Bayesian model fitting through Controlled Statistical Fusion with Exoplanet examples

Phil Gregory

Physics and Astronomy
Univ. of British Columbia

See papers for full details at http://www.physics.ubc.ca/~gregory/gregory.html
Abstract

Many different statistical techniques have been developed to facilitate the detection of a global minimum in chi-squared in the multi-modal environment commonly encountered in nonlinear model fitting. This is particularly challenging as the number of model parameters increase necessitating a MCMC approach.

This paper describes progress in the development of a new fusion Markov chain Monte Carlo algorithm which incorporates the advantages of parallel tempering, simulated annealing and the genetic algorithm. This statistical fusion approach has been achieved through the development of a unique multi-stage adaptive control system, hence the term controlled statistical fusion. Among other things the control system automates the tuning of the proposal distributions for efficient exploration of the model parameter space even when the parameters are highly correlated.

The fusion MCMC algorithm is implemented in Mathematica using parallelized code and run on an 8 core PC.
Fusion MCMC

The latest version of FMCMC for nonlinear model fitting incorporates:

- Parallel tempering
- Simulated annealing
- Genetic algorithm

Each of these methods was designed to facilitate the detection of a global minimum in $\chi^2$
  (in a Bayesian context: a global maximum of probability).
By combining all three we greatly increase the probability of realizing this goal.

This fusion has only been possible through the development of a unique adaptive control system. Among other things it automates the choice of an efficient set of MCMC proposal distributions even if the parameters are highly correlated.

Note: FMCMC was previously named hybrid MCMC (HMCMC) but I discovered this term already had a different meaning in the literature.
Adaptive Fusion MCMC

8 parallel tempering Metropolis chains

Output at each iteration

Monitor for parameters with peak probability

Peak parameter set: If \((\text{logprior} + \text{loglike}) > \) previous best by a threshold then update and reset burn-in

β = 1.0
β = 0.72
β = 0.52
β = 0.39
β = 0.29
β = 0.20
β = 0.13
β = 0.09

Parallel tempering swap operations

Anneal Gaussian proposal \(\sigma\)'s
Refine & update Gaussian proposal \(\sigma\)'s

2 stage proposal \(\sigma\) control system
error signal = (actual joint acceptance rate – 0.25)
Effectively defines burn-in interval

Genetic algorithm
Every 10th iteration perform gene crossover operation to breed larger \((\text{logprior} + \text{loglike})\) parameter set.

β = 1/ T
If you input a Kepler model the fusion MCMC becomes a Kepler periodogram. Optimum for finding Kepler orbits and evaluating their probabilities. Capable of simultaneously fitting multiple planet models. A multi-planet Kepler periodogram.
The radial velocity equation, a nonlinear model

\[ f_i = \text{model prediction} = V + K \left( \cos \left( \theta \left( t_i + \chi P \right) + \omega \right) + e \cos \omega \right) \]

\[ V = \text{systematic velocity} \]
\[ K = \text{velocity amplitude} = \frac{2 \pi a \sin i}{P \sqrt{1 - e^2}} ; \]
\[ \left( a = \text{semi-major axis} \right) \]
\[ i = \text{inclination} \]
\[ e = \text{orbital eccentricity} \]
\[ \omega = \text{longitude of periastron} \]
\[ \chi = \text{fraction of orbit prior to data reference time that periastron occurred at} \]
\[ \theta \left( t_i + \chi P \right) = \text{true anomaly} \]
\[ = \text{angle of star in orbit at time} \ t_i \ \text{relative to periastron} \]

\[ \theta_i \text{ and} \ t_i \text{ are related by the conservation of angular momentum equation} \]

\[ \partial_t \theta \left[ t \right] - \frac{2 \pi}{P \left( 1 - e^2 \right)^{3/2}} \left( 1 + e \cos \theta \right)^2 = 0 \]
Highly correlated parameters

Exoplanet example: for low eccentricity orbits the parameters $\omega$ and $\chi$ are not separately well determined. This shows up as a strong correlation between $\omega$ and $\chi$ as shown on the right.

One option re-parameterization

The combination $2\pi\chi + \omega$ is well determined for all eccentricities. Although $2\pi\chi - \omega$ is not well determined for low eccentricities, it is at least orthogonal to $2\pi\chi + \omega$ as shown.

Another option

Algorithm learns about the parameter correlations during the burn-in and generates proposals with these statistical correlations.
Adaptive Fusion MCMC

8 parallel tempering Metropolis chains

$\beta = 1.0$
$\beta = 0.72$
$\beta = 0.52$
$\beta = 0.39$
$\beta = 0.29$
$\beta = 0.20$
$\beta = 0.13$
$\beta = 0.09$

$\beta = 1/T$

Parallel tempering swap operations

Automatic proposal scheme that learns about parameter correlations during burn-in (for each chain)

Proposal I
Independent Gaussian proposal scheme employed 50% of the time

Proposal C
Proposal distribution with built in parameter correlations used 50% of the time

new parameter value
from proposal C

repeat
from proposal I

add every second to buffer

latest 300 values

difference of random pairs

multiply by constant <1

During burn-in control system adjusts constant so acceptance rate from Proposal C = 25%

MCMC adaptive control system
How to deal with highly correlated parameters

Using only independent Gaussian proposals the (‘l’ scheme) the \(\sigma\)‘s need to be very small for any proposal to be accepted and consequently convergence is very slow.

Learn about parameter correlations during burn-in

The accepted ‘l’ proposals will generally cluster along the correlation path so every 2\(^{nd}\) accepted ‘l’ proposal is appended to a correlated sample buffer (separate buffer for each tempering level).

Only the 300 most recent additions to the buffer are retained.

A ‘C’ proposal is generated using the difference between a pair of randomly selected samples drawn from the correlated sample buffer (for that tempering level), after multiplication by a constant.

Value of constant is computed automatically by another control system module which ensures that the ‘C’ proposal acceptance rate is close to 25%.
Left panels show marginal distributions for two orbital parameters $\chi$ and $\omega$. The black trace corresponds to a search in $\chi$ and $\omega$ using only ‘I’ proposals. Red trace corresponds to search with ‘C’ proposals turned on. Green trace, to an ‘I’ search using transformed orthogonal coordinates.

Right panels show a comparison of the MCMC autocorrelation functions.
Noise model

It is useful to incorporate an extra noise parameter, $s$, of unknown magnitude, added in quadrature to the known measurement uncertainties $\sigma_i$.

\[
\text{variance}_i = \sigma_i^2 + s^2
\]

In general, nature is more complicated than our model and known noise terms. Marginalizing $s$ has the desirable effect of treating anything in the data that can’t be explained by the model and known measurement errors as noise, leading to more conservative estimates of the parameters.

In the absence of detailed knowledge of the sampling distribution for the extra noise, pick an independent Gaussian model because for any given finite noise variance it is the distribution with the largest uncertainty as measured by the entropy, i.e., the maximum entropy distribution.

If there is no extra noise then the posterior probability distribution for $s$ will peak at $s = 0$. 
Annealing due to extra noise term, s

The inclusion of an extra noise term of unknown magnitude also gives rise to an annealing operation when the Markov chain is started far from the best-fit values.

If only known observational errors are included, the posterior probability distribution is often very “rough” with many local maxima throughout parameter space.

When $s$ is included, Bayesian Markov chain automatically inflates $s$ to include anything in the data that cannot be accounted for by the model with the current set of parameters and the known measurement errors.

This results in a smoothing out of the posterior surface and allows the Markov chain to explore the parameter space more quickly. The chain begins to decrease the value of $s$ as it settles in near the best-fit parameters. This behavior is similar to simulated annealing, but does not require choosing a cooling scheme.
47 Ursae Majoris is a solar twin ~46 light years away in Ursa Major

History

1) 1996, report of a P =1090 day companion

2) 2002, report of a 2nd companion, P = 2594±90 days

3) 2004-2009, several papers report either no 2nd planet or a 2nd planet with P = 9660 days when ecc. of 2nd planet set = 0.005, value found by Fischer et al., 2002.

   HMCMC confirms 2400 d planet and finds evidence for a 3rd planet of ~10000 d.
47 Ursae Majoris  3 planet FMCMC model fit, Lick telescope data

Arrow indicates starting periods

Evidence for 3 planets
Longest period not well constrained because period is longer than the data span
47 Ursae Majoris  3 planet FMCMC model fit
Lick + Hobby-Eberly + Harlem J Smith telescope data.
Total of 413 velocity measurements
47 Ursae Majoris

Parameter marginal probability distributions for 3 planet HMCMC fit of the combined Lick, HET, and HJS telescope data set.
47 Ursae Majoris Model selection

Odds ratio \( O_{i2} = \frac{p(M_i \mid D, I)}{p(M_2 \mid D, I)} = \frac{p(M_i \mid I)}{p(M_2 \mid I)} \times \frac{p(D \mid M_i, I)}{p(D \mid M_2, I)} = \frac{p(M_i \mid I)}{p(M_2 \mid I)} \times B_{i2} \)

\[
p(D \mid M_i, I) = \int d\bar{X} \ p(\bar{X} \mid M_i, I) \times p(D \mid \bar{X}, M_i, I)
\]

<table>
<thead>
<tr>
<th>Model</th>
<th>Bayes factor ( B_{i2} )</th>
<th>False Alarm Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M_{0s} )</td>
<td>( 1.6 \times 10^{-141} )</td>
<td>8 ( \times 10^{-114} )</td>
</tr>
<tr>
<td>( M_{1s} )</td>
<td>( 2.0 \times 10^{-28} )</td>
<td>2 ( \times 10^{-28} )</td>
</tr>
<tr>
<td>( M_{2s} )</td>
<td>1.0</td>
<td>2 ( \times 10^{-28} )</td>
</tr>
<tr>
<td>( M_{3s} )</td>
<td>( 2.0 \times 10^5 )</td>
<td>5 ( \times 10^{-6} )</td>
</tr>
</tbody>
</table>

False Alarm Probability (FAP): in the context of claiming the detection of a 3 planet model the FAP is the probability that there are actually 2 or less planets.

\[
FAP = \sum_{i=0}^{2} \text{prob. of } i \text{ planets} \quad p(M_i \mid D, I) = \frac{B_{i2}}{\sum_{j=0}^{N_{\text{mod}}} B_{j2}}
\]

\[
FAP = \frac{(B_{02} + B_{12} + B_{22})}{\sum_{j=0}^{3} B_{j2}} = 5 \times 10^{-6}
\]
Gliese 581 a star with two possible habitable zone planets

History
1) 2005 to 2009,
   Planet e is 1.9 Earth mass
   Planet b is 16 Earth mass
   Planet c is 5 Earth mass
   Planet d is 7 Earth mass
Latest paper:

2) 2010,
   Planet f is 7 Earth mass
   Planet g is 3.1 Earth mass
their analysis assumes all circular orbits.

   http://www.physics.ubc.ca/~gregory/gregory.html
   Don’t support claim for planet 581g.
   Find evidence that HIRES uncertainties are much larger than the quoted values,
an extra 1.8 m s⁻¹ added in quadrature.
Gliese 581 three planet fit to HARPS radial velocity data

3 planet Kepler periodogram

Arrow ↑ indicates starting periods

Eccentricity versus period
Gliese 581
5 planet fit to HARPS data

Marginal parameter distributions

5 planet Kepler periodogram

Eccentricity versus period
Gliese 581: The availability of multi-observatory data (HARPS and HIRES) permits a Bayesian analysis to distinguish and quantify any common extra noise (stellar jitter) from other unrecognized measurement uncertainties.

**Final Noise model**

\[
\sigma_{\text{HARPS}} = \sqrt{\sigma_{\text{quoted}}^2 + s^2}
\]

\[
\sigma_{\text{HIRES}} = \sqrt{\sigma_{\text{quoted}}^2 + ds_{\text{HIRES}}^2 + s^2}
\]

The figure shows the resulting marginal probability distributions for \(ds_{\text{HIRES}}\) and \(s\) obtained from a two planet fit to the combined HARPS/HIRES data set. The gray curve in the lower panel is the marginal for \(s\) from fit to HARPS only data.
Gliese 581: Find evidence that HIRES uncertainties are much larger than the quoted values by an extra 1.8 m s\(^{-1}\) added in quadrature.

**Final Noise model**

\[
\sigma_{\text{HARPS}_j} = \sqrt{\sigma_{\text{quoted}_j}^2 + s^2}
\]

\[
\sigma_{\text{HIRES}_j} = \sqrt{\sigma_{\text{quoted}_j}^2 + ds_{\text{HIRES}}^2 + s^2}
\]

**Extra noise parameters from fit to combined HIRES/HARPS data**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>2 planet full data</th>
<th>3 planet full data</th>
<th>4 planet full data</th>
<th>4 planet partial</th>
</tr>
</thead>
<tbody>
<tr>
<td>(ds_{\text{HIRES}})</td>
<td>(1.15^{+1.05}_{-0.35})</td>
<td>(1.55^{+0.58}_{-0.47})</td>
<td>(1.84^{+0.35}_{-0.33})</td>
<td>(1.76^{+0.45}_{-0.39})</td>
</tr>
<tr>
<td>(s)</td>
<td>(2.39^{+0.25}_{-0.11})</td>
<td>(2.0^{+0.2}_{-0.2})</td>
<td>(1.45^{+.17}_{-.17})</td>
<td>(1.39^{+.15}_{-.18})</td>
</tr>
</tbody>
</table>
Gliese 581 Model Selection: Marginal likelihoods computed using Nested Restricted Monte Carlo (NRMC) method

<table>
<thead>
<tr>
<th>Model</th>
<th>Periods (d)</th>
<th>Marginal Likelihood</th>
<th>Bayes factor nominal</th>
<th>False Alarm Probability</th>
<th>s (m s(^{-1}))</th>
<th>RMS residual (m s(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(M_0)</td>
<td></td>
<td>(6.10 \times 10^{-197})</td>
<td>2.0 (\times 10^{-59})</td>
<td></td>
<td></td>
<td>9.8</td>
</tr>
<tr>
<td>(M_1)</td>
<td>(5.37)</td>
<td>((4.221 \pm 0.003) \times 10^{-155})</td>
<td>1.4 (\times 10^{-17})</td>
<td>1.4 (\times 10^{-42})</td>
<td>3.5</td>
<td>3.6</td>
</tr>
<tr>
<td>(M_2)</td>
<td>(5.37, 12.9)</td>
<td>((1.94 \pm 0.01) \times 10^{-145})</td>
<td>6.5 (\times 10^{-8})</td>
<td>2.2 (\times 10^{-10})</td>
<td>2.4</td>
<td>2.6</td>
</tr>
<tr>
<td>(M_3)</td>
<td>(5.37, 12.9, 66.9)</td>
<td>((3.0^{+0.7}_{-0.5}) \times 10^{-142})</td>
<td>10(^{-4})</td>
<td>6.5 (\times 10^{-4})</td>
<td>1.9</td>
<td>2.2</td>
</tr>
<tr>
<td>(M_4)</td>
<td>(3.15, 5.37, 12.9, 66.9)</td>
<td>((3.0^{+1.1}_{-0.6}) \times 10^{-138})</td>
<td>1.0</td>
<td>10(^{-4})</td>
<td>1.4</td>
<td>1.7</td>
</tr>
<tr>
<td>(M_5)</td>
<td>(3.15, 5.37, 12.9, 66.9, 399)</td>
<td>((3.0^{+2.1}_{\times 0.65}) \times 10^{-136})</td>
<td>10(^{2})</td>
<td>0.01</td>
<td>1.2</td>
<td>1.5</td>
</tr>
<tr>
<td>(M_6)</td>
<td>(3.15, 5.37, 12.9, 34.4, 66.9, 399)</td>
<td>((6.7^{+2.4}_{\times 1/3}) \times 10^{-141})</td>
<td>2.2 (\times 10^{-3})</td>
<td>0.999978</td>
<td>1.0</td>
<td>1.4</td>
</tr>
</tbody>
</table>

Left panel shows the contribution of the individual nested intervals to the NRMC marginal likelihood for the 3 planet model (5 repeats). The right panel shows the integral of these contributions versus the parameter volume of the credible region.