21st-Century Statistical Computation for Exoplanet Studies

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

With David Chernoff, Merlise Clyde, Jim Berger, Floyd Bullard, Jim Crooks

Work very much in progress

MaxEnt — July 9, 2007
A Productive Methodological Union

Bayesian Statistics
Inference, Decision Theory, Information Theory

Multidimensional Integration
Cubature, Monte Carlo, MCMC/Metropolis-Hastings...

Applied Bayesian Statistics
An Unproductive Sociological Divide

Bayes in Statistics
Wald, Ramsey, de Finetti, Savage, Lindley...
(Econometrics, bioinformatics, sociology, psychology...)

Bayes in the Physical Sciences
Laplace..., Jeffreys..., Cox, Jaynes, Gull, Skilling, Bretthorst...

Information Science
An Example

Cosmological Parameters from WMAP & SDSS
Tegmark et al. 2004

- Looked at over 2 dozen models with 2 – 10 params
- Used random walk MCMC in “rotated” params
- Typical run length $3 – 5 \times 10^5$; a few runs trapped
- 30 CPU years of effort
1. Introducing Exoplanets

2. Bayesian Methods for Exoplanets

3. Some 21st Century Algorithms
   A Bayesian pipeline
   Improved MCMC for parameter estimation
   Marginal likelihood calculation
Outline

1. Introducing Exoplanets

2. Bayesian Methods for Exoplanets

3. Some 21st Century Algorithms
   - A Bayesian pipeline
   - Improved MCMC for parameter estimation
   - Marginal likelihood calculation
Extrasolar Planets

Exoplanet detection/measurement methods:

- **Direct**: Transits, gravitational lensing, imaging, interferometric nulling
- **Indirect**: Keplerian reflex motion (line-of-sight velocity, astrometric wobble)

The Sun’s Wobble From 10 pc
Radial Velocity Technique

As of July 2007, 245 planets found, including 26 multiple-planet systems. Vast majority (233) found via Doppler radial velocity (RV) measurements.
Parameters for an Orbit — Single Planet

Intrinsic geometry: semimajor axis $a$, eccentricity $e$
Orientation: 3 Euler angles, $i$, $\omega$, $\Omega$
Time: period $\tau$, origin $M_p$

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

Physics: min. mass $m \sin i$, size $a$, eccentricity $e$

(Ultimate goal: multiple planets, astrometry $\rightarrow \sim 30+$ parameters!)
Conventional RV Data Analysis

Analysis method: Identify best candidate period via periodogram; fit parameters with nonlinear least squares

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

Fisher et al. 2003
Issues with Periodogram + Least Squares

- Multimodality, nonlinearity, sparse data $\rightarrow \Delta \chi^2$
  uncertainties not valid
- Reporting uncertainties in derived parameters ($m \sin i, a$)
- Lomb-Scargle periodogram not optimal for eccentric and multiple planets
- Handling marginal detections
- Combining info from many systems for pop’n studies
- Scheduling future observations

*Bayesian statistics can address all these within a unified framework!*
Outline

1. Introducing Exoplanets

2. Bayesian Methods for Exoplanets

3. Some 21st Century Algorithms
   - A Bayesian pipeline
   - Improved MCMC for parameter estimation
   - Marginal likelihood calculation
Scientific Goals ↔ Bayesian Methodology

- **Improved system inferences** — Detection (marginal likelihoods, Bayes factors), estimation (joint & marginal posteriors)

- **Improved population inferences** — Propagate system uncertainties (including detection uncertainties and selection effects) to population level (hierarchical/multi-level Bayes)

- **Adaptive scheduling** — Plan future observations to most quickly reduce uncertainties (Bayesian experimental design)
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 of pericenter passage $M_p$
- System center-of-mass velocity $v_0$

Velocity vs. time

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

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

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

A strongly nonlinear model!
The Likelihood Function

Keplerian velocity model with parameters \( \theta = \{K, \tau, e, M_p, \omega, v_0\} \):

\[
d_i = v(t_i; \theta) + \epsilon_i
\]

For measurement errors with std dev’n \( \sigma_i \), and additional “jitter” with std dev’n \( \sigma_J \),

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

\[
= \prod_{i=1}^{N} \frac{1}{2\pi \sqrt{\sigma_i^2 + \sigma_J^2}} \exp \left[ -\frac{1}{2} \frac{(d_i - v(t_i; \theta))^2}{\sigma_i^2 + \sigma_J^2} \right]
\]

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

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

Ignore jitter for now . . .
What To Do With It

Parameter estimation

Posterior dist’n for parameters of model $M_i$ with $i$ planets:

$$p(\theta|D, M) \propto p(\theta|M) \mathcal{L}_i(\theta)$$

Summarize with mode, means, credible regions (found by integrating over $\theta$)

Detection

Calculate probability for no planets ($M_0$), one planet ($M_1$) . . . . Let $I = \{M_i\}$.

$$p(M_i|D, I) \propto p(M_i|I) \mathcal{L}(M_i)$$

where $\mathcal{L}(M_i) = \int d\theta p(\theta|M_i) \mathcal{L}(\theta)$

Marginal likelihood $\mathcal{L}(M_i)$ includes “Occam factor”
**Design**

Predict future datum $d_t$ at time $t$, accounting for model uncertainties:

$$p(d_t|D, M_i) = \int d\theta \ p(d_t|\theta, M_i) \ p(\theta|D, M_i).$$

1st factor is Gaussian for $d_t$ with known model; 2nd term & integral account for uncertainty. Information theory $\rightarrow$ best time has largest $d_t$ uncertainty.

**Population analysis**

Multi-level ("hierarchical") model: Parameterize the orbit parameter prior and infer it.
Include probabilities for star having (1, 2, . . . ) planets $\rightarrow$ Bayes factors weight system contributions.
Need *all* data, well-characterized surveys.
Current MCMC Methods

Random Walk Metropolis (Ford)

<table>
<thead>
<tr>
<th>$T_{obs}/P$</th>
<th>$u_1$</th>
<th>$u_2$</th>
<th>$u_3$</th>
<th>$(u_3, u_4, u_5)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon = 0.01$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.252.5</td>
<td>7.8</td>
<td>5.4</td>
<td>7.7</td>
<td>5.2</td>
</tr>
<tr>
<td>$\epsilon = 0.1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.252.5</td>
<td>6.2</td>
<td>4.7</td>
<td>6.1</td>
<td>4.2</td>
</tr>
</tbody>
</table>

Performance depends a lot on parameterization

Parallel Tempering (Gregory)

General purpose, but inefficient/unwieldy
Outline

1. Introducing Exoplanets

2. Bayesian Methods for Exoplanets

3. Some 21st Century Algorithms
   - A Bayesian pipeline
   - Improved MCMC for parameter estimation
   - Marginal likelihood calculation
Improving Exoplanet Bayesian Calculation

Seek more efficient and less unwieldy samplers → abandon “general purpose” algorithms

Keep all tasks in mind: Estimation, detection, design, pop’n . . .

Guidelines for modest dimensional Bayesian computation:

- **Know thine enemy** — Understand scales in various dimensions, multimodality, locate modes; significant exploration should precede & guide sampling

- **Reduce dimensionality** — Analytic/numerical techniques may help

- **Clever priors** — If a clever prior (that doesn’t do too much violence to real prior info) helps, use it; you can weight and resample/Rao-Blackwellize later

- **Try different samplers** — Nearly 20 years of ideas
Know Thine Enemy I: Linear vs. Nonlinear Parameters

\[ v(t) = v_0 + K (e \cos \omega + \cos[\omega + v(t)]) \]
\[ = A_1 + A_2[e + \cos v(t)] + A_3 \sin v(t) \]

\[ A_1 = v_0; \quad A_2 = K \cos \omega; \quad A_3 = -K \sin \omega \]

Model is linear wrt \( A_i \), nonlinear wrt \( e, M_p, \omega \)

\[ p(\theta) = p(\tau, e, M_p) \ p(\{A_i\}|\tau, e, M_p) \quad \parallel D, M_1 \]

with \( p(\{A_i\}|\tau, e, M_p) \) conditionally normal if we adopt a flat or conjugate prior for \( \{A_i\} \).

Flat prior \( \rightarrow p(K|M_1) \propto K \). Using this interim prior, dimension can be reduced by 3.
Know Thine Enemy II — Slices

Extreme multimodality in $\tau$; challenging multimodality in $M_p$; smooth in $e$
Periodogram Connections

Bayesian periodograms (Jaynes-Bretthorst algorithm)

Data are superposition of periodic functions + noise:

\[ d_i = \sum_{\alpha=1}^{M} A_\alpha g_\alpha(t_i; \omega; \theta) + e_i \]

Calculate \( \mathcal{L}(\{A\}, \omega, \theta) \) using \( \chi^2 \).
Integrate out \( A \)'s \( \rightarrow \) least squares + volume factors:

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

\( r^2(\omega, \theta) \) is squared residual, found by diagonalizing metric

\[
\eta_{\alpha\beta} = g_\alpha \cdot g_\beta
\]

Integrate out \( \theta \) numerically \( \rightarrow p(\omega|D) \);

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

Generalizes Schuster periodogram & LSP.
Circular Orbits

Assume circular orbits: $\theta = \{K, \tau, \phi, v_0\}$

Frequentist

For given $\tau$, maximize likelihood over $K$ and $\phi$ (set $v_0$ to data mean, $\bar{v}$) $\rightarrow$ profile likelihood:

$$\log L_p(\tau, \bar{v}) \propto \text{Lomb-Scargle periodogram}$$

Bayesian

For given $\tau$, integrate ("marginalize") likelihood $\times$ prior over $K$ and $\phi$ (set $v_0$ to data mean, $\bar{v}$) $\rightarrow$ marginal posterior:

$$\log p(\tau, \bar{v}|D) \propto \text{Lomb-Scargle periodogram}$$

Additionally marginalize over $v_0 \rightarrow$ floating-mean LSP
Radial Kepler Periodogram for Eccentric Orbits

\[ \nu(t) = A_1 + A_2[e + \cos \nu(t)] + A_3 \sin \nu(t) \]

For given \((\tau, e, M_p)\), analytically marginalize \(\{A_i\}\):

\[
\ln p(\tau, e, M_p|D, M_p) = (\text{Radial}) \text{ Kepler-gram}
\]

For given \(\tau\), marginalize over \((e, M_p)\):

\[
\log p(\tau|D) \propto (\text{Radial}) \text{ Kepler periodogram}
\]

This requires a 2-d numerical integral.
Terminology for Generalized Periodograms

**Sinusoid model**

\[ A_1 \cos \omega t + A_2 \sin \omega t \]

\[ \ln p(\omega|D, M) \propto \text{Schuster periodogram} \]

**Chirp model**

\[ A_1 \cos(\omega t + \alpha t^2) + A_2 \sin(\omega t + \alpha t^2) \]

\[ \ln p(\omega, \alpha|D, M) \propto \text{Chirpogram (Jaynes)} \]

**Keplerian reflex motion model**

\[ \ln p(\tau, e, M_p|D, M_p) = \text{(Radial) Kepler-gram} \]

\[ \log p(\tau|D) \propto \text{(Radial) Kepler periodogram} \]
A Bayesian Pipeline

- Choose priors to enable analytic calculations to improve MCMC performance; fix priors with importance weighting
- Use modern, adaptive MCMC algorithms
A Bayesian Pipeline

- Choose priors to enable analytic calculations to improve MCMC performance; fix priors with importance weighting
- Use modern, adaptive MCMC algorithms
Some Modern MCMC Themes

- Adaptive MCMC: Proposal dist’ns that change scale/shape
- Coupled chains: Parallel tempering, parallel marginalization
- Population-based MCMC: Coupled chains, all sampling the posterior; pop’n defines the proposal
- Mixing samplers: Local exploration + mode-hopping

(Note: Skilling’s BayeSys3 has some of this; large statistics literature)
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 occasionally switch to $\gamma = 1$ for mode-swapping
Differential Evolution for Exoplanets

Use Kepler periodogram to reduce dimensionality to 3-d ($\tau$, $e$, $M_p$).

Use Kepler periodogram results ($\rho(\tau)$, moments of $e$, $M_p$) to define initial population for DEMC:
- Marsaglia’s 5-table sampler for $\tau$
- Conditional beta and Von Mises samplers for $e$, $M_p$

Once we have $\{\tau, e, M_p\}$, get associated $\{K, \omega, v_0\}$ samples from their exact conditional distribution.

Advantages vs. PT & RWM:
- Only 2 tuning parameters (# of parallel chains; mode swapping)
- 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; high e

Kepler Periodogram

![Kepler Periodogram](image-url)
Differential Evolution MCMC Performance

Reaches convergence dramatically faster than PT or RWM, with only one tunable parameter (pop’n size — unexplored here!)

Conspiracy of three factors: Reduced dimensionality, adaptive proposals, good starting population (from K-gram)
Results for HD 73526 (Early Data)

Exoplanet system with 3 important modes in \( \tau \)

DEMC can quickly explore \textit{within} modes, but swaps between modes \( \rightarrow \) longer convergence. Partitioning allows fast exploration within modes, can be automated.

Does seem to be swapping at “right” frequency; the two minor modes are well-separated (though weak).

What should our goal be for multimodal sampling? Should we weight small but not-insignificant modes to increase their sampling frequency?
Exoplanet MCMC Work-In-Progress

- Normal kernel coupler (Warnes 2000)
- Stochastic approximation Monte Carlo (SAMC) — Better explore modes
- Derivatives of the Jaynes-Bretthorst algorithm:
  - May accelerate MCMC via hybrid Monte Carlo
  - Score-based output analysis (Fan, Brooks & Gelman 2006)
- Gradient/geodesic mode-hopping
Marginal Likelihood Calculation

**Unpromising methods**

- Harmonic mean
- Weighted harmonic mean
- Thermodynamic integration
- Nested sampling

**Promising methods**

- Restricted importance sampler
- Regional mixture importance sampler
- Hybrid ratio estimator
- Adaptive kernel density importance sampler
- Adaptive simplex cubature
Adaptive Simplex Cubature
(With a nod to Ken Hanson)

Motivation: Use MCMC sample locations and densities

Suppose you were given \( \{\theta_i, q_i\} \) and told to estimate \( Z = \int d\theta q(\theta) \) for this 1-d \( q \).

Use a quadrature approximation that doesn't require specific abscissas: histogram, trapezoid, etc.. These weight by “volume” factors:

\[
Z = \sum_{\text{intervals}} \text{(length)} \times \text{(avg. height)}
\]

In 2-d intervals are triangles (2-simplices); length→area. Make the triangles via Delaunay triangulation.
Higher dimensions: Combine $n$-d Delaunay triangulation and $n$-d simplex trapezoidal rules of Lyness & Genz
Performance

Explored up to 6-d with a variety of standard test-case normal mixtures, using samples as vertices. Qhull used for triangulation. Triangulation is expensive → use a small number of vertices. In few-d, requires many fewer points that subregion-adaptive cubature (DCUHRE), but underestimates integrals in > 4-D. There is lots of volume in the outer “shell” so even though density is low, it contributes a lot.

Modifications

• Tempered/derivative-weighted resampling (seems to work to 6- or 7-D)
• Non-optimal triangulations
Summary: Guidelines for Modest-Dimension Bayes

- **Know thine enemy** — Understand scales in various dimensions, multimodality, locate modes; *significant exploration should precede & guide sampling*

- **Reduce dimensionality** — Analytic/numerical techniques may help

- **Clever priors** — If a clever prior (that doesn’t do too much violence to real prior info) helps, use it; you can weight and resample/Rao-Blackwellize later

- **Try different samplers** — Nearly 20 years of ideas

*Cross the sociological divide! It’s worth it!*