Optimal Scheduling of Exoplanet Observations via Bayesian Adaptive Exploration

Tom Loredo
Dept. of Astronomy, Cornell University
http://www.astro.cornell.edu/staff/loredo/

GISS Workshop — 25 Feb 2011

First, a word from our sponsor (NASA!)

Google

nasa astrophysics program review | Search

About 723,000 results (0.24 seconds)   Advanced search

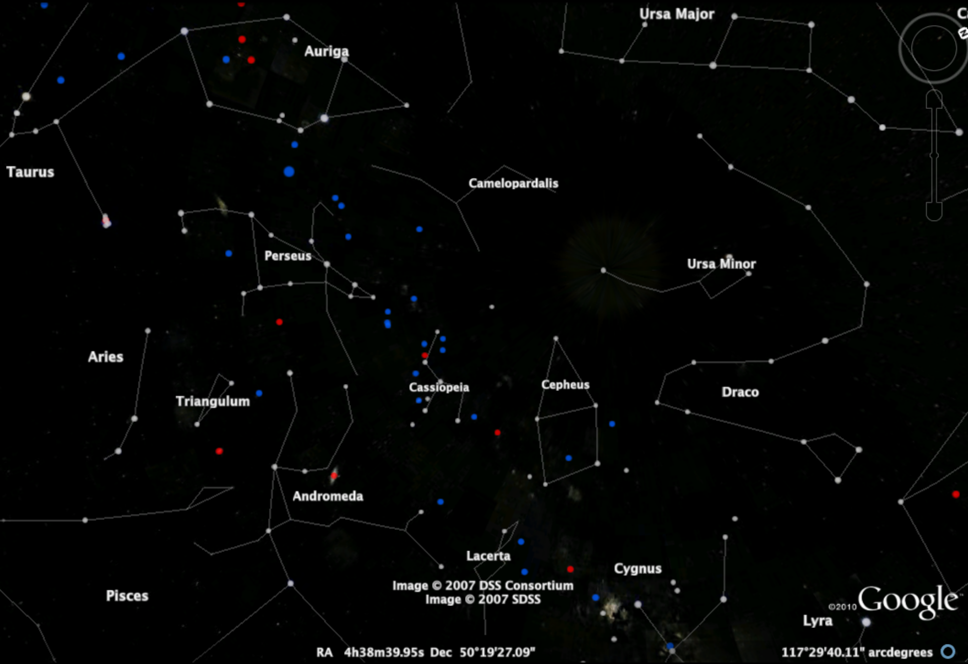
Astrophysics Research Program Review

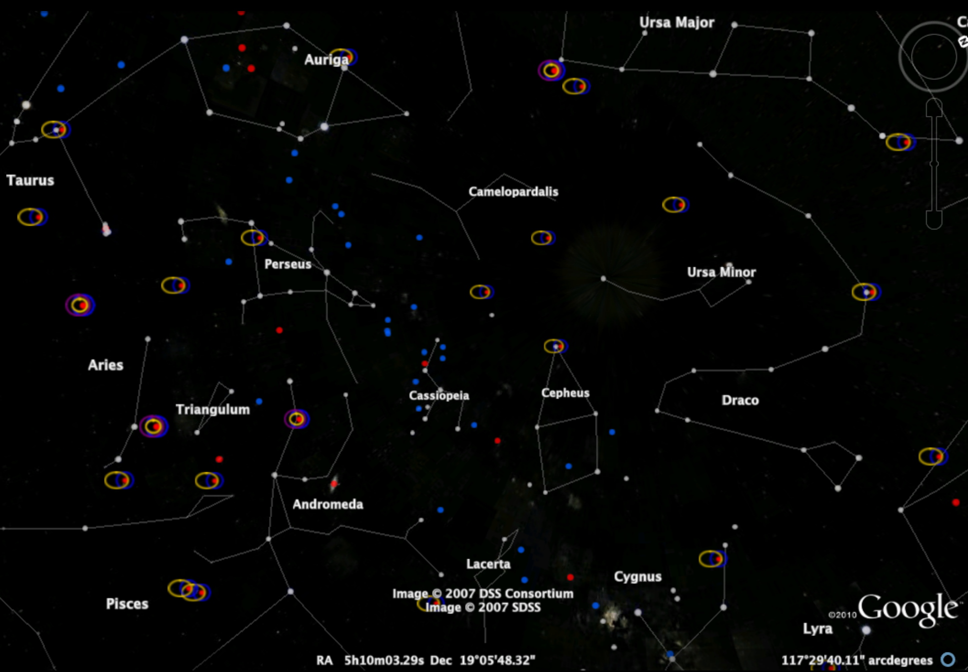
The Astrophysics Division of NASA's Science Mission Directorate will conduct a review of its programs in Research, Analysis and Enabling Technology. This review will assess the effectiveness of these programs in maximizing the scientific productivity from NASA's current and future missions, in the context of the Science Mission Directorate's Science Plan and the Astro2010 Decadal Survey → . Read the charter for this review.

The Astrophysics Division has appointed a review panel, which held its first meeting in mid-December 2010. A Splinter Meeting was held at the American Astronomical Society's January 2011 meeting in Seattle for public comments from the astronomical community: read the presentation (PDF). Comments are also welcome electronically, via the website at http://astroresearchreview.nsstc.nasa.gov/portal/ → from 18 January 2011. The review committee should report its findings in the summer of 2011.

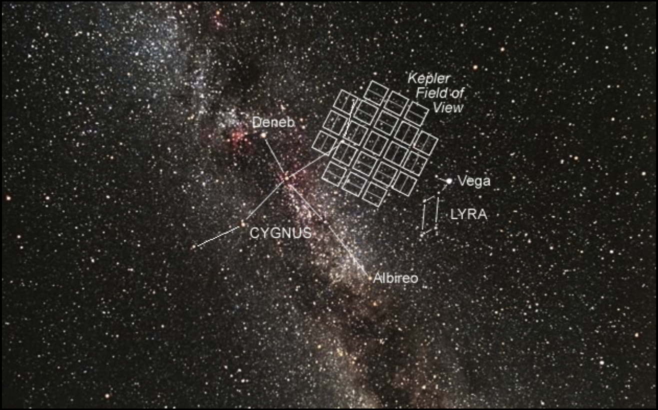
Linda Sparke
Astrophysics Research Program Manager, NASA HQ

Jon Morse
Director, Astrophysics Division, NASA HQ

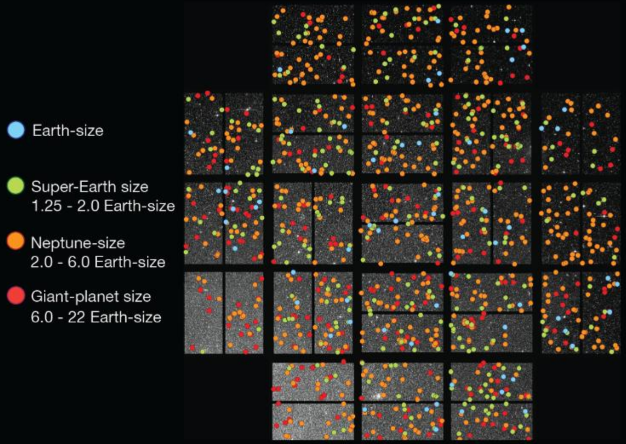

Ursa Major
Auriga
Taurus
Camelopardalis
Perseus
Ursa Minor
Aries
Cassiopeia
Cepheus
Draco
Triangulum
Andromeda
Lacerta
Cygnus
Pisces
Image © 2007 DSS Consortium
Image © 2007 SDSS
Lyra
©2010 Google
RA 4h38m39.95s Dec 50°19'27.09"
117°29'40.11" arcdegrees



Ursa Major
Auriga
Taurus
Camelopardalis
Perseus
Ursa Minor
Aries
Triangulum
Cassiopeia
Cepheus
Draco
Andromeda
Lacerta
Cygnus
Pisces
Image © 2007 DSS Consortium
Image © 2007 SDSS
©2010 Google
Lyra
RA 5h10m03.29s Dec 19°05'48.32"
117°29'40.11" arcdegrees


Kepler
Field of
View
Deneb
Vega
LYRA
CYGNUS
Albireo

Locations of Kepler Planet Candidates



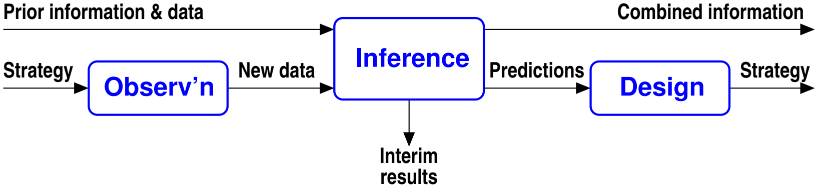
Scientific method

Science is more than a body of knowledge; it is a way of thinking.
The method of science, as stodgy and grumpy as it may seem,
is far more important than the findings of science.
—Carl Sagan

Classic hypothetico-deductive approach

- Form hypothesis (based on past observation/experiment)
- Devise experiment to test predictions of hypothesis
- Perform experiment
- Analysis $\rightarrow$
  - Devise new hypothesis if hypothesis fails
  - Devise new experiment if hypothesis corroborated

Bayesian Adaptive Exploration



- Observation — Gather new data based on observing plan
- Inference — Interim results via posterior sampling
- Design — Predict future data; explore where expected information from new data is greatest

Agenda

1. Motivation: Exoplanets via Doppler RV observations

2. Bayesian adaptive exploration

3. Toy problem: Bump hunting

4. BAE for HD 222582

Agenda

1. **Motivation: Exoplanets via Doppler RV observations**

2. Bayesian adaptive exploration

3. Toy problem: Bump hunting

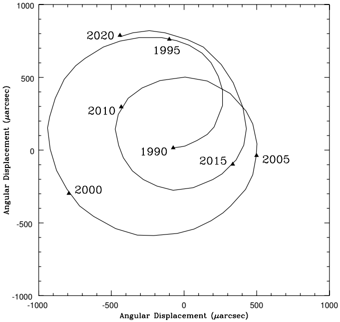
4. BAE for HD 222582

Finding Exoplanets via Stellar Reflex Motion

All bodies in a planetary system orbit wrt the system's center of mass, *including the host star*:

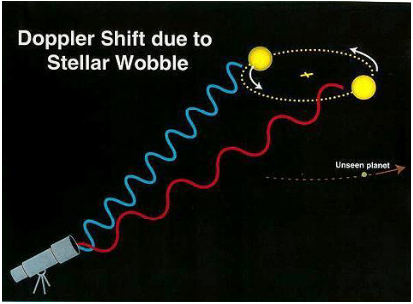
Astrometric Method

Sun's Astrometric Wobble from 10 pc



Doppler Radial Velocity (RV) Method

Doppler Shift Along Line-of-Sight

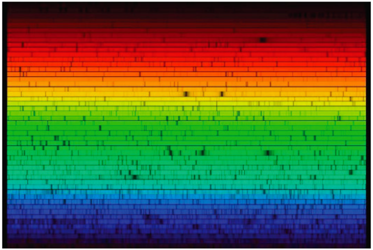


≈ 490 of ≈ 530 currently confirmed exoplanets found using RV method

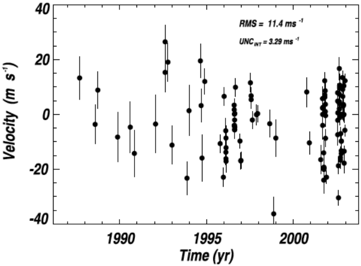
RV method is used to confirm & measure transiting exoplanet candidates

RV Data Via Precision Spectroscopy

Millipixel spectroscopy



Meter-per-second velocities



A Variety of Related Statistical Tasks

- *Planet detection* — Is there a planet present? Are multiple planets present?
- *Orbit estimation* — What are the orbital parameters? Are planets in multiple systems interacting?
- *Orbit prediction* — What planets will be best positioned for follow-up observations?
- *Population analysis* — What types of stars harbor planets? With what frequency? What is the distribution of planetary system properties?
- **Optimal scheduling** — How may astronomers best use limited, expensive observing resources to address these goals?

Bayesian approach tightly integrates these tasks

Agenda

1. Motivation: Exoplanets via Doppler RV observations
2. **Bayesian adaptive exploration**
3. Toy problem: Bump hunting
4. BAE for HD 222582

Experimental Design as Decision Making

When we perform an experiment we have choices of actions:

- What sample size to use
- What times or locations to probe/query
- Whether to do one sensitive, expensive experiment or several less sensitive, less expensive experiments
- Whether to stop or continue a sequence of trials
- . . .

We must choose amidst uncertainty about the data we may obtain and the resulting consequences for our experimental results.

$\Rightarrow$ Seek a principled approach for optimizing experiments, accounting for all relevant uncertainties.

Bayesian Decision Theory

Decisions depend on consequences

Might bet on an improbable outcome provided the payoff is large if it occurs and/or 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.

$$\text{Utility} = U(a, o) \qquad \begin{aligned} a &= \text{Choice of action (decide b/t these)} \\ o &= \text{Outcome (what we are uncertain of)} \end{aligned}$$

Loss $L(a, o) = U_{\max} - U(a, o)$

Bayesian Decision Theory

Uncertainty & expected utility

We are uncertain of what the outcome will be
$\rightarrow$ *average over outcomes*:

$$\mathbb{E}U(a) = \sum_{\text{outcomes}} P(o|\ldots)\, U(a,o)$$

The best action *maximizes the expected utility*:

$$\hat{a} = \arg\max_{a} \mathbb{E}U(a)$$

I.e., minimize expected loss.

Axiomatized: von Neumann & Morgenstern; Ramsey, de Finetti, Savage

Bayesian Experimental Design

Actions = $\{e\}$, possible experiments (sample sizes, sample times/locations, stopping criteria . . .).

Outcomes = $\{d_e\}$, values of future data from experiment e.

Utility measures value of d_e for achieving experiment goals, possibly accounting for the cost of the experiment.

Choose the experiment that maximizes

$$\mathbb{E}U(e) = \sum_{d_e} p(d_e|\ldots)\, U(e, d_e)$$

To predict d_e we must consider various hypotheses, H_i, for the data-producing process $\rightarrow$ Average over H_i uncertainty:

$$\mathbb{E}U(e) = \sum_{d_e} \left[\sum_{H_i} p(H_i|\ldots)p(d_e|H_i,\ldots)\right] U(e, d_e)$$

A Hint of Trouble Ahead

Multiple sums/integrals

$$\mathbb{E}U(e) = \sum_{d_e} \left[\sum_{H_i} p(H_i|I)p(d_e|H_i, I) \right] U(e, d_e)$$

Average over both hypothesis and data spaces

Plus an optimization

$$\hat{e} = \arg\max_{e} \mathbb{E}U(e)$$

Aside: The dual averaging—over hypothesis and data spaces—hints (correctly!) of connections between Bayesian and frequentist approaches.

Information-Based Utility

Many scientific studies do not have a single, clear-cut goal.

Broad goal: Learn/explore, with resulting information made available for a variety of future uses.

Example: Astronomical measurement of orbits of minor planets or exoplanets

- Use to infer physical properties of a body (mass, habitability)
- Use to infer distributions of properties among the population (constrains formation theories)
- Use to predict future location (collision hazard; plan future observations)

Motivates using a "general purpose" utility that measures *what is learned* about the H_i describing the phenomenon

Information Gain as Entropy Change

Entropy and uncertainty

Shannon entropy = a scalar measure of the degree of uncertainty expressed by a probability distribution

$$\begin{aligned} \mathcal{S} &= \sum_i p_i \log \frac{1}{p_i} \qquad \text{"Average surprisal"} \\ &= -\sum_i p_i \log p_i \end{aligned}$$

Information gain

Existing data $D \rightarrow$ interim posterior $p(H_i|D)$
Information gain upon learning d = decrease in uncertainty:

$$\begin{aligned} \mathcal{I}(d) &= \mathcal{S}[\{p(H_i|D)\}] - \mathcal{S}[\{p(H_i|d, D)\}] \\ &= \sum_i p(H_i|d, D) \log p(H_i|d, D) - \text{Const (wrt } d) \end{aligned}$$

Lindley (1956, 1972) and Bernardo (1979) advocated using $\mathcal{I}(d)$ as utility

A 'Bit' About Entropy

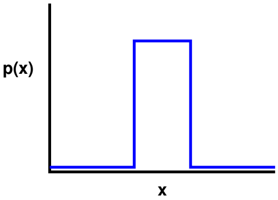
Entropy of a Gaussian

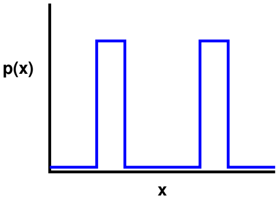
$$p(x) \propto e^{-(x-\mu)^2/2\sigma^2} \quad \rightarrow \quad \mathcal{I} \propto -\log(\sigma)$$

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

$\rightarrow$ Asymptotically like Fisher matrix criteria

Entropy is a log-measure of "volume," not range





These distributions have the same entropy/amount of information.

Prediction & expected information

Information gain from datum d_t at time t:

$$\mathcal{I}(d_t) = \sum_i p(H_i|d_t, D) \log p(H_i|d_t, D)$$

We don't know what value d_t will take $\rightarrow$ average over prediction uncertainty

Expected information at time t:

$$\mathbb{E}\mathcal{I}(t) = \int dd_t \; p(d_t|D) \, \mathcal{I}(d_t)$$

Predictive distribution for value of future datum:

$$\begin{aligned} p(d_t|D) &= \sum_i p(d_t, H_i|D) = \sum_i p(H_i|D) \, p(d_t|H_i) \\ &= \sum \text{Interim posterior} \times \text{Single-datum likelihood} \end{aligned}$$

There is a heck of a lot of averaging going on!

Computation

MaxEnt sampling for parameter estimation cases

Setting:

- We have specified a model, M, with uncertain parameters θ
- We have data $D \to$ current posterior $p(\theta|D, M)$
- The entropy of the noise distribution doesn't depend on θ,

$$\to \mathbb{E}\mathcal{I}(t) \;\; = \;\; \text{Const} - \int dd_t \; p(d_t|D, I) \log p(d_t|D, I)$$

Maximum entropy sampling.

(Sebastiani & Wynn 1997, 2000)

To learn the most, sample where you know the least.

Nested Monte Carlo integration for $\mathbb{E}\mathcal{I}$

Entropy of predictive dist'n:

$$\mathcal{S}[d_t|D, M] = -\int dd_t \; p(d_t|D, M_1) \log p(d_t|D, M)$$

- *Sample* predictive via θ ∼posterior, d_t ∼sampling dist'n given θ
- *Evaluate* predictive as θ-mixture of sampling dist'ns

Posterior sampling in parameter space

- Many models are (linearly) *separable* → handle linear "fast" parameters analytically
- When priors prevent analytical marginalization, use interim priors & importance sampling
- Treat nonlinear "slow" parameters via adaptive or population-based MCMC; e.g., diff'l evolution MCMC

Agenda

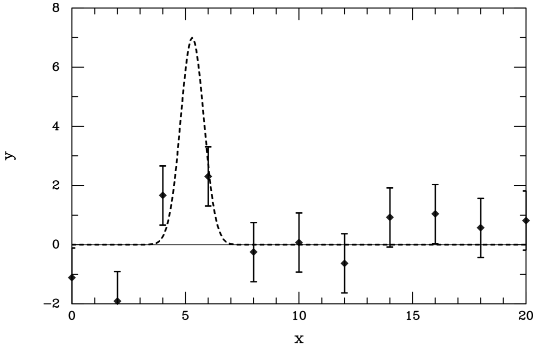
1. Motivation: Exoplanets via Doppler RV observations
2. Bayesian adaptive exploration
3. **Toy problem: Bump hunting**
4. BAE for HD 222582

Locating a bump

Object is 1-d Gaussian of unknown loc'n, amplitude, and width. True values:

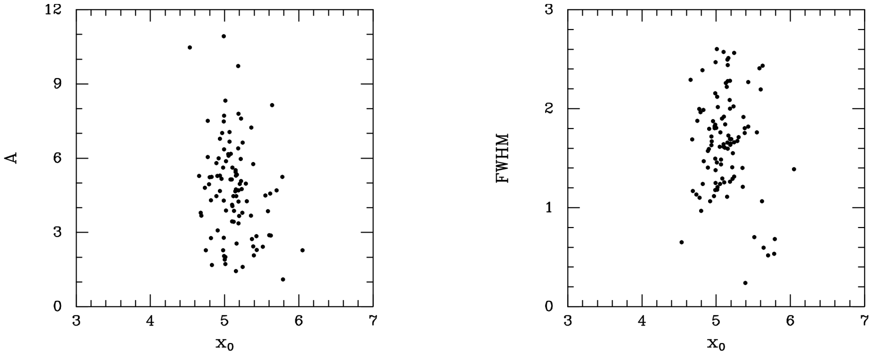
$$x_0 = 5.2, \qquad \text{FWHM} = 0.6, \qquad A = 7$$

Initial scan with crude ($\sigma = 1$) instrument provides 11 equispaced observations over $[0, 20]$. Subsequent observations will use a better ($\sigma = 1/3$) instrument.

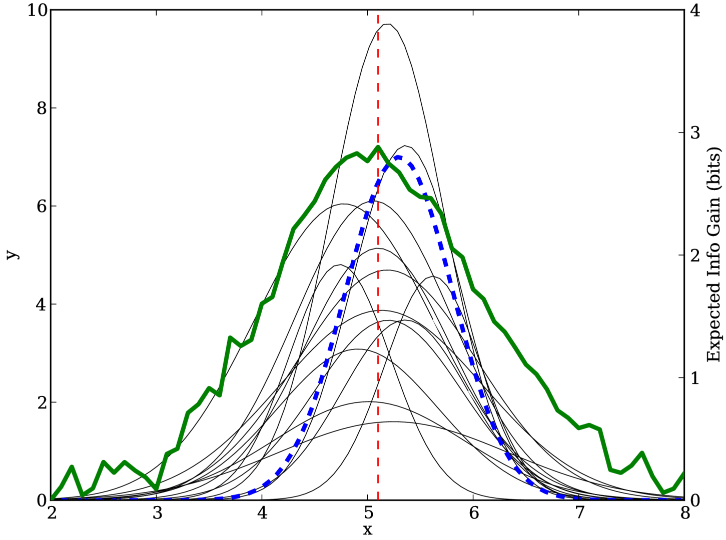


Cycle 1 Interim Inferences

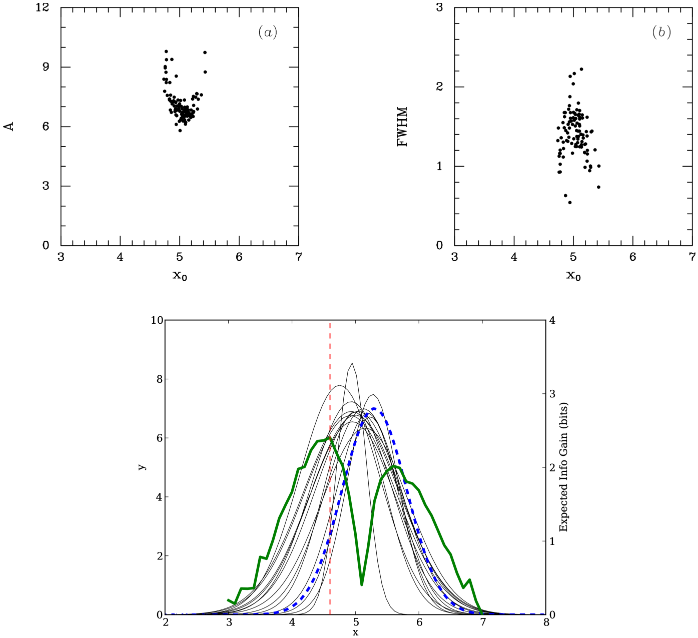
Generate $\{x_0, FWHM, A\}$ via posterior sampling.



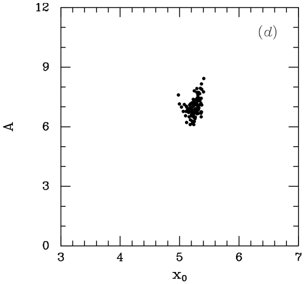
Cycle 1 Design: Predictions, Entropy

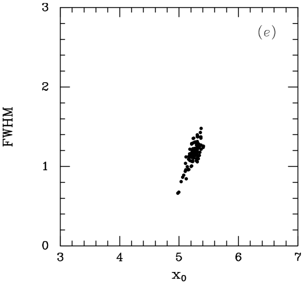


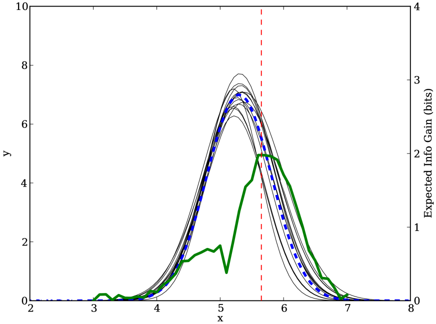
Cycle 2: Inference, Design



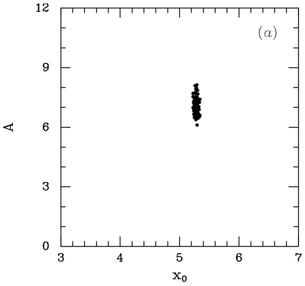
Cycle 3: Inference, Design

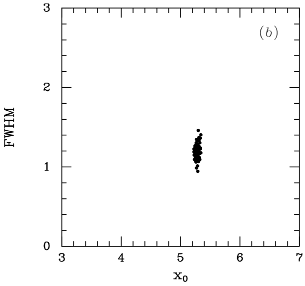




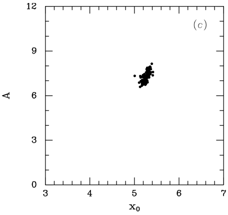


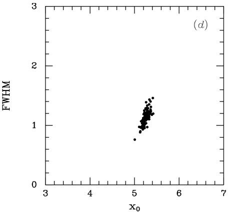
Cycle 4: Inferences





Inferences from *non-optimal* datum

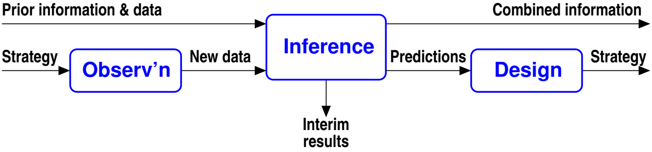




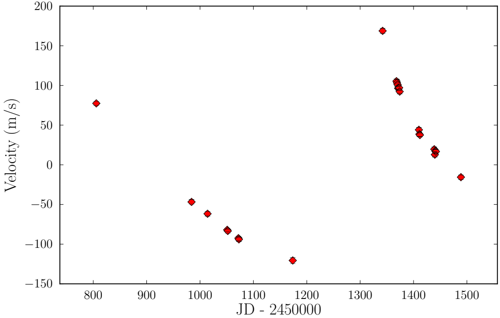
Agenda

1. Motivation: Exoplanets via Doppler RV observations
2. Bayesian adaptive exploration
3. Toy problem: Bump hunting
4. **BAE for HD 222582**

BAE for HD 222582: Cycle 1



HD 222582: G5V at 42 pc in Aquarius, $V = 7.7$
Vogt^{+} (2000) reported planet discovery based on 24 RV measurements



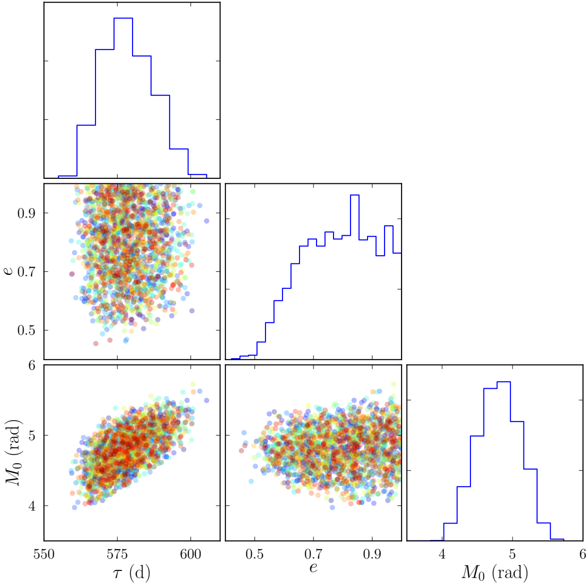
Keplerian Radial Velocity Model

Parameters for single planet

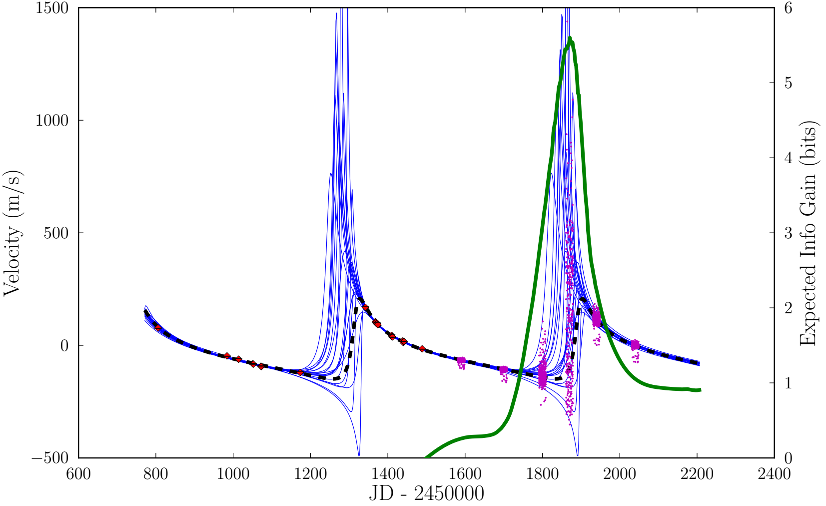
- τ = orbital period (days)
- e = orbital eccentricity
- K = velocity amplitude (m/s)
- Argument of pericenter ω
- Mean anomaly at $t = 0$, M_0
- Systemic velocity v_0

Requires solving Kepler's equation for every (τ, e, M_0)—A strongly nonlinear model!

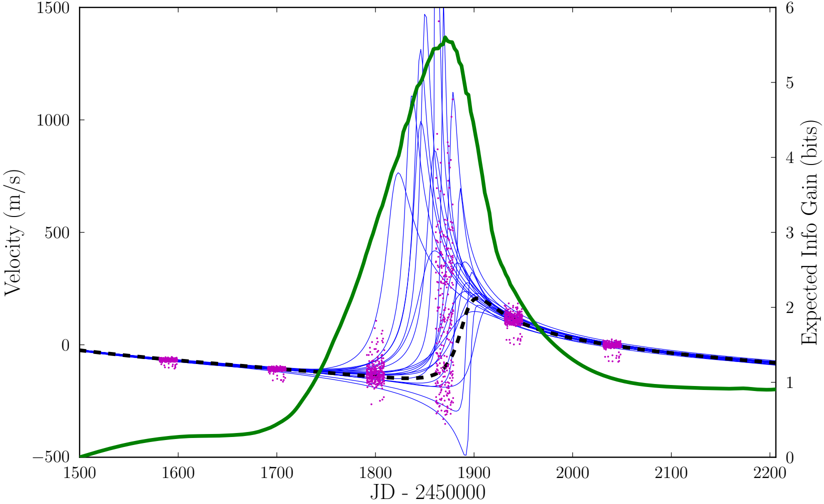
Cycle 1 Interim inferences



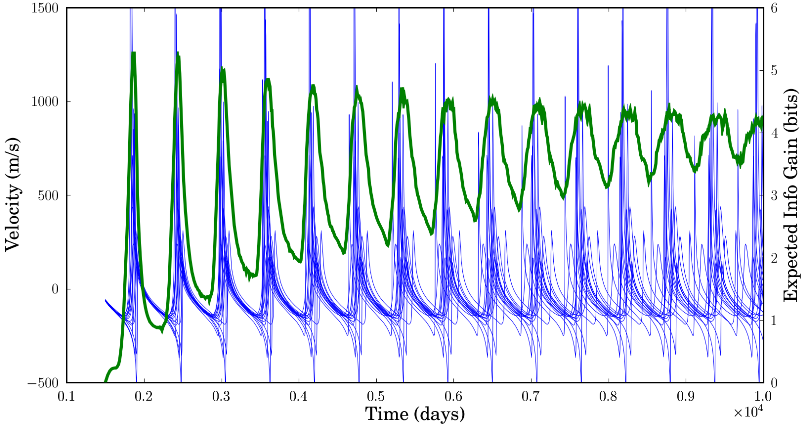
Cycle 1 Design



The next period

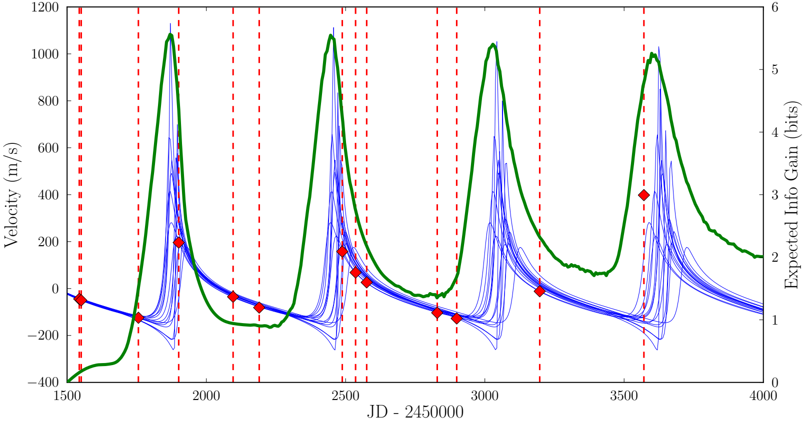


The distant future



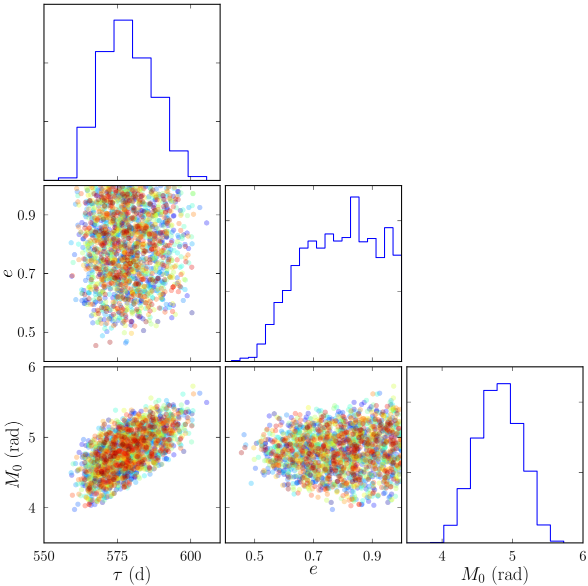
New Data

Red points = 13 subsequent observations, Butler^{+}(2006)

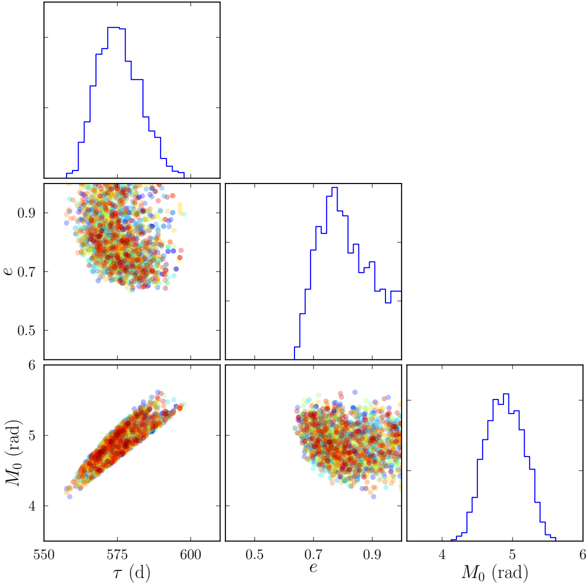


- Use 37-point best fit to simulate three new optimal observations
- Compare $24 + 3$ & all-data inferences

Cycle 1 Interim inferences (24 pts)

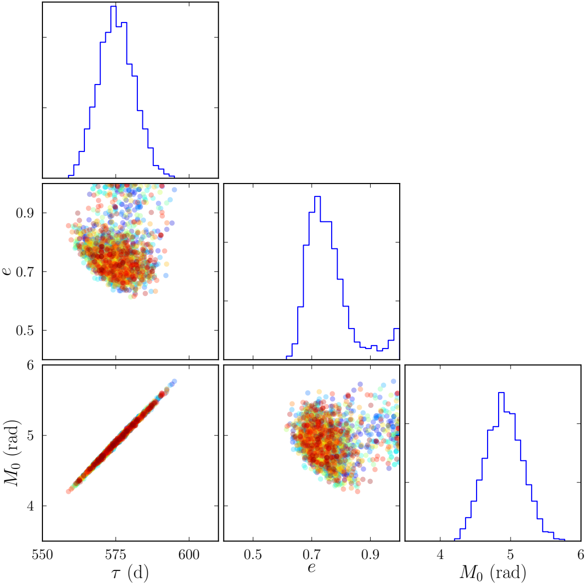


Cycle 2 Interim inferences (25 pts)



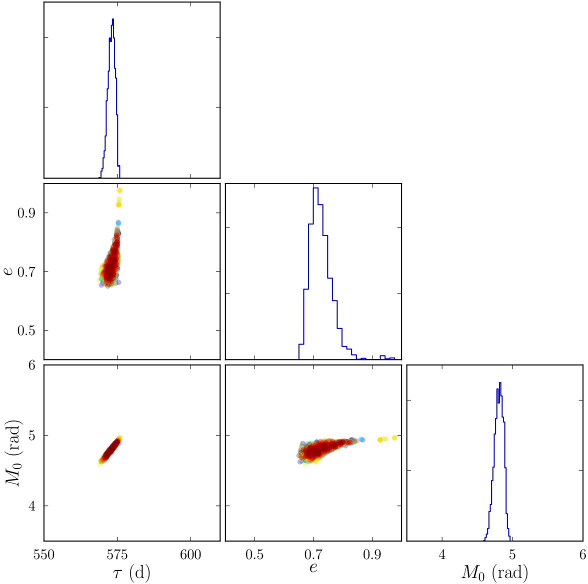
$\prod \sigma_i$ is reduced 2.4x

Cycle 3 Interim inferences (26 pts)



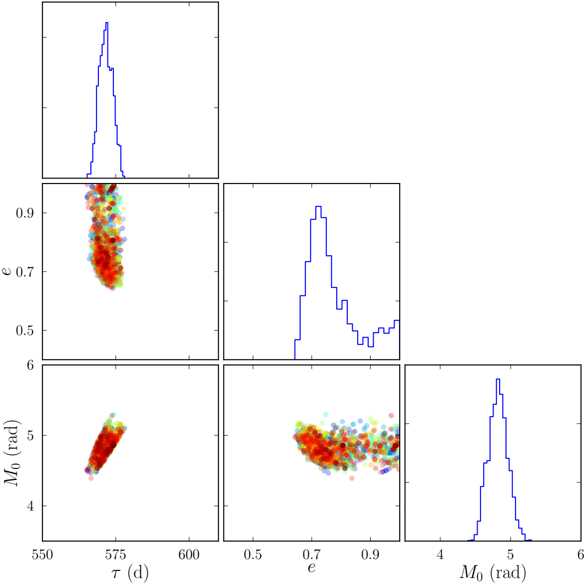
$\prod \sigma_i$ is reduced further 1.5x

Cycle 4 Interim inferences (27 pts)



$\prod \sigma_i$ is reduced further 30x

All-data inferences (37 pts)



$\prod \sigma_i$ is 7x larger than 24 + 3 BAE pts

Outlook

- Explore more cases, e.g., multiple planets, marginal detections
- Explore other adaptive MCMC algorithms
- Extend to include planet *detection*:
  - Total entropy criterion smoothly moves between detection & estimation
  - MaxEnt sampling no longer valid
  - Marginal likelihood computation needed
  - Non-greedy designs likely needed

Thanks to my collaborators!

Cornell Astronomy

David Chernoff

Duke Statistical Sciences

Merlise Clyde, Jim Berger, Bin Liu, Jim Crooks

Finally, a word from our sponsor (NASA!)

Google

nasa astrophysics program review | Search

About 723,000 results (0.24 seconds) | Advanced search

Astrophysics Research Program Review

The Astrophysics Division of NASA's Science Mission Directorate will conduct a review of its programs in Research, Analysis and Enabling Technology. This review will assess the effectiveness of these programs in maximizing the scientific productivity from NASA's current and future missions, in the context of the Science Mission Directorate's Science Plan and the Astro2010 Decadal Survey → . Read the charter for this review.

The Astrophysics Division has appointed a review panel, which held its first meeting in mid-December 2010. A Splinter Meeting was held at the American Astronomical Society's January 2011 meeting in Seattle for public comments from the astronomical community: read the presentation (PDF). Comments are also welcome electronically, via the website at http://astroresearchreview.nsstc.nasa.gov/portal/ → from 18 January 2011. The review committee should report its findings in the summer of 2011.

Linda Sparke
Astrophysics Research Program Manager, NASA HQ

Jon Morse
Director, Astrophysics Division, NASA HQ

Final Provocation

Much data analysis is *sequential*:

- Sequential experimentation/exploration
- Chains of discovery (individual objects/events $\rightarrow$ population)

Herman Chernoff on sequential analysis (1996):

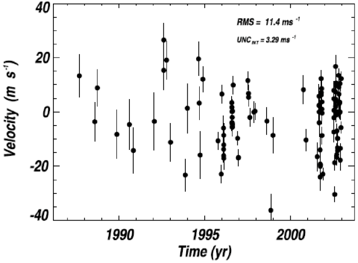
> *I became interested in the notion of experimental design in a much broader context, namely: what's the nature of scientic inference and how do people do science? The thought was not all that unique that it is a sequential procedure...*
>
> *Although I regard myself as non-Bayesian, I feel in sequential problems it is rather dangerous to play around with non-Bayesian procedures.... Optimality is, of course, implicit in the Bayesian approach.*

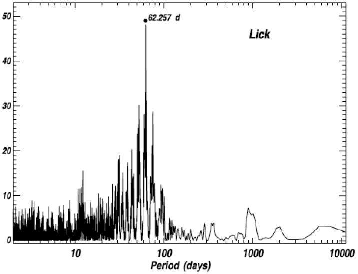
Jetsam

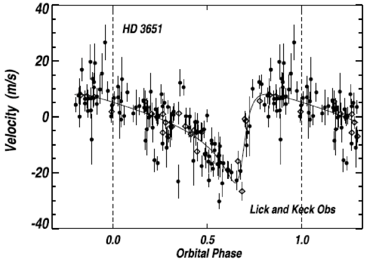
jetsam: material that has been thrown overboard from a ship, esp. material discarded to lighten the vessel

Conventional RV Orbit Fitting

Analysis method: Identify best candidate period via periodogram; fit parameters with nonlinear least squares/min χ^2







System: HD 3651

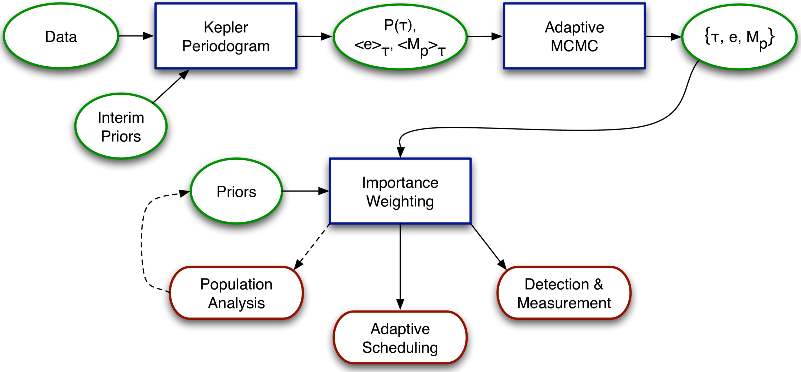
P = 62.23 d
e = 0.63
m sin i = 0.20 M_J
a = 0.28 AU

Fischer et al. 2003

Challenges for Conventional Approaches

- Multimodality, nonlinearity, nonregularity, sparse data $\rightarrow$ Asymptotic uncertainties not valid
- Reporting uncertainties in derived parameters ($m \sin i$, a) and predictions
- Lomb-Scargle periodogram not optimal for eccentric orbits or multiple planets
- Accounting for marginal detections
- Combining info from many systems for pop'n studies
- Scheduling future observations

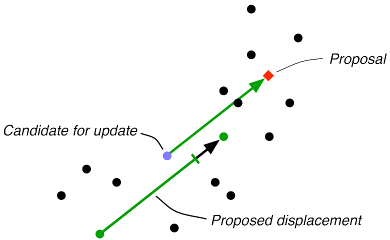
Periodogram-Based Bayesian Pipeline



Differential Evolution MCMC

Ter Braak 2006 — Combine evolutionary computing & MCMC

Follow a population of states, where a randomly selected state is considered for updating via the (scaled) vector difference between two other states.



Behaves roughly like RWM, but with a proposal distribution that automatically adjusts to shape & scale of posterior

Step scale: Optimal $\gamma \approx 2.38/\sqrt{2d}$, but occassionally switch to $\gamma = 1$ for mode-swapping

Expected Information via Nested Monte Carlo

Assume we have posterior samples $\theta_i \sim p(\theta|D, M)$

Evaluating predictive dist'n:

$$\begin{aligned} p(d_e|D, M) &= \int d\theta \; p(\theta|D, M)\, p(d_e|\theta, M) \\ \rightarrow \hat{p}(d_e) &= \frac{1}{N_\theta} \sum_{i=1}^{N_\theta} p(d_e|\theta_i, M) \end{aligned}$$

Sampling predictive dist'n:

$$\begin{aligned} &\theta_i \sim p(\theta|D, M) \\ &d_{e,j} \sim p(d_e|\theta, M) \end{aligned}$$

Entropy of predictive dist'n:

$$\begin{aligned} \mathcal{S}[d_e|D, M] &= -\int dd_e \; p(d_e|D, M_1) \log p(d_e|D, M) \\ &\approx -\frac{1}{N_d} \sum_{j=1}^{N_d} \log \hat{p}(d_{e,j}) \end{aligned}$$