Locations of Kepler Planet Candidates

- Earth-size
- Super-Earth size
  1.25 - 2.0 Earth-size
- Neptune-size
  2.0 - 6.0 Earth-size
- Giant-planet size
  6.0 - 22 Earth-size
Scientific method: hypothetico-deductive approach

- Form hypothesis (based on theory/past experiment)
- Devise experiment to test predictions of hypothesis
- Perform experiment
- Analysis →
  - Devise new hypothesis if hypothesis fails
  - Devise new experiment if hypothesis corroborated
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 scientific 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.*
Bayesian Adaptive Exploration

Bayesian inference + Bayesian decision theory + Information theory

(Plus some computational algorithms...)
Optimal Scheduling of Exoplanet Observations via Bayesian Adaptive Exploration

Tom Loredo
Dept. of Astronomy, Cornell University

Based on work with David Chernoff, Merlise Clyde, Jim Berger & Bin Liu

Supported by the NSF MSPA-Astronomy program

CfA — 15 Feb 2012
Agenda

1. Decision theory & experimental design
2. BAE: Information-maximizing seq’l design
3. Toy problem: Bump hunting
4. BAE for exoplanet RV observations
5. Jetsam
Agenda

1. Decision theory & experimental design
2. BAE: Information-maximizing seq’l design
3. Toy problem: Bump hunting
4. BAE for exoplanet RV observations
5. Jetsam
A Bayesian analysis results in probabilities for two hypotheses:

\[ p(H_1|I) = \frac{5}{6}; \quad p(H_2|I) = \frac{1}{6} \]

Equivalently, the odds favoring \( H_1 \) over \( H_2 \) are

\[ O_{12} = 5 \]

We must base future actions on either \( H_1 \) or \( H_2 \).

Which should we choose?

Naive decision maker: *Choose the most probable, \( H_1 \).*
Naive Decision Making—Deadly!

Russian Roulette

\( H_1 = \text{Chamber is empty}; \quad H_2 = \text{Bullet in chamber} \)

What is your choice now?

Decisions should depend on consequences!

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.

⇒ 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.

\[ \text{Utility} = U(a, o) \]
\[ a = \text{Choice of action (decide b/t these)} \]
\[ o = \text{Outcome (what we are uncertain of)} \]

\[ \text{Loss} \ L(a, o) = U_{\text{max}} - U(a, o) \]
Russian Roulette Utility

<table>
<thead>
<tr>
<th>Actions</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Empty (click)</td>
</tr>
<tr>
<td>Play</td>
<td>$6,000</td>
</tr>
<tr>
<td>Pass</td>
<td>0</td>
</tr>
</tbody>
</table>
Uncertainty & expected utility

We are uncertain of what the outcome will be → 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
### Russian Roulette Expected Utility

<table>
<thead>
<tr>
<th>Actions</th>
<th>Empty <em>(click)</em></th>
<th>Bullet <em>(BANG!)</em></th>
<th>$\mathbb{E}U$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Play</strong></td>
<td>$6,000$</td>
<td>$-$Life$</td>
<td>$5000-$Life/6$</td>
</tr>
<tr>
<td><strong>Pass</strong></td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
</tr>
</tbody>
</table>

As long as $\$Life > \$30,000, *don’t play!*
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
Agenda

1. Decision theory & experimental design
2. BAE: Information-maximizing seq’l design
3. Toy problem: Bump hunting
4. BAE for exoplanet RV observations
5. Jetsam
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 Hi 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

\[ S = \sum_i p_i \log \frac{1}{p_i} \]

“Average surprisal”

\[ = - \sum_i p_i \log p_i \]

Information gain

Existing data \( D \rightarrow \) interim posterior \( p(H_i|D) \)

Information gain upon learning \( d = \) decrease in uncertainty:

\[ \mathcal{I}(d) = S[\{p(H_i|D)\}] - S[\{p(H_i|d, D)\}] \]

\[ = \sum_i p(H_i|d, D) \log p(H_i|d, D) - \text{Const (wrt } d) \]

Lindley (1956, 1972) and Bernardo (1979) advocated using \( \mathcal{I}(d) \) as utility
As an argument of a functional, let $H_i|d, I$ stand for the whole distribution $\{p(H_i|d, I)\}$.

Use the Skilling conditional:

$$\mathcal{I}[H_i|d, I] = \sum_i p(H_i|d, I) \log p(H_i|d, I)$$

$$\rightarrow \mathcal{I}[H_i] = \sum_i p(H_i) \log p(H_i) \quad \| d, I$$
Continuous spaces (e.g., parameter space, $\theta$) need a measure:

- Proper treatment as a limit
- Parameterization invariance
- Makes argument of $\log(\cdot)$ dimensionless

$$I[\theta] = \int d\theta \, p(\theta) \log \frac{p(\theta)}{m(\theta)}$$

For simplicity, we adopt a uniform measure and drop $m(\cdot)$ below (changing it doesn’t affect results).

Aside: Measuring information gain via Kullback-Leibler divergence between prior & posterior does not change results (MacKay 1992).
A ‘Bit’ About Entropy

Entropy of a Gaussian

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

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

→ 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:

$$p(d_t|D) = \sum_i p(d_t, H_i|D) = \sum_i p(H_i|D) p(d_t|H_i)$$

$$= \sum_i \text{Interim posterior} \times \text{Single-datum likelihood}$$
Computational challenge!

**Expected Information**

\[
\mathbb{E} \mathcal{I}(e) = \sum_{d_e} p(d_e|l) \mathcal{I}[H_i|d_e, l]
\]

\[
= \sum_{d_e} \sum_{H_i} p(H_i|l) p(d_e|H_i, l) \times \sum_{H_i'} p(H_i'|d_e, l) \log \left[ p(H_i'|d_e, l) \right]
\]

*There is a heck of a lot of averaging going on!*  
*Plus an optimization!*
Simplification: Maximum entropy sampling

Parameter estimation setting

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

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

Maximum entropy sampling
(Sebastiani & Wynn 1997, 2000)

To learn the most, sample where you know the least
**Nested Monte Carlo integration for **\( \mathcal{E}_I \)

Entropy of predictive dist’n:

\[
S[d_t|D, M] = - \int d d_t \ p(d_t|D, M_1) \log p(d_t|D, M)
\]

- **Sample** predictive via \( \theta \sim \text{posterior}, \ d_t \sim \text{sampling dist’n given } \theta \)
- **Evaluate** predictive as \( \theta \)-mixture of sampling dist’n’s

**Posterior sampling in parameter space**

- Many models are (linearly) **separable** \( \rightarrow \) 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
Bayesian Adaptive Exploration

Greedy information-maximizing sequential design

- 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. Decision theory & experimental design
2. BAE: Information-maximizing seq’l design
3. Toy problem: Bump hunting
4. BAE for exoplanet RV observations
5. Jetsam
Locating a bump

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

\[ x_0 = 5.2, \quad \text{FWHM} = 0.6, \quad A = 7 \]

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

\( A \)

\( x_0 \)

\( \text{FWHM} \)

\( x_0 \)

\( y \)

Expected Info Gain (bits)

\( x \)
Cycle 3: Inference, Design
Cycle 4: Inferences

Inferences from non-optimal datum
Agenda

1. Decision theory & experimental design
2. BAE: Information-maximizing seq’l design
3. Toy problem: Bump hunting
4. BAE for exoplanet RV observations
5. Jetsam
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

Doppler Shift due to Stellar Wobble

\[
\approx 490 \text{ of } \approx 530 \text{ currently confirmed exoplanets found using RV method}
\]

RV method is used to confirm & measure transiting exoplanet candidates
RV Data Via Precision Spectroscopy

Millipixel spectroscopy

Meter-per-second velocities

HD 3651

RMS = 11.4 ms⁻¹

UNCₓᵧ = 3.29 ms⁻¹

Fischer et al. 2003
Keplerian Radial Velocity Model

Parameters for single planet

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

Requires solving Kepler’s equation for every ($\tau, e, M_0$)—A strongly nonlinear model!
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*
BD 222582: G5V at 42 pc in Aquarius, \( V = 7.7 \)

Vogt\(^+\) (2000) reported planet discovery based on 24 RV measurements
Cycle 1 Interim inferences
The next period

![Graph showing velocity and expected info gain over time.](image-url)
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)

\[ \prod \sigma_i \text{ is reduced } 2.4x \]
Cycle 3 Interim inferences (26 pts)

\[ \prod \sigma_i \] is reduced further 1.5x
Cycle 4 Interim inferences (27 pts)

\[ \prod \sigma_i \text{ is reduced further } 30x \]
All-data inferences (37 pts)

\[ \prod \sigma_i \text{ is 7x larger than } 24 + 3 \text{ BAE pts} \]
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
Agenda

1. Decision theory & experimental design
2. BAE: Information-maximizing seq’l design
3. Toy problem: Bump hunting
4. BAE for exoplanet RV observations
5. Jetsam
jetsam: material that has been thrown overboard from a ship, esp. material discarded to lighten the vessel
Parameters for an Orbit — Single Planet

Size & shape: semimajor axis $a$, eccentricity $e$
Orientation: 3 Euler angles, $i$, $\omega$, $\Omega$
Time evolution: period $\tau$, origin $M_0$
Center-of-mass position & velocity

RV parameters: semi-amplitude $K(a, e, \tau)$, $\tau$, $e$, $M_0$, $\omega$, COM velocity $v_0$

Ultimate goal: multiple planets, astrometry $\rightarrow$ dozens of parameters!
Keplerian Radial Velocity Model

Parameters for single planet
- $\tau = \text{orbital period (days)}$
- $e = \text{orbital eccentricity}$
- $K = \text{velocity amplitude (m/s)}$
- Argument of pericenter $\omega$
- Mean anomaly at $t = 0$, $M_0$
- Systemic velocity $v_0$

Keplerian reflex velocity vs. time

$$v(t) = v_0 + K \left( e \cos \omega + \cos[\omega + \nu(t)] \right)$$

True anomaly $\nu(t)$ found via Kepler’s equation for eccentric anomaly:

$$E(t) - e \sin E(t) = \frac{2\pi t}{\tau} - M_0; \quad \tan \frac{\nu}{2} = \left( \frac{1 + e}{1 - e} \right)^{1/2} \tan \frac{E}{2}$$

A strongly nonlinear model!
The Likelihood Function

Keplerian velocity model with parameters $\theta = \{K, \tau, e, M_0, \omega, v_0\}$:

$$d_i = \nu(t_i; \theta) + \epsilon_i$$

For measurement errors with std dev'n $\sigma_i$, and additional “jitter” with std dev'n $\sigma_J$,

$$L(\theta, \sigma_J) \equiv p(\{d_i\} | \theta, \sigma_J)$$

$$= \frac{1}{\sqrt{2\pi(\sigma_i^2 + \sigma_J^2)}} \exp \left( -\frac{1}{2} \frac{(d_i - \nu(t_i; \theta))^2}{\sigma_i^2 + \sigma_J^2} \right)$$

$$\propto \prod_i \frac{1}{2\pi \sqrt{\sigma_i^2 + \sigma_J^2}} \exp \left( -\frac{1}{2} \chi^2(\theta) \right)$$

where $\chi^2(\theta, \sigma_J) \equiv \sum_i \frac{(d_i - \nu(t_i; \theta))^2}{\sigma_i^2 + \sigma_J^2}$

Ignore jitter for now . . .
Know Thine Enemy: Likelihood Slices

\[ d_i = \nu(t_i; \theta) + \epsilon_i \quad \Rightarrow \quad \mathcal{L}(\theta) \propto \exp \left[ -\frac{1}{2} \chi^2(\theta) \right] \quad \text{(include jitter)} \]

Bayesian calculations must \textit{integrate over} \( \theta \).
Conventional RV Orbit Fitting

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

System: HD 3651

$P = 62.23 \text{ d}$
$e = 0.63$
$m \sin i = 0.20 \text{ M}_J$
$a = 0.28 \text{ AU}$

Fischer et al. 2003
Challenges for Conventional Approaches

- Multimodality, nonlinearity, nonregularity, sparse data → 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
Computational Tasks

**Posterior sampling**

Draw \( \{\theta_i\} \) from

\[
p(\theta|D, M_p) = \frac{\pi(\theta|M_p) L(\theta)}{Z} \equiv \frac{q(\theta)}{Z}
\]

An “oracle” is available for \( q(\theta) \); \( Z \) is not initially known. Use samples to approximate \( \int d\theta \ p(\theta|D, M_p) f(\theta) \).

**Model (marginal) likelihood computation**

\[
L(M_p) \equiv p(D|M_p) = Z = \int d\theta \ q(\theta)
\]

**Information functional computation**

\[
\mathcal{I}[H_j] = \sum_j p(H_j) \log p(H_i) \quad \text{(over } \theta \text{ or } M_p)\]
Two New Directions

Bayesian periodograms + population-based MCMC

- Use periodograms to:
  - Reduce dimensionality (requires interim priors)
  - Create an initial population of candidate orbits
- Evolve the candidate population using interactive chains

Annealing adaptive importance sampling (SAIS)

- Abandon MCMC!
- Use sequential Monte Carlo to build importance sampler from $q(\theta)$
- Gives posterior samples and marginal likelihood
- Blind start (currently . . . )
Periodogram-Based Bayesian Pipeline

Data → Kepler Periodogram → $P(\tau)$, $<e>_\tau$, $<M_\rho>_\tau$ → Adaptive MCMC → $\{\tau, e, M_\rho\}$

Interim Priors

Priors → Importance Weighting

Population Analysis → Adaptive Scheduling

Detection & Measurement
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 $\gamma \approx \frac{2.38}{\sqrt{2d}}$, but occasionally switch to $\gamma = 1$ for mode-swapping
Differential Evolution for Exoplanets

Use Kepler & harmonic periodogram results to define initial population for DEMC.

Augment final \( \{\tau, e, M_0\} \) with associated \( \{K, \omega, v_0\} \) samples from their exact conditional MVN distribution.

Advantages:

- Only 2 tuning parameters (\# of parallel chains; mode swapping)
- Good initial sample \( \rightarrow \) fast “burn-in”
- Updates all parameters at once
- Candidate distribution adapts its shape and size
- All of the parallel chains are usable
- Simple!
Results for HD 222582

24 Keck RV observations spanning 683 days; long period; hi e

Reaches convergence dramatically faster than PT or RWM

Conspiracy of three factors: Reduced dimensionality, adaptive proposals, good starting population (from K-gram)
Expected Information via Nested Monte Carlo

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

Evaluating predictive dist’n:

$$p(d_e|D, M) = \int d\theta \ p(\theta|D, M) \ p(d_e|\theta, M)$$

$$\rightarrow \hat{p}(d_e) = \frac{1}{N_\theta} \sum_{i=1}^{N_\theta} p(d_e|\theta_i, M)$$

Sampling predictive dist’n:

$$\theta_i \sim p(\theta|D, M)$$

$$d_{e,j} \sim p(d_e|\theta, M)$$

Entropy of predictive dist’n:

$$S[d_e|D, M] = - \int dd_e \ p(d_e|D, M_1) \log p(d_e|D, M)$$

$$\approx - \frac{1}{N_d} \sum_{j=1}^{N_d} \log \hat{p}(d_{e,j})$$
Importance sampling

\[
\int d\theta \, \phi(\theta) q(\theta) = \int d\theta \, \phi(\theta) \frac{q(\theta)}{P(\theta)} P(\theta) \approx \frac{1}{N} \sum_{\theta_i \sim P(\theta)} \phi(\theta_i) \frac{q(\theta_i)}{P(\theta_i)}
\]

Choose \( Q \) to make variance small. (Not easy!)

Can be useful for both model comparison (marginal likelihood calculation), and parameter estimation.
Building a Good Importance Sampler

Estimate an annealing target density, $\pi_n$, using a mixture of multivariate Student-t distributions, $q_n$:

$$q_n(\theta) = [q_0(\theta)]^{1-\lambda_n} \times [q(\theta)]^{\lambda_n}, \quad \lambda_n = 0 \ldots 1$$

$$P_n(\theta) = \sum_j \text{MVT}(\theta; \mu^n_j, S^n_j, \nu)$$

Adapt the mixture to the target using ideas from sequential Monte Carlo.

**Initialization**
Sample, weight, refine

Sample & calculate weights

Refine IS: EM + Birth/Death

Overall algorithm

Target $q_0$ → Anneal $q_1$ → AAIS Step $q_2$

Design $P_0$ → Sample $\{\theta_i, w_i\}$ → Adapt $P_1$ → Sample $P_1$ → Anneal $q_2$
2-D Example:
Many well-separated correlated normals

\[ \lambda_1 = 0.01 \]

\[ \lambda_3 = 0.11 \]

\[ \lambda_8 = 1 \]

d.o.f. = 5; weights vary

samples from \( q_1 \)
scales
locations
Observed Data: HD 73526 (2 planets)

Data and RV Curve for 2-Planet Fit
Periods: 188 d, 377 d (weakly resonant)

Bayes factors:
1 vs 0 planet: $6.5 \times 10^6$
2 vs 1 planet(s): $8.2 \times 10^4$

1-D and 2-D Marginals for Orbital Parameters
(longer-period planet)

Sampling efficiency of final mixture $\text{ESS}/N \approx 65\%$
Design for Model Comparison

For comparing $M_1$ to $M_0$ (e.g., signal detection) again consider information as utility, but information in model posterior, $p(M_i|d_e, D, I)$.

The predictive is now a finite mixture:

$$p(d_e|D, I) = p(M_0|D, I)p(d_e|D, M_0) + p(M_1|D, I)p(d_e|D, M_1)$$

The conditional predictive is also a mixture (for parametric models):

$$p(d_e|D, M_i) = \int d\theta_i \ p(\theta_i|D, M_i) \ p(d_e|\theta_i, M_i)$$

Parameter uncertainty → this typically depends on $e$
Three Complications

- **Marginal likelihoods appear**: $p(M_k|D, I) \rightarrow$ Need ML algorithm
- **No MaxEnt sampling**: The conditional predictive is $p(d_e|D, M_k)$; its entropy *does* depend on $M_k$. $\rightarrow$ Utility is computationally expensive
- **Non-greedy design**: Greedy algorithms typically behave poorly for model discrimination (Bayes factors may not change much with just a single new sample). $\rightarrow$ Design space is higher dimensional

$\Rightarrow$ There is limited work in this direction.
Total Entropy Criterion

*Can we automate switching between detection & estimation in a principled way?*

Look at information in joint posterior for \((M_k, \theta_k)\):

\[
p(M_k, \theta_k | D) = p(M_k | D)p(\theta_k | D, M_k) \equiv p_k q_k(\theta_k)
\]

Calculate information:

\[
\mathcal{I}[M_k, \theta_k | D] = \sum_k \int d\theta_k p_k q_k(\theta_k) \log[p_k q_k(\theta_k)]
\]

\[
= \sum_k p_k \log p_k + \sum_k p_k \int d\theta_k q_k(\theta_k) \log q_k(\theta_k)
\]

Balances entropy changes in the model posterior and the parameter posteriors (Borth 1975).