Exoplanet Detection from Bayesian Inference

Marion Dierickx, Xinyi Guo, Philip Mocz
AM207 Final Project

May 13, 2013

Abstract

Analyzing stellar Radial Velocity (RV) curves to determine planetary orbital properties is a statistical challenge. Important degeneracies, sparse sampling, measurement errors and the possible presence of multiple planets suggest the use of a Bayesian framework to extract parameters and select best-fit models. We implement a Nested Sampling method to compute Bayesian evidence and sample the posteriors of the orbital parameters. We first test our algorithm on mock data and show that it successfully converges at low computational cost despite strong multimodality. Next, we apply our method to study the contested HD159868 system, where we identify two low-eccentricity Jupiter mass planets in a limited data set (rather than a single high-eccentricity planet as identified by previous studies.) Our model is in good agreement with recent extended RV observations of the same system, demonstrating the power of the Bayesian framework and our sampling method.

Contents

1 Introduction

Finding planets outside of our own Solar System, known as exoplanets, is one of the most active fields of research in astronomy today. Characterizing these planets has vast consequences

for our understanding of Earth, the phenomenon of life, and our place in the Universe. This is however a challenging goal, as unlike stars, planets are not powered by nuclear fusion and are rarely luminous enough for direct imaging. Most detection methods are therefore inherently indirect.

Historically the most prolific technique for exoplanet detection has been the Radial Velocity (RV) method. With over a dozen ongoing search programs, RV totals approximately 500 finds out of the total ∼800-strong sample of known exoplanets[1]. As its name suggests, the method relies on measuring the line-of-sight velocity of a star with high precision. As the star and its planetary companion(s) orbit their common center of mass, stellar velocity deviations caused by the planets' presence can be detected through spectroscopy. Through the Doppler effect, spectral lines in the stellar spectrum are frequency-shifted depending on the source's velocity. The mechanism is illustrated in Fig. 1. As the star recedes from the observer, the signal is shifted to longer wavelengths, a phenomenon known as redshift. Conversely, as the star approaches the observer in the line-of-sight, the lines are blueshifted, corresponding to negative radial velocity by convention. The resulting signal is a periodic pattern known as an RV curve, which carries a lot of information on the companions' orbital and physical properties. It is worth noting that the method is naturally more sensitive to massive planets close to their host stars, because they cause larger velocity deviations.

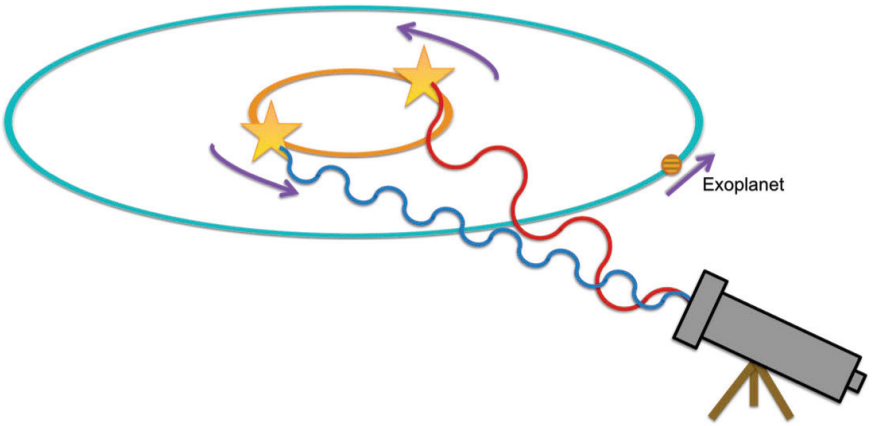

Figure 1: Detection of RV 'wobble' caused by exoplanet companion through Doppler frequency shifting.

RV curves can be modeled following the formalism given in Balan and Lahav [2009] as well as Feroz et al. [2011]. The radial velocity at instant t_i can be calculated as:

$$v(t_i) = V - \sum_{p=1}^{N_p} K_p \left\{ \sin\left[f_{i,p}(t_i, e_p, P_p, \chi_P) + \varpi_p \right] + e_p \sin(\varpi_p) \right\} \tag{1}$$

where:

- $f_{i,p}(t_i, e_p, P_p, \chi_P)$ is the true anomaly, an angle that encodes at what point in the orbit the planet is situated at time t_i. The true anomaly is related to time through Kepler's equation.

[1] www.exoplanet.eu

- K_p is the semi-amplitude of the velocity signature of planet p.
- e_p is the orbital eccentricity.
- ϖ_p is the longitude of periastron (the moment of closest approach between the star and planet): it is a compound angle as it encodes where periastron occurs along the orbit as well as help define the orientation of the orbital plane with respect to the plane of the sky.
- χ_p is the time of periastron - more specifically, the fraction of the orbit that has already elapsed after periastron when data-taking started.
- P_p is the orbital period. Hence the product $\chi_p P_p$ is the number of days that have elapsed since periastron of planet p when data-taking started.
- V is the systemic velocity of the stellar system with respect to the observer.

These orbital parameters can then be used to infer physical properties of the planetary system given a value of the stellar mass, m_s. In particular, we can place lower limits on the semi-major axis of the stellar orbit, a_s, the mass of planet p, m_p, and estimate its semi-major axis, a_p, by the following equations

$$a_s \sin i = \frac{K_p P_p \sqrt{1 - e_p^2}}{2\pi} \tag{2}$$

$$m_p \sin i \approx \frac{K_p m_s^{2/3} P_p^{1/3} \sqrt{1 - e_p^2}}{(2\pi G)^{1/3}} \tag{3}$$

$$a_p \approx \frac{m_s a_s \sin i}{m_p \sin i} \tag{4}$$

We also model the stellar jitter s, which accounts for intrinsic stellar variability, as a source of uncorrelated Gaussian noise. Therefore, for each planet we must model the five free parameters $[K_p, e_p, \varpi_p, \chi_p, P_p]$. In addition, we must also fit the two nuisance parameters s and V common to all planets. We note that in the model described by Equation 1 above, planet-planet interactions are neglected such that the velocity contributions of each planet are simply superimposed. This approximation greatly simplifies the treatment of RV curves and is valid in most cases because the mass of the star greatly dominates the orbital dynamics of the planetary companions.

It is now clearer why the analysis of RV curves is a statistical challenge. Firstly, the orbital parameters outlined above highly degenerate. For example, two low-eccentricity planets can camouflage as one higher-eccentricity companion [O'Toole et al., 2007] Secondly, data-taking occurs only over a finite timespan, and potentially at a infrequent, uneven rate. Through the sampling theorem, this translates into limits on the frequency components that can be recovered from the data. For longer-period planets the data may very well not cover a full orbital cycle, in which case the planet's signal would likely go unnoticed or have to be modeled as a linear trend, as for example in O'Toole et al. [2007]. Together, the frequency component and parameter degeneracy problems cause traditional methods, such as constructing Lomb-Scargle periodograms to fix the period and then applying a χ^2 minimization algorithm, to be

unreliable. Bayesian inference provides a mathematically rigorous answer to both of these issues by treating all parameters as random variables and building up posterior samples.

2 The Radial Velocity Curve Model

As mentioned in Section 1, we model the radial velocity curve with N_p planets assuming no planet-planet interaction with the equation:

$$v(t_i) = V - \sum_{p=1}^{N_p} K_p \left\{ \cos\left[f_{i,p}(t_i, e_p, P_p, \chi_P) + \varpi_p \right] + e_p \cos(\varpi_p) \right\} \tag{5}$$

Here we outline the details of calculating the true anomaly $f_{i,p}(t_i, e_p, P_p, \chi_P)$, which has to be done iteratively as follows.

First, we define the mean anomaly of the planet at time step t_i:

$$M_{i,p} = \frac{2\pi}{P_p} \left(t_i + \chi_p P_p \right) \tag{6}$$

Next, we solve Kepler's equation with a Newton-Raphson iterator to find the eccentric anomaly $E_{i,p}$ at time t_i:

$$E_{i,p} - e_p \sin(E_{i,p}) = M_{i,p} \tag{7}$$

We use an initial guess of $E_{i,p} = \pi$ for the iterator, and continue updating this value until a tolerance level of 10^{-8} for the change of $E_{i,p}$ is achieved.

Finally, the true anomaly $f_{i,p}$ is related to the eccentric anomaly according to:

$$f_{i,p} = 2\arg\left(\sqrt{(1 - e_p)} \cos(E_{i,p}/2), \sqrt{(1 + e_p)} \sin(E_{i,p}/2) \right) \tag{8}$$

3 Bayesian Framework

We use a Bayesian framework for parameter estimation and model selection. Given some data D and a model (hypothesis) H, we can estimate the posterior distribution of the set of parameters θ in the model using Bayes' Theorem:

$$P(\theta \mid D, H) = \frac{P(D \mid \theta, H) P(\theta \mid H)}{P(D \mid H)} \tag{9}$$

where $P(\theta \mid D, H) \equiv P(\theta)$ is the posterior probability distribution of the parameters, $P(D \mid \theta, H) \equiv L(\theta)$ is the likelihood, $P(\theta \mid H) \equiv P(\theta)$ is the prior, and $P(D \mid H) \equiv \mathcal{Z}$, the normalization, is the Bayesian evidence.

If one is interested in parameter estimation only, the normalization factor – the Bayesian evidence $\mathcal{Z}$ – need not be computed. Standard Metropolis-Hastings MCMC sampling of the posterior ignores the normalization.

However, for model selection, which is one of our primary interests, the Bayesian evidence has a central role. It can be evaluated as the average of the likelihood over the prior:

$$\mathcal{Z} = \int L(\theta) P(\theta) \, d^D\theta \tag{10}$$

Model selection between two models H_1 and H_2 can then be decided by a comparison of their posterior probabilities:

$$B = \frac{P(H_2 \mid D)}{P(H_1 \mid D)} = \frac{P(D \mid H_2)P(H_2)}{P(D \mid H_1)P(H_1)} = \frac{\mathcal{Z}_2}{\mathcal{Z}_1}\frac{P(H_2)}{P(H_1)} \tag{11}$$

where the a priori probability ratio of the two models $\frac{P(H_2)}{P(H_1)}$ may be set to 1.

The ratio $\frac{P(H_2|D)}{P(H_1|D)}$ is called the Bayes factor B, and corresponds to the probability of one model being preferred over another. Ratios above 20 are considered strong evidence for one model over the other. A table of the interpretation of Bayes factors [Efstathiou, 2008, Feroz et al., 2011] is presented in Table 1.

Table 1: Scale used for the interpretation of model probabilities

$\mid \Delta \ln B \mid$	Odds	Probability	Interpretation
< 1.0	$\lesssim 3:1$	< 0.750	Inconclusive
1.0	$\sim 3:1$	0.750	Weak Evidence
2.5	$\sim 12:1$	0.923	Moderate Evidence
5.0	$\sim 15:1$	0.993	Strong Evidence

3.1 Likelihood

The errors on the RV curve may be treated as Gaussian. Therefore, we write the likelihood function as:

$$\mathcal{L}(\theta) = \prod_i \frac{1}{\sqrt{2\pi(\sigma_i^2 + s^2)}} \exp\left[-\frac{(v(\theta;\ t_i) - v_i)^2}{2(\sigma_i^2 + s^2)}\right] \tag{12}$$

The σ_i are measurement errors on the individual data points (which include instrumental and observational effects such as weather conditions, seeing, etc.) The stellar jitter s adds additional Gaussian noise to the data-points. If one fits a many-planet system with a model that does not have enough planets, this can systematically push up the value of s. Bad data points will also contribute to s.

3.2 Choice of priors

The choice of priors is an important component of the Bayesian methodology. The choice is to some extent arbitrary and often a point of criticism of the Bayesian method, despite the Bayesian formalism's overwhelming success and elegant framework for handling uncertainties and errors [Efstathiou, 2008, Starck et al., 2013]. For discovery of planetary signals in RV data, we require flat, wide, physically realistic priors on the orbital parameters. A number of different choices for priors with these qualitative features is found in the literature [Balan and Lahav, 2009, Feroz et al., 2011, Wittenmyer et al., 2012]. Adapting these priors, we settle on the prior distributions listed in Table 2 when we investigate real RV data.

The priors allow for a wide range of semi-amplitude K (up to 200 m s^{-1}) and periods P (up to 2000 days). The parameters χ and ϖ related to the angular position of the planet

Table 2: Priors of parameters

	Prior	Mathematical Form	Lower Bound	Upper Bounds
K [m s^{-1}]	Uniform	$\frac{1}{K_{\max}-K_{\min}}$	0	200
ϖ	Uniform	$\frac{1}{2\pi}$	0	2π
e	Gaussian	$\mathcal{N}(0, 0.3)$	0	1
P [d]	Uniform	$\frac{1}{P_{\max}-P_{\min}}$	0.2	2000
χ	1	0.5	0	1
V [m s^{-1}]	Uniform	$\frac{1}{V_{max}-V_{min}}$	-100	100
s [m s^{-1}]	Gaussian	$\mathcal{N}(2.3, 0.1)$	0	10

are uniformly distributed. High orbital eccentricities are allowed but penalized by the choice of prior, as in Wittenmyer et al. [2012]. Systematic velocity shifts V up to $\pm$100 m s^{-1} are allowed. Finally, when investigating the HD159868 system the stellar jitter prior is assumed to be a Gaussian centered at 2.3 m s^{-1}, a physically motivated value for jitter from spots, granulations, and stellar oscillations for a star of this type [O'Toole et al., 2007, Wittenmyer et al., 2012]. For our mock data the prior on s is simply uniform.

Other works, such as Feroz et al. [2011], use an even more extended range on priors, (for example, allow orbital periods up to 365000 days and stellar jitter up to 2000 m s^{-1}.) The extended priors have the effect of increasing the allowed parameter space volume and consequently penalizing higher-number planetary models more in the calculation of the Bayesian evidence. Our choice of priors are somewhat more restricted (though still quite broad) so that we could reasonably sample the posterior distribution given the available computational power of our laptops.

4 Nested Sampling

We implement the Nested Sampling technique in Matlab[2] to estimate parameter posterior distributions and the Bayesian evidence for each model [Skilling, 2004]. This recently developed method excels at efficiently calculating the Bayesian evidence, even in the case where the posterior distribution is multi-modal, and also produces parameter posterior distributions. The method has a number of attractive features: (1) it transforms the multi-dimensional integral in the calculation of the Bayesian evidence into a one-dimensional integral, (2) it samples the posterior distribution with many active particles instead of just one (unlike standard MCMC methods), and (3) the method is generally applicable to arbitrary likelihoods.

The basic idea of Nested Sampling is to sample the posterior space with a number of active points (model parameters) drawn from the prior. Points with the lowest likelihood are then iteratively replaced with points of strictly higher likelihood (also drawn from the prior.) In doing so, one shrinks the 'prior volume' with each iteration, hence the algorithm is called 'nested'.

[2]Our code is included. See the README.txt for a description of the Matlab files.

The 'prior volume' X is defined by:

$$X(\lambda) = \int_{L(\theta)>\lambda} P(\theta) d^D\theta \tag{13}$$

If we define $L(X)$ to be the inverse of $X(\lambda)$, then, the Bayesian evidence can be written as:

$$\mathcal{Z} = \int_0^1 L(X)\, dX. \tag{14}$$

To estimate this integral, we can evaluate a sequence of increasing likelihoods $L_j = L(X_j)$ for a sequence of decreasing values of X_j and use the trapezoid rule:

$$\mathcal{Z} = \sum_{i=1}^{M} L_i w_i \tag{15}$$

where the $w_i = \frac{1}{2}(X_{i-1} - X_{i+1})$ are the trapezoid rule weights.

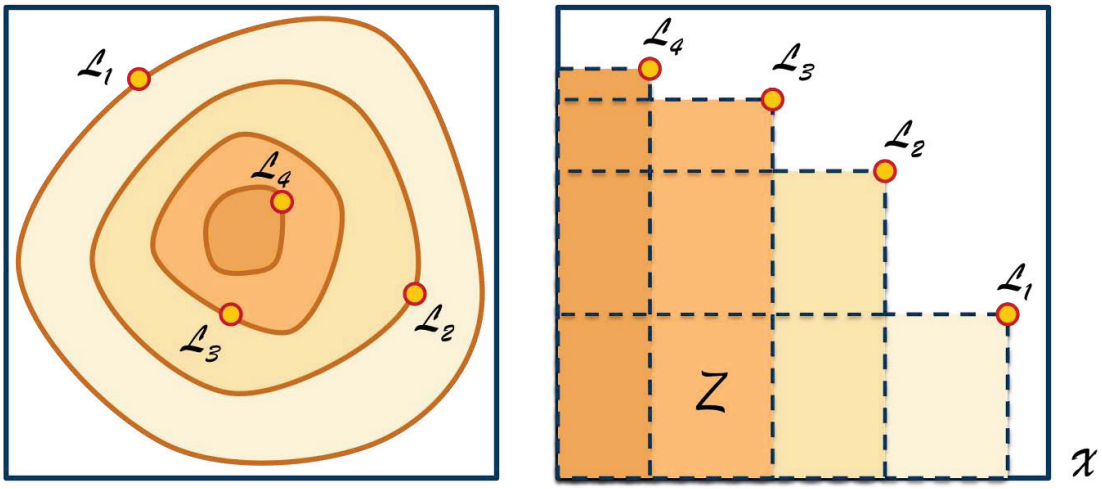

Figure 2: Nested Sampling transforms a multidimensional integral over the prior volume X to a 1D integral that can be evaluated through the trapezoid rule.

4.1 Bayesian Evidence

The calculation of the Bayesian evidence as in Equation 15 is performed as follows. Initially, we draw N 'live'/'active' particles from the prior $P(\theta)$, which can be accomplished with standard techniques such as the inversion method or rejection method. The initial prior volume is $X_0 = 1$. The particles are sorted by their likelihoods and the lowest likelihood particle (likelihood L_0) is removed from the live sample and stored. A new particle with likelihood $L > L_0$ is drawn to replace the particle thrown out. We iterate the discarding of the lowest-likelihood (L_i) active point and replacement by a point of higher likelihood ($L > L_i$). With each iteration i, it can be shown that the prior volume shrinks approximately as $X_i = \exp(-i/N)$.

The Bayesian evidence $\mathcal{Z}$ may be calculated by repeating the iterations until further iterations change $\mathcal{Z}$ by a negligible amount.

4.2 Posteriors

The posterior distributions of model parameters may be obtained as a by-product of the Nested Sampling algorithm. Once $\mathcal{Z}$ is calculated, the posteriors can be found by a weighted resampling of the discarded points, with weights given by:

$$p_i = \frac{L_i w_i}{\mathcal{Z}} \tag{16}$$

4.3 Adaptive Metropolis Nested Sampling

The most challenging part of the Nested Sampling algorithm is the drawing of points with a strict increase in the likelihood $L > L_i$ to replace the discarded points at each iteration. A number of methods may be employed for this sampling procedure. Feroz and Hobson [2008] use a specially developed clustered ellipsoidal sampling method. In the present work, we use an adaptive Metropolis MCMC sampling method. At each iteration, one of the live points θ is picked at random and is used to generate a trial point θ', which is accepted with probability:

$$\alpha = \begin{cases} 1 & \text{if } P(\theta') > P(\theta)\,\&\,L(\theta') > L_i \\ P(\theta')/P(\theta) & \text{if } P(\theta') < P(\theta)\,\&\,L(\theta') > L_i \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

For each parameter, we use a Gaussian distribution as the proposal distribution with standard deviation σ specific to the parameter. We take 20 MCMC steps away from the starting point to avoid correlations in the sampling.

The value of σ is updated at the end of each nested sampling iteration in order to maintain the acceptance rate at around 50 percent. σ is updated as:

$$\sigma \rightarrow \begin{cases} \sigma \exp(1/N_{\text{accept}}) & \text{if } N_{\text{accept}} > N_{\text{reject}} \\ \sigma \exp(-1/N_{\text{reject}}) & \text{if } N_{\text{accept}} \leq N_{\text{reject}} \end{cases} \tag{18}$$

where N_{accept} and N_{reject} ate the number of accepted and rejected samples in the previous MCMC sampling phase.

A pseudocode of our Nested Sampling algorithm is presented below:

Nested Sampling(){
% initialize active points by sampling prior
Draw N points $\{\theta_1, ..., \theta_N\}$ from the prior distribution
for $i = 1 : M$
  % find active point with least likelihood
  L_i = min(likelihoods of the N active points $\{\theta_1, ..., \theta_N\}$);
  $\theta_{\text{min},i} = \theta_k \in \{\theta_1, ..., \theta_N\}$ with the minimum likelihood
  $X_i = \exp(-i/N)$;
  % replace point with higher likelihood point with Metropolis MCMC
  Pick a $\theta \in \{\theta_1, ..., \theta_N\}, \theta \neq \theta_{\text{min},i}$ randomly
  $\theta_{\text{replace}} = \theta$;

N_accept $= 0$
while N_accept $<$ N_mcmc
Propose a new point θ' around $\theta_{\rm replace}$ using multivariate Gaussian proposal distribution
Perform MCMC with acceptance probability

$$\alpha = \begin{cases} 1 & \text{if } P(\theta') > P(\theta) \,\&\, L(\theta') > L_i \\ P(\theta')/P(\theta) & \text{if } P(\theta') < P(\theta) \,\&\, L(\theta') > L_i \\ 0 & \text{otherwise} \end{cases}$$

if θ' is accepted
$\theta_{\rm replace} = \theta'$
N_accept $+= 1$;
end
end
Store $\theta_{{\rm min},i}$
Replace $\theta_{{\rm min},i}$ with $\theta_{\rm replace}$ in the set of active particles $\{\theta_1, ..., \theta_N\}$
% adaptive MCMC stepping
update multivariate Gaussian proposal distribution standard deviations σ with

$$\sigma \rightarrow \begin{cases} \sigma * \exp(1/\text{N\_accept}) & \text{if N\_accept} > \text{N\_reject} \\ \\ \sigma * \exp(-1/\text{N\_reject}) & \text{if N\_accept} \leq \text{N\_reject} \end{cases}$$

end
% Calculate Bayesian Evidence with trapezoid rule
$\mathcal{Z} = \sum_{i=1}^{M} L_i w_i$
% Obtain posteriors
to obtain posterior distribution, resample the points $\theta_{{\rm min},i}$ with weights
$p_i = \frac{L_i w_i}{\mathcal{Z}}$
}

5 Results

We test our Bayesian nested sampling approach on both synthetically generated data and real observations.

5.1 Mock Data

First, we test our method on synthetic data sets containing 1 and 2 planets.

5.1.1 Single planet synthetic RV curve

We generate a single planet RV curve with 40 data points with Gaussian noise of approximately 3 m s^{-1} and with planetary orbital parameters listed in the first column of Table 3.

We employ our nested sampling algorithm on the data, first fitting a single planet model. For these simple test cases, we use $N = 30$ live particles and iterate for 1500 steps. The data and the maximum a posteriori (MAP) model found by our run is shown in Fig. 3. The

Table 3: Mock data planet parameters

	planet 1	planet 2
K [m s^{-1}]	60	57
ϖ	1.0	1.23
e	0.2	0.52
P [d]	100	345.72
χ	0.3	0.5
V [m s^{-1}]	0.0	–
s [m s^{-1}]	1.0	–

posteriors for the orbital parameters are presented in Fig. 4, which show good agreement with the true values. Fig. 5 shows the calculation/update of the Bayesian evidence $\mathcal{Z}$ as a function of iteration number. The figure shows that our estimate for $\mathcal{Z}$ has converged.

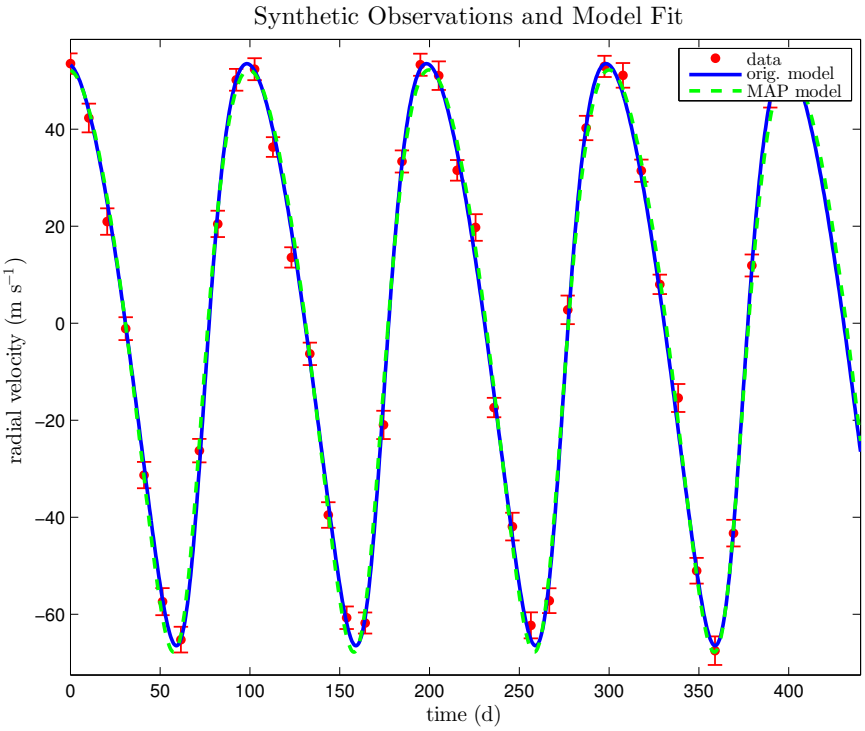

Figure 3: Single planet mock data and MAP fit for a single planet model. The model shows good agreement with the data.

Next we consider fitting a two planet model to the single planet data, and comparing the Bayesian evidence. The method converges on two planets whose signal adds up to appear like a single planet. The 2 planetary parameters are listed in Table 6.

The Bayesian evidences for the 1 planet and 2 planet models are listed in Table 4. We find that a 1 planet model is e^{22} times more probable than a two-planet model, which qualifies as strong evidence for the 1 planet model over the 2 planet model.

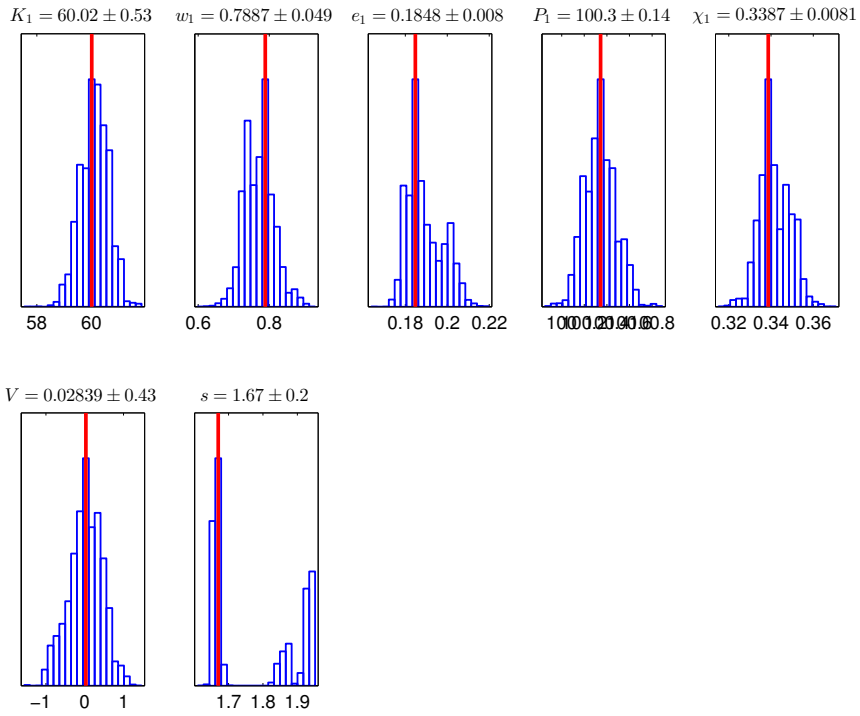

Figure 4: Orbital parameters posterior distributions for a single planet model fit to single planet mock data.

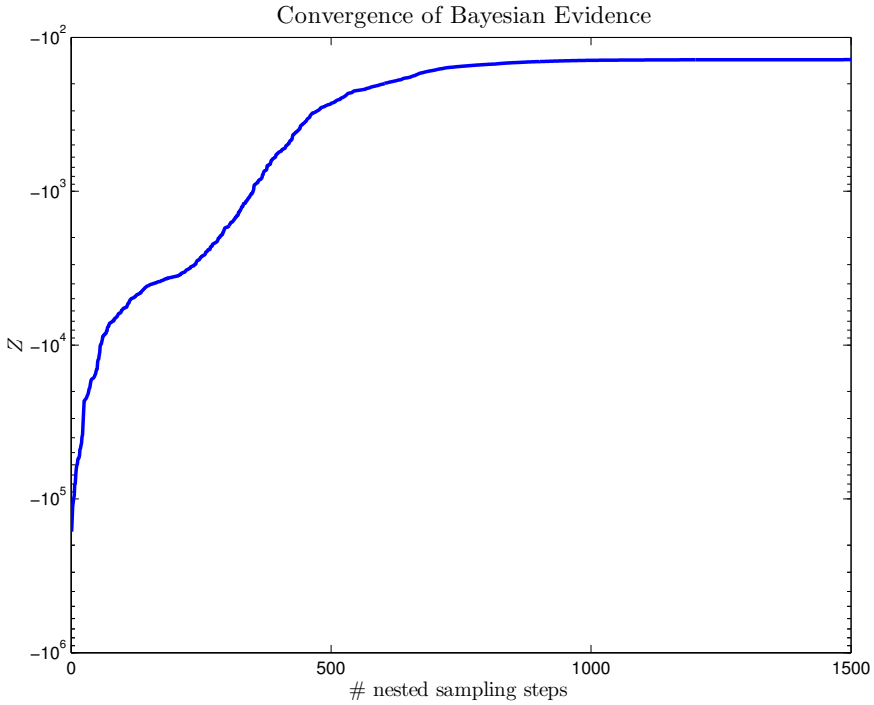

Figure 5: Convergence of the calculation of $\mathcal{Z}$ for a one-planet fit to a one-planet mock dataset.

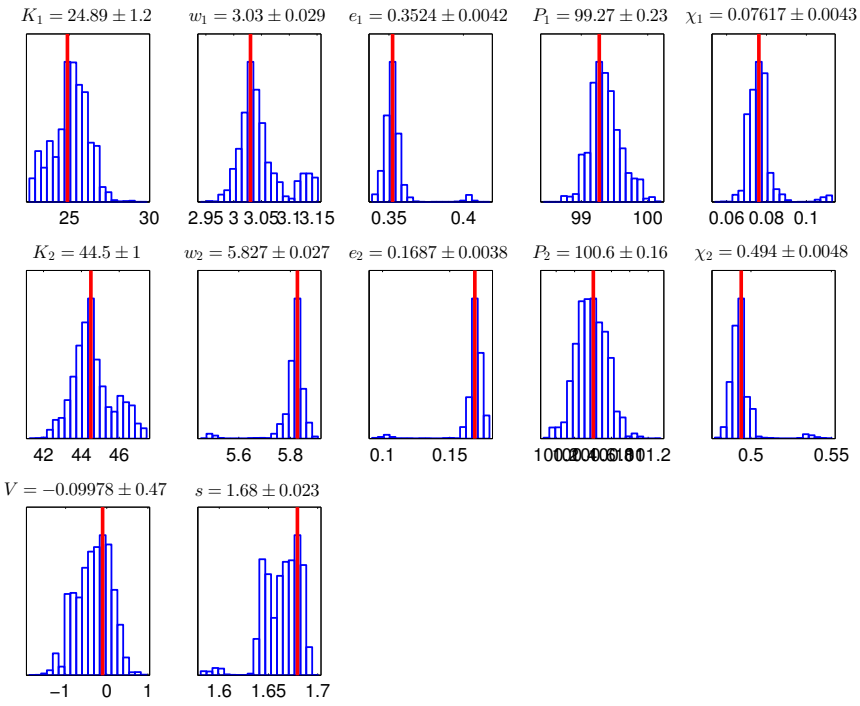

Figure 6: Posterior distributions of parameters of a double planet model fit to single planet mock data.

Table 4: Bayesian evidences for 1 planet mock data

planets in model	$\log \mathcal{Z}$
1	−139.57
2	−162.51

The single planet is preferred over the two-planet fit despite the fact that we could have generated exactly the same mock dataset considering two identical planets of half semi-amplitude. The Bayesian evidence incorporates Occam's razor and prefers simpler models (smaller volume of parameter space) over more complicated models.

5.1.2 Two-planet synthetic RV curve

Next, we generate mock RV data for the signal expected from both planets listed in Table 3. The orbital parameters of the second planet are inspired by observations from Lo Curto et al. [2010].

We carry out the strategy of considering first a one-planet model and then repeating our analysis by adding an additional planet until the Bayesian evidence decreases. The MAP fits to the data for 1, 2, and 3 planet models are shown in Fig. 7. The Bayesian evidences are listed in Table 5. We find that the two planet model is preferred over the single planet model by a factor of e^{1986}. In addition, the two-planet model is preferred over the three-planet model by a factor of e^{29}. The two-planet model is strongly supported over the other models by the Bayesian evidence.

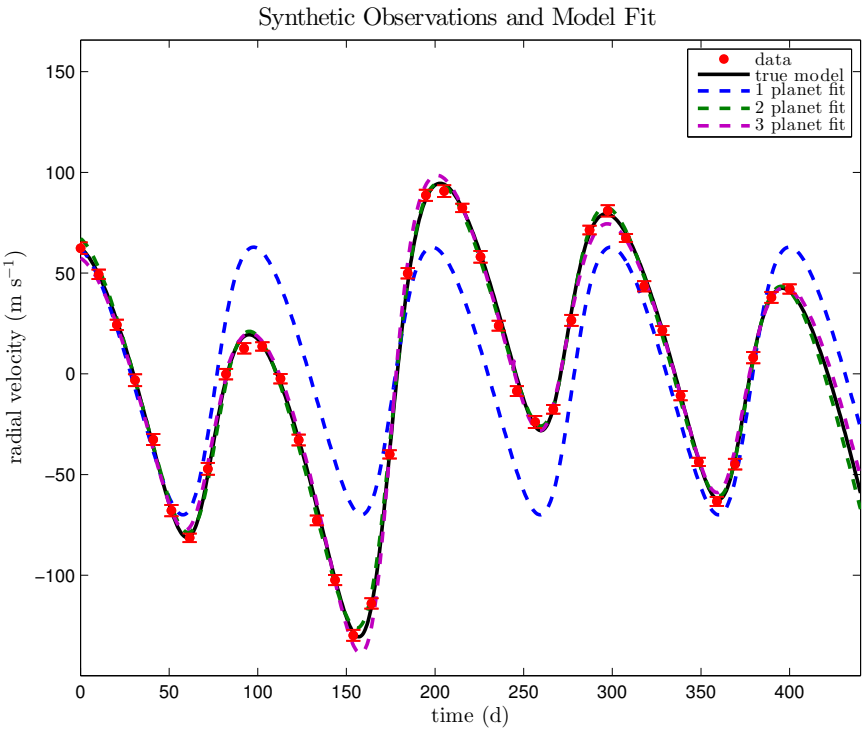

Figure 7: Two planet mock data with 1,2 and 3 planet MAP fits to the data.

Table 5: Bayesian evidences for two-planet mock data

planets in model	$\mathcal{Z}$
1	−2141.0
2	−154.79
3	−184.28

The posteriors of the two-planet fit to the mock data are shown in Fig. 8, which shows we recover the true parameters nicely, validating our method and implementation.

We graphically illustrate the nested sampling algorithm in Fig. 9 at different stages in its iteration for sampling the two-planet model posterior distribution to provide visual intuition about how the method works.

5.2 Application to Real Data: HD159868

The existence of a planetary system orbiting the star HD159868 was first announced by O'Toole et al. [2007]. At the time of publication, the star had been observed 28 times since 2002 through the Anglo-Australian Planet Search program (AAPS). The authors first performed a two-dimensional Keplerian Lomb-Scargle periodogram analysis on the RV curve to find the period with maximum spectral power and use it as an initial guess for a non-linear least-squares fit. With a one-planet model, their best fit still has a reduced-chi-squared fit of $\chi^2_\nu = 15.22$. The large residual motivated them to search for an additional

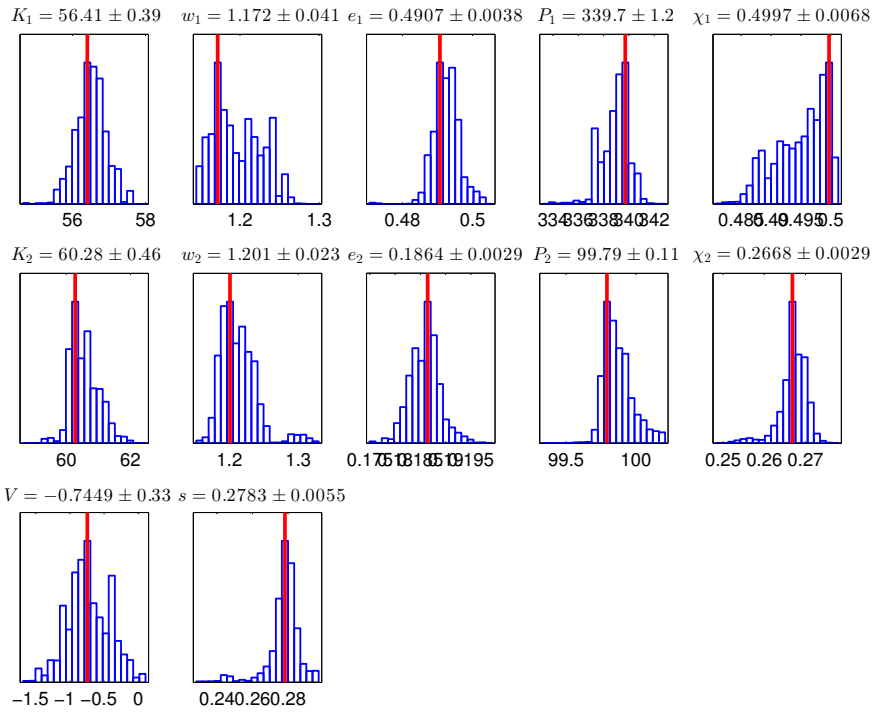

Figure 8: Posterior distributions of parameters of a two planet model fit to two-planet mock data.

planet. With a two-planet model, they managed to reduce the χ^2_ν to 4.4. But without the Bayesian framework, they were not able to make a definite statement on the model selection. Balan and Lahav [2009] used their software EXOFIT to estimate the orbital parameters of the HD159868 system in a Bayesian framework utilizing a Metropolis-Hasting MCMC algorithm. Using the same data set as O'Toole et al. [2007], they noted that the stellar jitter s, an indicator for noise, is 10 times higher than the errors in the measurement, suggesting the presence of an additional planet. However, lacking the ability to calculate the Bayesian evidence, they did not investigate further on whether a two-planet model fits the data better. Recently, Wittenmyer et al. [2012] found a second planet around HD159868 using extended observations from the AAPS (total 47 epochs) combined with 37 additional Keck data. These previous works suggest that HD159868 is an exciting case for testing our algorithm's ability to select models and fitting orbital parameters. We run our analysis on the original (non-extended) data set provided by [O'Toole et al., 2007] and then check whether it successfully selects the two-planet model as Wittenmyer et al. [2012] concluded.

We investigated a number of prior choices, particularly for the stellar jitter. In fitting this particular case, we settled on a more informed prior on s (compared to a wide modified-Jeffery's prior used by Feroz et al. [2011]), because the data is sparsely sampled and a very generous prior on s will yield high likelihood for a lot more combination of parameters, giving unreliable results. O'Toole et al. [2007] assumed $s = 2$ m s^{-1} in performing the least-squared fit while Wittenmyer et al. [2012] fixed $s = 2.65$ m·s^{-1} in their Bayesian analysis. We thus adopt a Gaussian prior, $s \sim \mathcal{N}(2.3, 0.1)$, which covers both estimates on the stellar jitter of HD159868.

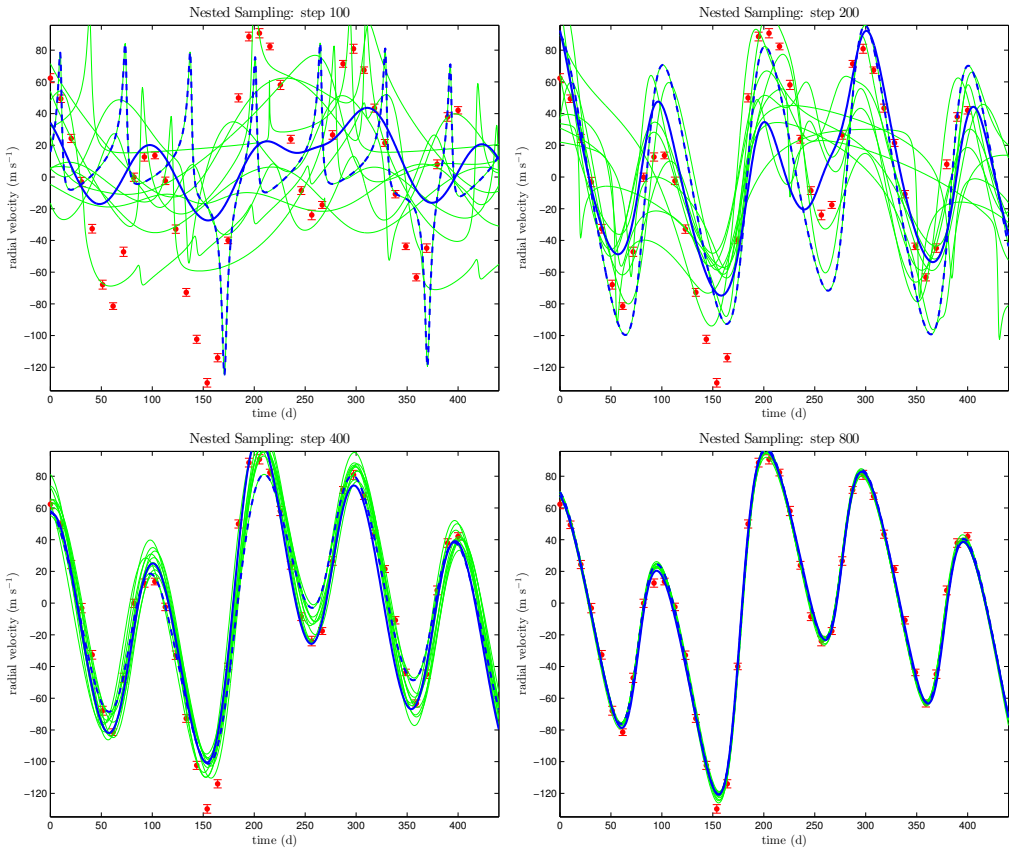

Figure 9: Illustration of the Nested Sampling algorithm in action. The figures illustrate the algorithm at iteration steps 100, 200, 400, and 800. The data points are shown as red points with error bars. The thin green lines show the previous 20 models with lowest-likelihoods that were discarded from the set of active points. The thick dashed blue line shows the model corresponding to the current point that is being discarded. The thick solid blue line shows the model of strictly greater likelihood that is replacing the discarded model. As time progresses, the sampling identifies the correct model parameters.

We run our nested sampling algorithm on the data with $N = 250$ live particles and iterate for 20000 steps.

The first column of Table 6 summarizes the evidence of 1-, 2-, 3-planet models on the old RV data set of HD159868. Using a 1-planet model, our MAP set of parameters is almost identical to the least-squares fit by [O'Toole et al., 2007]. Using a 2-planet model, we obtain a significantly better fit as illustrated in Fig. 11. The marginalized posterior distribution of the parameters are shown in Fig. 10 and the MAP values are summarized in Table 7. Even with the old data, our result is very comparable to that of Table 6 in Wittenmyer et al. [2012]. Most importantly, by comparing the evidence, we find that the 2-planet model is $e^{30.6}$ more probable than the 1-planet model, allowing us to make a decisive statement that there is a second planet in the system. In addition, the parameters of the 2-planet model we inferred by fitting the old data set predict the additional data points very well, as illustrated in Fig. 12.

Interestingly, there is some lingering probability that the 3-planet model is more favorable than the 2-planet model. Considering that the old data set is both limited in number of data points and sampling interval, we decided to compare the Bayesian evidence of 2- and 3-planet model using the new extended data set. The second column in Table 6 summarizes the evidence. We see that the preference of a 3-planet model is a lot weaker with more data. Furthermore, if we compare the posterior distribution of the 2-planet and 3-planet models using the new data, we can clearly identify 2 planets that share about the same e, P, K values and a third planet with a suspiciously small value of $K_3 = 0.26 \pm 0.39$, which is a lot smaller than both the measurement error and the stellar jitter, and also consistent with a semi-amplitude of 0. This suggests that the "third planet" is fitting the noise in the signal, and as a result we do not believe the existence of a third planet given the current analysis on the available data.

Table 6: Bayesian evidences of various models for the RV data sets of HD159868

planets in model	$\log \mathcal{Z}$ (old data)	$\log \mathcal{Z}$ (new data)
1	-187.2	
2	-146.6	-241.0
3	-137.3	-237.5

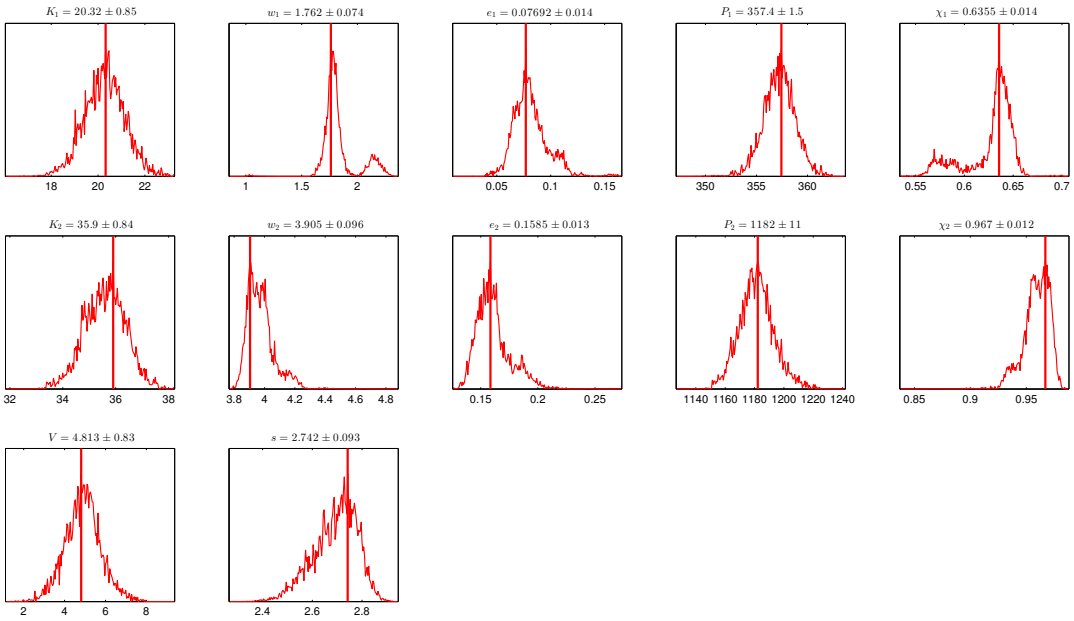

Figure 10: Posterior distributions for the two-planet model orbital parameters derived for the restricted dataset for HD159868.

Based on this model selection, we infer the physical properties of the planetary system HD159868 in a Bayesian framework. We assumed $m_s = 1.087 M_\odot$ [Wittenmyer et al., 2012]

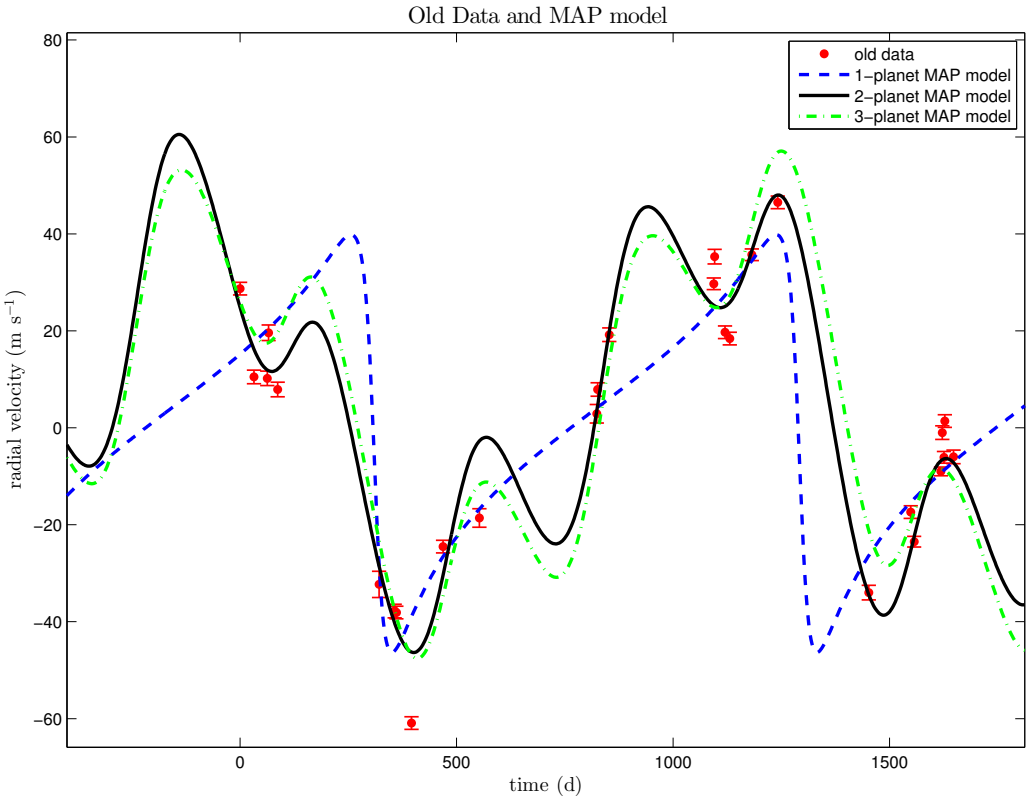

Figure 11: The 1-, 2- and 3-planet MAP fits overplotted on the restricted dataset for HD159868.

and apply Equations 2, 3, 4 to the posterior distribution of K, P, and e obtained from the Bayesian analysis on the new data set assuming a 2-planet model (shown in the upper panel of Fig. 13.) Fig. 14 shows the posterior distribution of $a_s \sin_i$, $m_p \sin i$ and a_p. Based on the relatively large value of semi-major axes and lower limit of the masses, we can rule out the possibility of any of them being Hot Jupiter-type planets. In addition, the first planet has a very large lower limit on mass, which indicates a slight possibility of it being a brown dwarf, which has a lower mass limit of $\sim 13 M_J$. In particular if the inclination angle i_1 is less than 8.85°, then the first "planet" can qualify as a brown dwarf.

6 Discussion

The calculation of the Bayesian evidence incorporates the principle of Occam's razor (all things being equal, a simpler hypothesis is selected.) One can generate a simple mock data set as we have done in multiple ways: for example with a single planet or with two identical planets of half the semi-amplitude (these produce identical RV curves.) The calculation of the Bayesian evidence leads us to favor the simpler 1-planet model over the 2-planet model, because the volume and dimension of the parameter space is much smaller. A two-planet

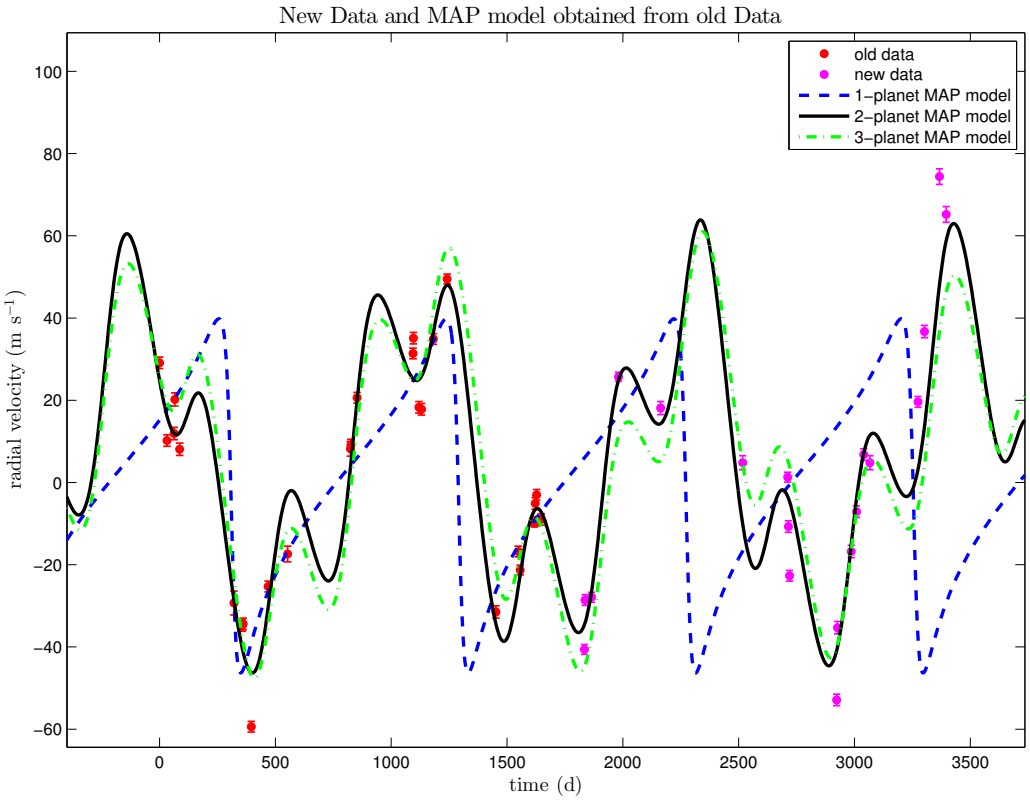

Figure 12: The 1-, 2- and 3-planet MAP models obtained from the restricted dataset over-plotted on the entire dataset for HD159868.

model would need to significantly increase the likelihood of the fit in order to be preferred.

The Bayesian formalism can be quite powerful even when the number of data points is small and comparable to the number of parameters in the model. In such a case, the model posteriors end up being wider than if there were more data points, but the Bayesian evidence may still be used to select between models.

We applied our nested sampling method to the planetary system HD159868, which has limited, potentially noisy RV data. Ideally, when the data is constraining, the assumption of prior should have minimal impact on the distribution of the posterior. However, in the application to the sparsely sampled, older data set of HD159868, we found that the posterior can vary significantly with the choice of priors on s. A very generous prior on s (such as the wide modified-Jeffreys prior used by Feroz et al. [2011]) increases the prior volume, and our choice of number of live particles based on the capacity of our laptops might not be sufficient to fully explore the larger volume. We plan to investigate this issue further by optimizing our code for parallel computation and exploiting more powerful computational resources so that we can increase the number of active particles used to sample the prior volume by several orders of magnitude. Exploring a larger parameter space for the priors will likely reduce the ratio of the Bayesian evidence of the three-planet fit to the two-planet fit by Occam's razor principle built-in to the Bayesian model selection formalism. The added prior volume

Table 7: Estimated parameters of the two-planet model of the old data of HD159868.

	planet 1	planet 2
K [m s^{-1}]	20.32 ± 0.85	35.9 ± 0.8
ϖ	1.762 ± 0.074	3.905 ± 0.096
e	0.077 ± 0.014	0.159 ± 0.013
P [d]	357.4 ± 1.5	1182 ± 11
χ	0.636 ± 0.014	0.967 ± 0.012
V [m s^{-1}]	4.81 ± 0.83	–
s [m s^{-1}]	2.742 ± 0.0934	–

is expected to compensate for the seemingly larger likelihood of the fit. We note that in our three-planet fit of the data, the third planet has small semi-amplitude consistent with 0 and noticeably wider posterior distributions on its orbital parameters than the other two planets.

7 Conclusions

We implemented the nested sampling method to select multi-planet models based on Bayesian evidence and infer the orbital parameters. We applied our method to synthetic data sets and successfully recover the parameters used to generate the data, validating our implementation.

We applied our method to the system HD159868 and found overwhelming preference of a two-planet model over a one-planet model. Our 2-planet MAP fit obtained from the old data set fits the RV curve extremely well even when new data points are added (with no re-fitting!) Despite a slight increase in evidence of the 3-planet model, we prefer the 2-planet model based on careful study of the posterior distribution of the "third planet", which we discarded as being a case of fitting the noise. We used the posterior distributions of the orbital parameters of the two-planet model to infer the mass and semi-major axes of the HD159868 companions in a Bayesian framework. We ruled out the possibility that any one of them is a hot Jupiter and conclude that the more massive one has some chance of being a brown dwarf if the inclination is small.

8 Acknowledgments

We would like to thank P. Protopapas, R. Dave, and A. Goulding for their feedback during the Harvard SEAS Design Fair poster session.

References

S. T. Balan and O. Lahav. EXOFIT: orbital parameters of extrasolar planets from radial velocities. *MNRAS*, 394:1936–1944, April 2009. doi: 10.1111/j.1365-2966.2008.14385.x.

G. Efstathiou. Limitations of Bayesian Evidence applied to cosmology. *MNRAS*, 388:1314–1320, August 2008. doi: 10.1111/j.1365-2966.2008.13498.x.

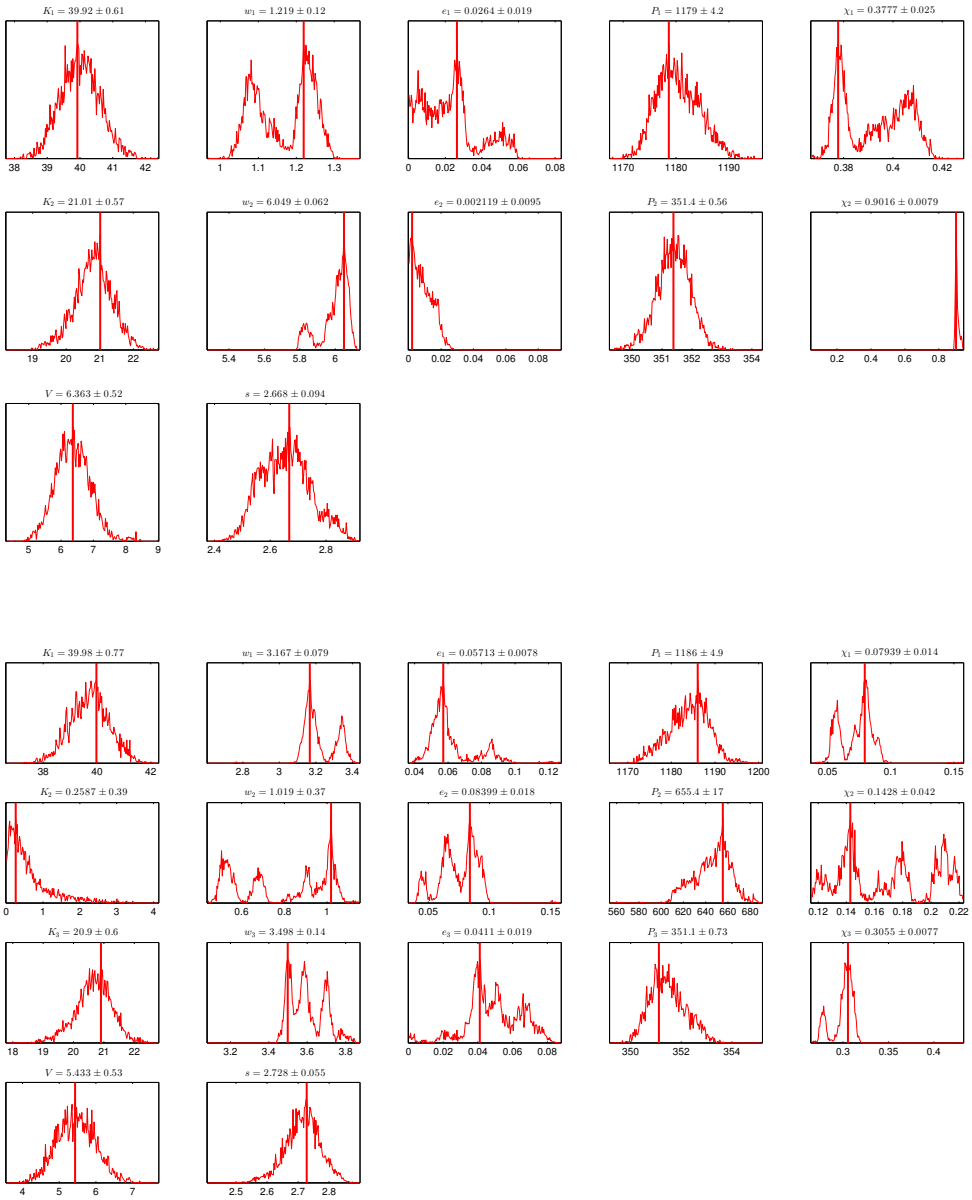

Figure 13: The upper panel shows the posterior distributions of parameters estimated from the 2-planet model on the extended data. The lower panel of plots shows posterior distributions for parameters estimated from the 3-planet model on the same data set.

F. Feroz and M. P. Hobson. Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. *MNRAS*, 384: 449–463, February 2008. doi: 10.1111/j.1365-2966.2007.12353.x.

F. Feroz, S. T. Balan, and M. P. Hobson. Detecting extrasolar planets from stellar radial velocities using Bayesian evidence. *MNRAS*, 415:3462–3472, August 2011. doi: 10.1111/j.1365-2966.2011.18962.x.

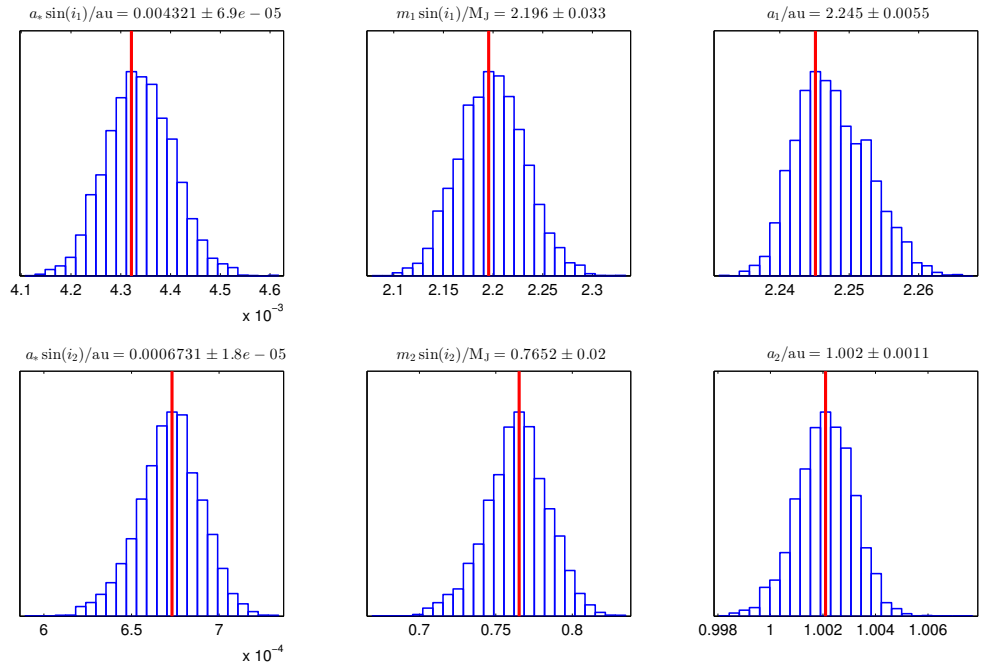

Figure 14: Posterior distribution of the physical properties of the planetary system HD159868.

G. Lo Curto, M. Mayor, W. Benz, F. Bouchy, C. Lovis, C. Moutou, D. Naef, F. Pepe, D. Queloz, N. C. Santos, D. Segransan, and S. Udry. The HARPS search for southern extra-solar planets . XXII. Multiple planet systems from the HARPS volume limited sample. *A&A*, 512:A48, March 2010. doi: 10.1051/0004-6361/200913523.

S. J. O'Toole, R. P. Butler, C. G. Tinney, H. R. A. Jones, G. W. Marcy, B. Carter, C. McCarthy, J. Bailey, A. J. Penny, K. Apps, and D. Fischer. New Planets around Three G Dwarfs. *ApJ*, 660:1636–1641, May 2007. doi: 10.1086/513563.

J. Skilling. Nested Sampling. In R. Fischer, R. Preuss, and U. V. Toussaint, editors, *American Institute of Physics Conference Series*, volume 735 of *American Institute of Physics Conference Series*, pages 395–405, November 2004. doi: 10.1063/1.1835238.

J.-L. Starck, D. L. Donoho, M. J. Fadili, and A. Rassat. Sparsity and the Bayesian perspective. *A&A*, 552:A133, April 2013. doi: 10.1051/0004-6361/201321257.

R. A. Wittenmyer, J. Horner, M. Tuomi, G. S. Salter, C. G. Tinney, R. P. Butler, H. R. A. Jones, S. J. O'Toole, J. Bailey, B. D. Carter, J. S. Jenkins, Z. Zhang, S. S. Vogt, and E. J. Rivera. The Anglo-Australian Planet Search. XXII. Two New Multi-planet Systems. *ApJ*, 753:169, July 2012. doi: 10.1088/0004-637X/753/2/169.