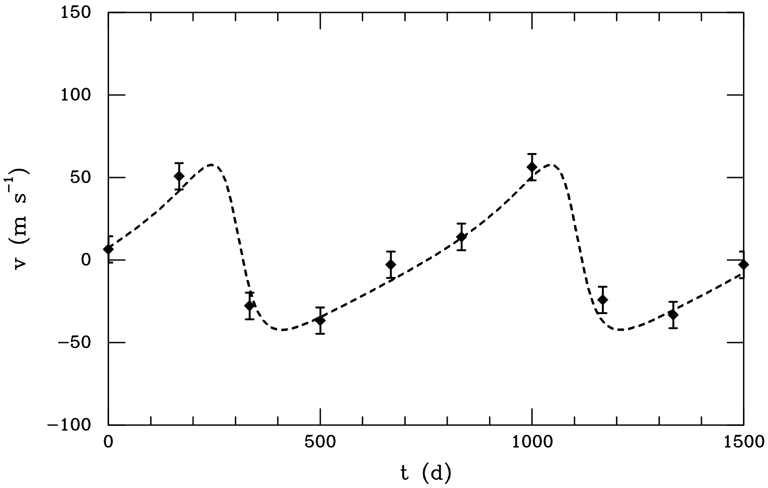
$$
\mathcal{EI}(t) = -\int dd\, p(d|t,D,I) \log[p(d|t,D,I)]
$$

Maximum entropy sampling.
(Sebastiani & Wynn 1997, 2000)

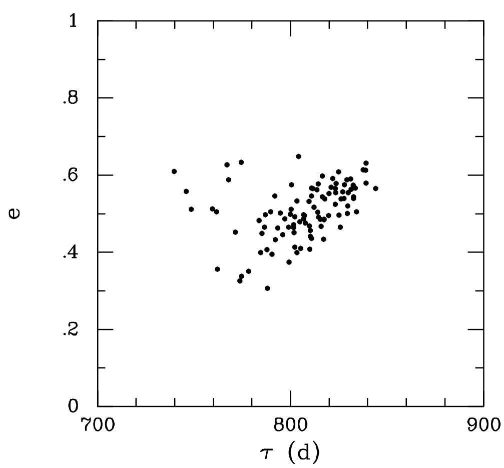
Evaluate by Monte Carlo using posterior samples & data samples. Pick t to maximize $\mathcal{EI}(t)$.

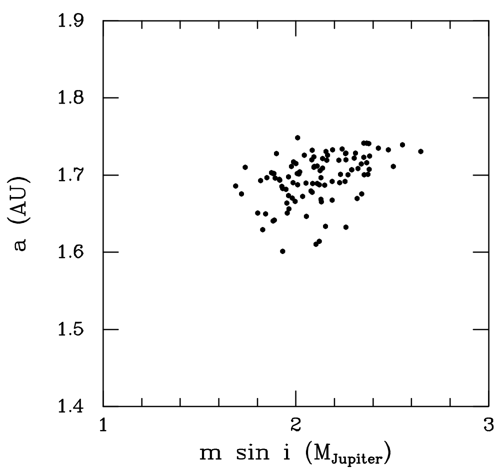
Cycle 1

Observation

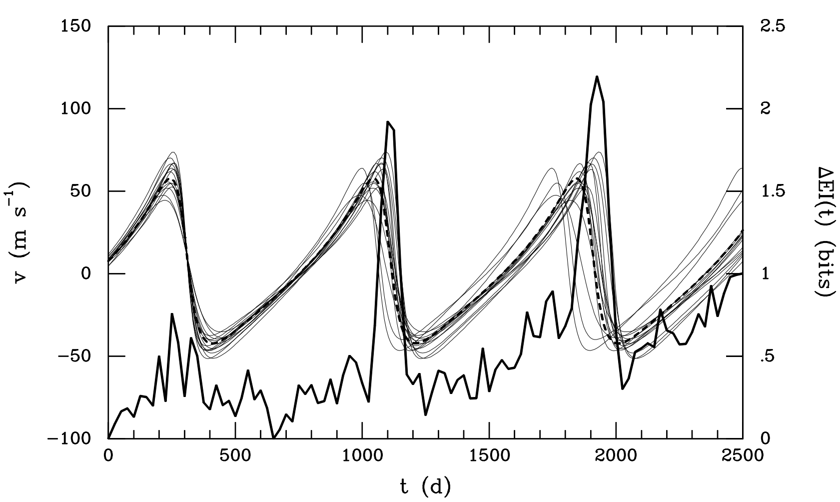


Inference



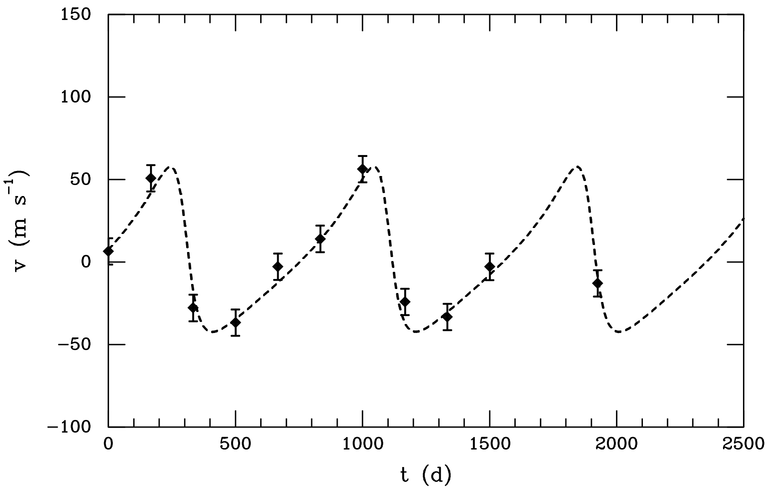


Design

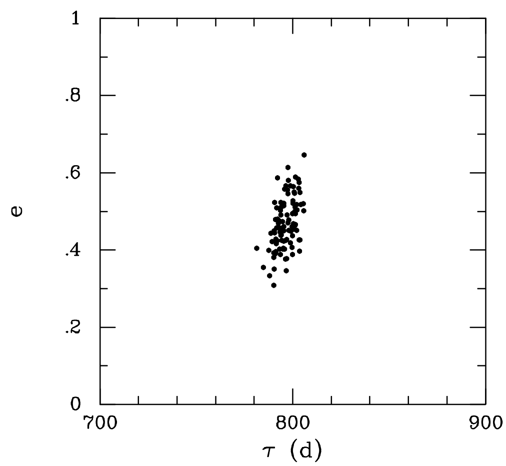


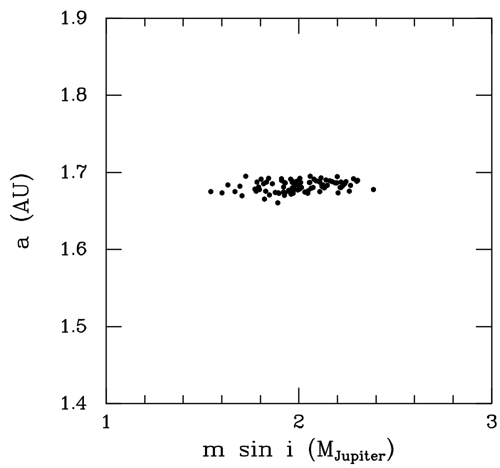
Cycle 2

Observation

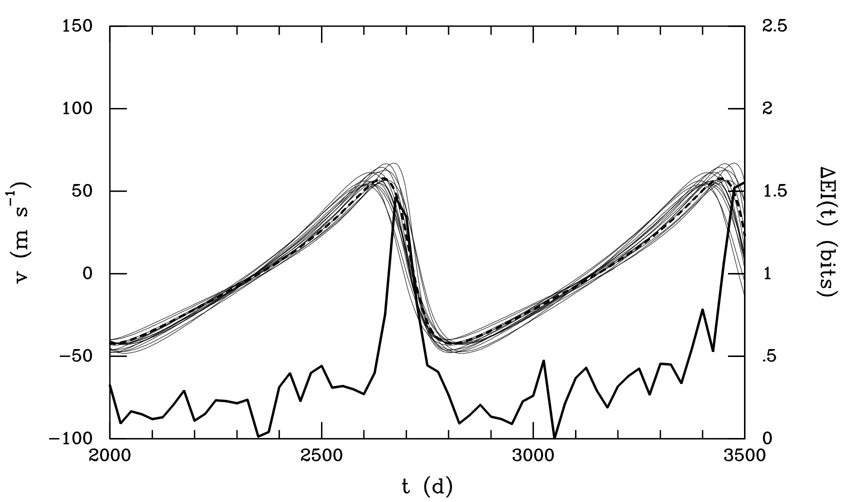


Inference



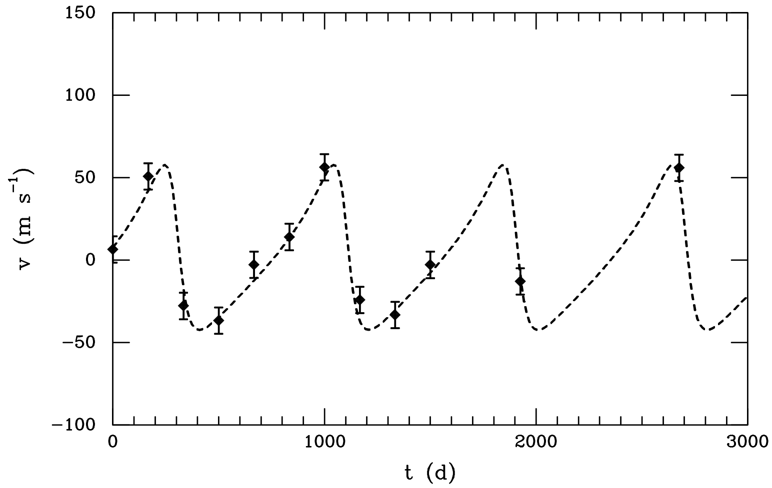


Design

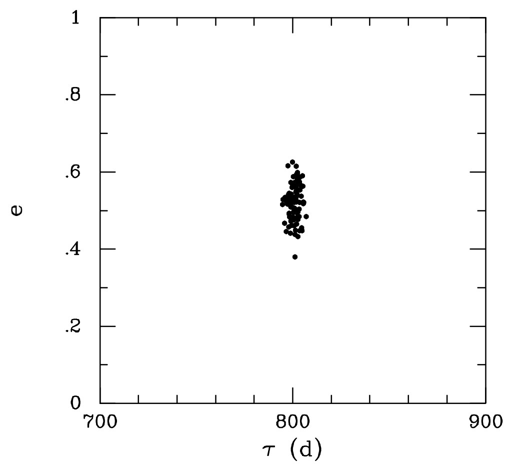


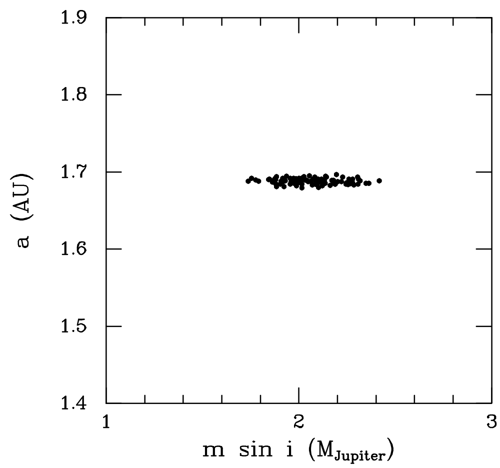
Cycle 3

Observation

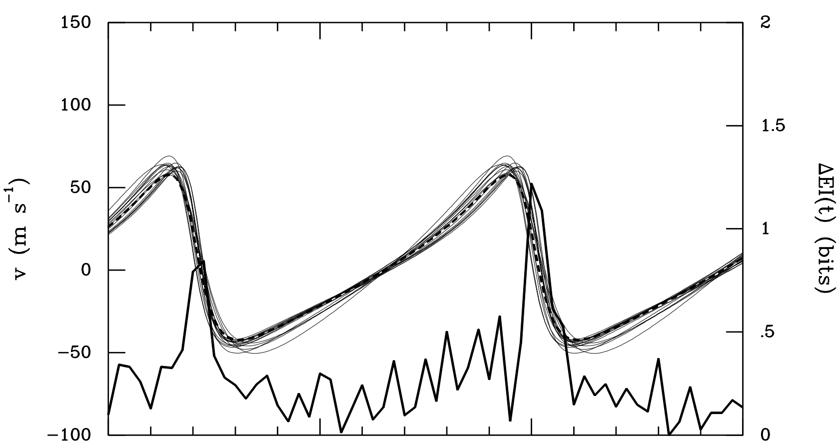


Inference



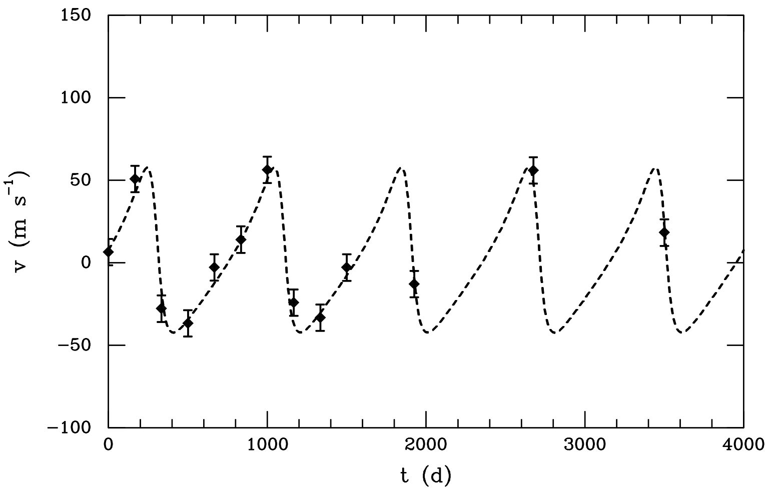


Design

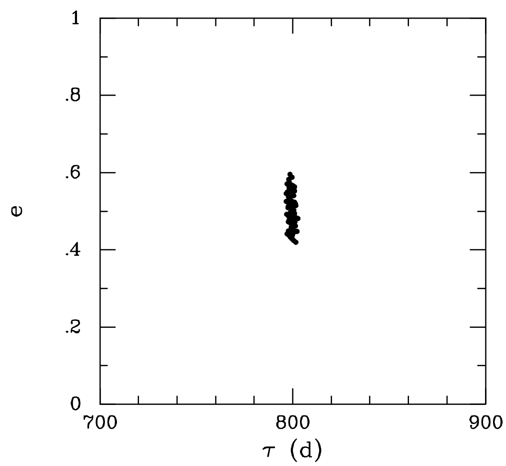


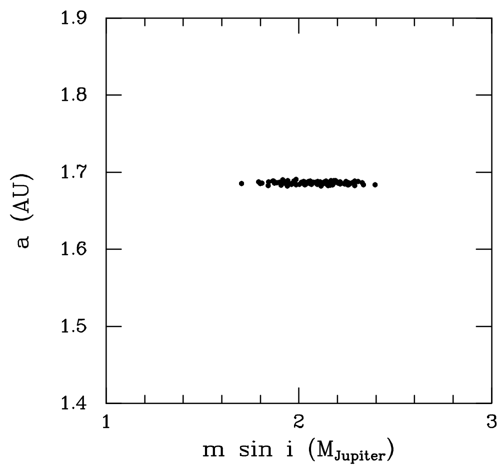
Cycle 4

Observation



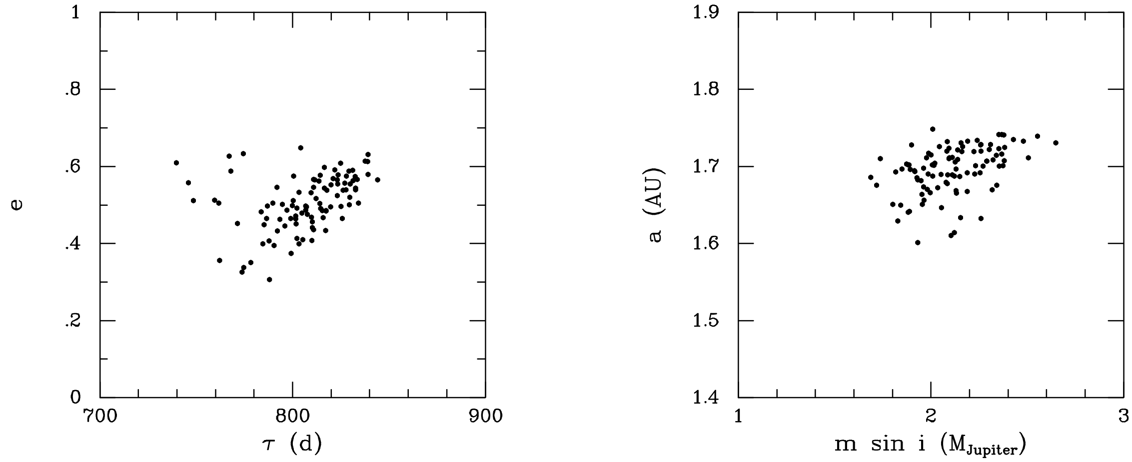
Inference



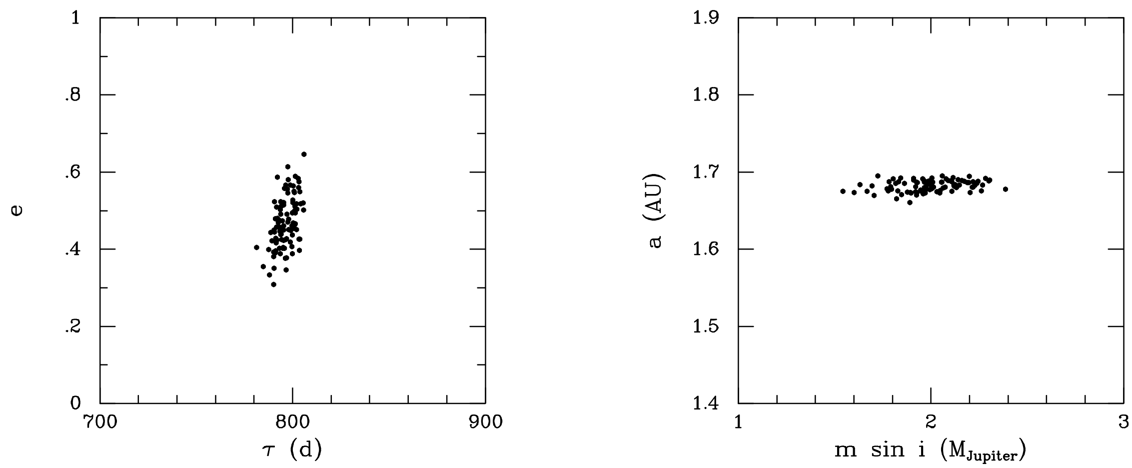


Evolution of Inferences

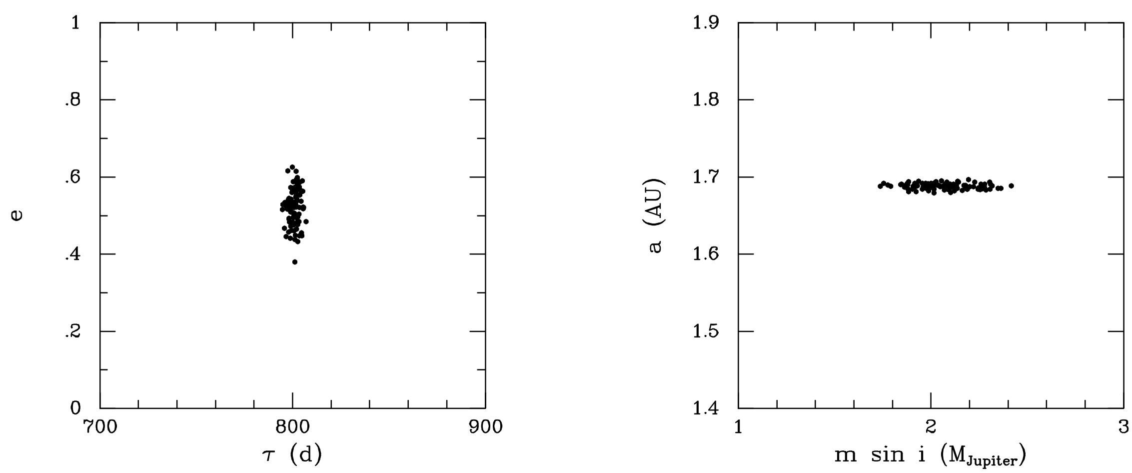
Cycle 1 (10 samples)



Cycle 2 (11 samples)

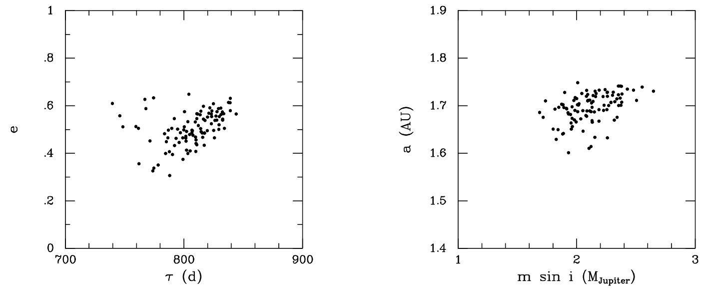


Cycle 3 (12 samples)

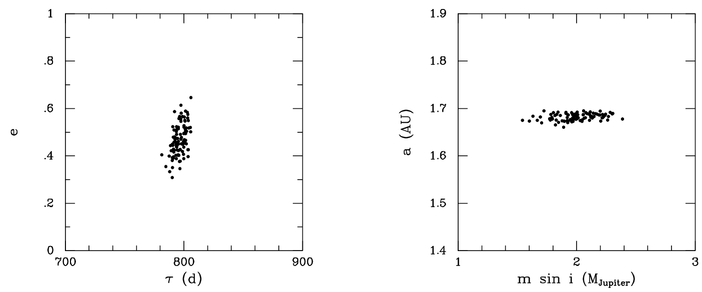


Evolution of Inferences

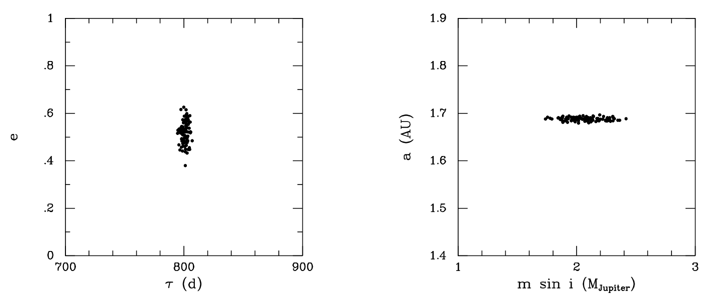
Cycle 1 (10 samples)



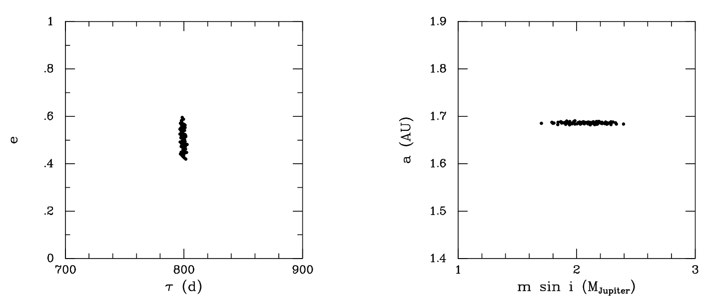
Cycle 2 (11 samples)



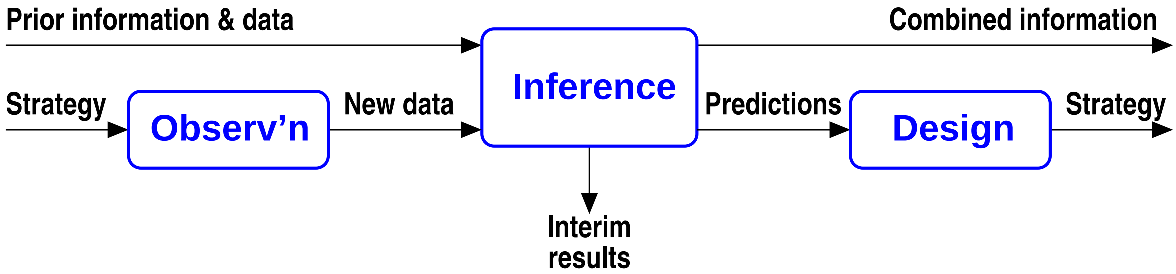
Cycle 3 (12 samples)



Cycle 4 (13 samples)



Challenges



Evolving goals for inference

Goal may originally be detection (model comparison), then estimation. How are these related? How/when to switch?

Generalizing the utility function

Cost of a sample vs. time or costs of samples of different size could enter utility. How many bits is an observation worth?

Computational algorithms

Are there MCMC algorithms uniquely suited to adaptive exploration? When is it smart to linearize?

Design for the "setup" cycle

What should the size of a setup sample be? Can the same algorithms be used for setup design?

Related fields

Sequential design, active data selection, and active, adaptive, incremental learning...