ExoFit User's Guide

Sreekumar Thaithara Balan & Ofer Lahav

September 15, 2008

1 Introduction

ExoFit is a software for extracting orbital parameters of extra-solar planets from radial velocity data. It can search for either *one or two planets* and uses Markov Chain Monte Carlo(MCMC) method to estimate orbital parameters and their uncertainties. *ExoFit* is presented in [1].

2 Installing *ExoFit*

- **Step 1:** Download the tar ball ExoFit.v2.tar.gz from

  http://zuserver2.star.ucl.ac.uk/~lahav/exofit.html

- **Step 2:** Go to directory where ExoFit.v2.tar.gz is downloaded and in the terminal type:

  ```
  tar xvzf ExoFit.v2.tar.gz
  ```

  This will extract all the source files into a directory named ExoFit.vX.XX.

- **Step 3:** *ExoFit* needs GSL - GNU Scientific Library. In a typical installation of GSL, GSL header files are located in a directory named `gsl` under `/usr/include/` and the `lib` files are located in `/usr/lib/`. If you have installed gsl somewhere else please specify the path to `gsl` in `Makefile`. The first two lines of the `Makefile` should be edited to specify `gsl` header files and libraries. For example if you have installed GSL in `/home/visitor/usr/local/`, the first two lines should be modified to:

  ```
  INCDIR=/home/visitor/usr/local/include
  LIBDIR=/home/visitor/usr/local/lib
  ```

  Now you can compile the source files.

- **Step 4:** We assume that you have g++ from GCC,the GNU Compiler Collection. To compile and make the executable type:

  ```
  make
  ```

This will create two executable files *exofit* and *plotmaker*. Now we are ready to go. You may put these executable files in `~/bin` or any other directory in your `PATH` for ease of use.

3 Using *ExoFit*

To use exofit type[1]

`exofit` [OPTION] *path/rvdata.dat*

Where *rvdata.dat* is the file containing radial velocity data, *path* is the path to the file *rvdata.dat* and [OPTION] can be either *-p=1* or *-p=2* indicating the radial velocity model (1-planet or 2-planets) that should be chosen for the search. If you do not provide any options here, `exofit` runs with the default option of *1-planet*. For example if your radial velocity data is in */home/data/HD187085.dat* and you would like to search for 1-planet, the you should type:

```
exofit -p=1 /home/data/HD187085.dat
```

Modelling of Radial velocity is described in Appendix [A]. A brief introduction to Bayesian analysis given in Appendix [B]. Appendix [C] explains the MCMC method.

4 Input Data

The input to *exofit* is the radial velocity data and the state data.

4.1 Radial Velocity Data

120.9170	-12.1	5.3
411.0753	-2.5	6.5
683.1693	16.1	5.7
743.0494	9.3	5.5
767.0046	8.8	4.7
769.0652	5.3	4.4
770.1153	5.3	5.2
855.9477	8.6	8.9
1061.2140	-4.5	5.3

Table 1: Form of the radial velocity data. The entries shown here are from [17] for HD187085. Columns 1, 2 and 3 show time, radial velocity and uncertainty in measurement respectively.

[1] We assume that `exofit` is in your `PATH`. Otherwise specify the full path to `exofit` like `/home/visitor/programs/exofit`. You can copy `exofit` and `plotmaker` to the present working directory and type `./exofit /home/data/HD187085.dat`

Radial velocity data is a simple text file with format shown in Table[1]. The data has 3 columns. Column 1 is the time coordinate, Column 2 is the radial velocity in ms^{-1} and Column 3 is the uncertainty in measurement also in ms^{-1}.

4.2 State Data

The state data *state.dat* defines the starting points of the Markov Chain and the prior boundaries. If the file *state.dat* is present in the same directory as the *exofit*, then the values from *state.dat* is taken as input parameters. If it is not found then *exofit* runs with default state values which are identical to the values shown in the table.If you are not familiar with MCMC do not keep this file in the same directory as *exofit*. *The program works fine even if this file is not present.* Each column in *state.dat* has the following format:

Parameter
Minimum
Start Value
Maximum
Step Size

V	T1	K1	e1	w1	X1	s
-2000.0	0.2	0.00001	0.0	3.0	0.0	0.0
0.0	7500.0	1000.0	0.50	4.145	0.5	1000.0
2000.0	15000.0	2000.0	0.99999	6.28318	0.99999	2000.0
400.0	1500.0	200.0	0.1	0.628	0.1	200.0

Table 2: Form of the text file *state.dat* for a single planet model, were $V, T1, K1, e1, w1, X1, s$ stands for *period* in days, *amplitude*, in ms^{-1}, *systematic velocity* in ms^{-1}, *eccentricity*, *longitude of periastron* in radians, *periastron passage factor* and *noise factor* in ms^{-1} respectively. Similarly, for a 2-planet model, state data should contain prior boundaries and step sizes for 12 parameters. Sample state data is provided with `exofit` package and can be found in the directory *state_data*

A sample state data is shown in Table[2]. The fist row in the data file is called a *header* and it contains the names of each parameter in the state. The second row contains the minimum values of each parameter, third row contains the starting points of MCMC[2], the forth row specifies the maximum values of each parameter and the fifth row defines the step sizes parameters. Each column should be separated by white spaces.

5 Output

The output of *exofit* is again a text file called *extract.dat*. This file, as shown in Table[3] has a header which contains the names of the parameters in the state and current strength of the state (G in Table[3]). After the header each row represents the state of the parameters (in the same order as in the header) at

[2] Starting values of MCMC should be between minimum and maximum vales of the parameter.

each iteration in the MCMC and the strength at that iteration. We call this an MCMC extract. The program also produces two other files namely *burn.dat* and *diag.dat*. *burn.dat* has the same format as the MCMC extract. However these values are considered as burn in and should not be used for the calculation of densities. *diag.dat* has again the same format, but shows the mean and standard deviation of each parameter for every 10000 iterations. These files shows the progress of the Markov Chain towards its stationary distribution. *extract.dat* is the final output of *ExoFit*. You may use your own statistical visualisation programs to analyse *extract.dat* or make use of the scripts provided below.

V	T1	K1	e1	w1	X1	s1	G
-2.0020	1037.0840	14.3680	0.2761	0.4046	0.0661	5.3891	-162.1100
-0.9926	1003.8610	15.5710	0.1887	0.0411	0.1066	4.9867	-162.2000
-0.7553	1080.5220	15.1040	0.3282	0.3089	0.1552	5.2289	-160.8500
0.3385	1005.3890	17.5910	0.2941	0.2238	0.0303	4.6077	-165.5800

Table 3: A sample MCMC extract for 1-planet model.

6 Statistics and Visualisation

6.1 Basics

The output of the *exofit* is used to estimate the orbital parameters and their uncertainties. This can be done with the help of any statistical packages available. We used R to find out the statistical summaries and produce their visualisation. R is a robust environment for statistical computation which is freely available. It can be obtained from http://www.r-project.org/. We have provide an R script that calculates the statistical summaries, plots the posterior densities of each parameter and finally make the radial velocity curve using median of samples from the MCMC extract. This file could be found under the directory `scripts`. The procedure is explained below. We assume that your machine has R installed on it.

- **Step 1:** Start an R session. Open a terminal session and type

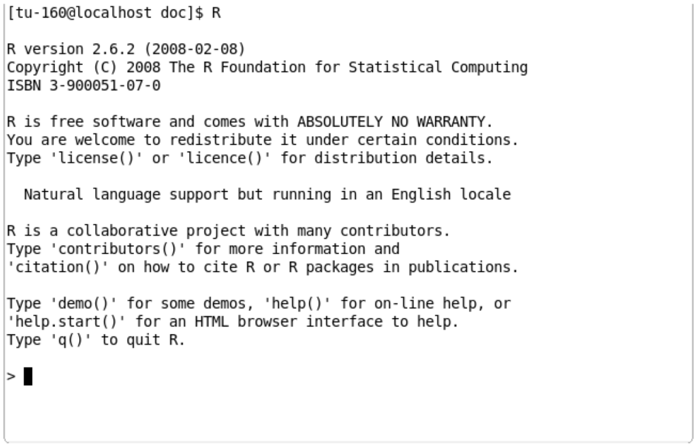
  ```
  R
  ```

  This will start a new R session. You should get an output as shown in Figure[1]. To make the density plots and errorbars, the following packages must be present in R.

  1. **lattice**
  2. **grid**
  3. **MASS**
  4. **coda**

  To install these packages use the command:

  ```
  available.packages()
  ```



```
[tu-160@localhost doc]$ R

R version 2.6.2 (2008-02-08)
Copyright (C) 2008 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> █
```

Figure 1: R session

This will open up a dialogue box as shown in Figure[2]. Choose an appropriate mirror to download the packages from. To install a packages use the command *install.packages('package')*. For example, to install the package *coda* type:

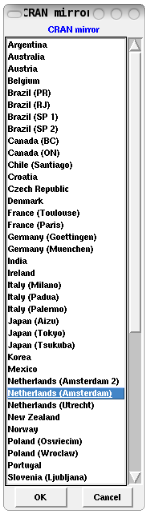


Figure 2: CRAN mirrors

```
install.packages('coda')
```

To install some of the packages you might need the R header files. You can get help for a particular command with *help()*. For example to get help for *install.packages()* type:

```
help(install.packages)
```

You may also use the documentation available at the CRAN page at R Installation and Administration. To quit an R session use the command:

```
q()
```

- **Step 2:** If you are running *exofit* and *plotmaker* from the present working directory, load the R script file onto the current R session by typing:

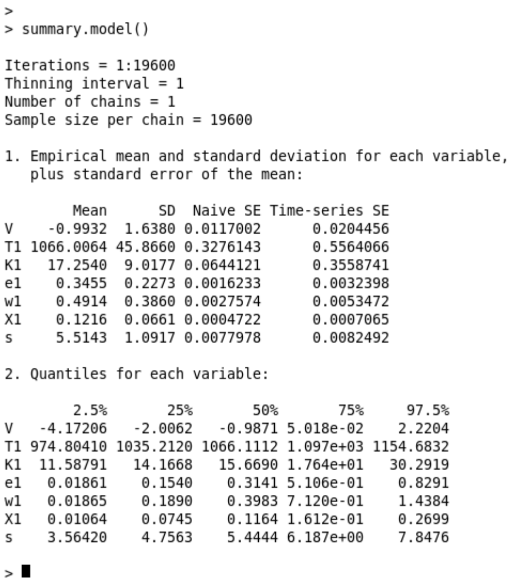
  `source(‘‘`*path/*`orbit_plot.R’’)`

  Where *path* refers to the path to *orbit_plot.R*. This scripts loads the data from *extract.dat* and calculates the statistical summaries. You need to have *extract.dat* and the executable `plotmaker` in the present working directory for this R script to work! Otherwise you need to specify the `PATH` to *plotmaker* in the R script *orbit_plot.R* by modifying the line `system('./plotmaker')` to `system('PATH/plotmaker')` where `PATH` refers to the path to *plotmaker*.

- **Step 3:** Display the estimates of orbital parameters in the radial velocity model. Type:

  `summary.model()`

  This should produce an output as shown in the Figure [3][3] and the medians of all the samples from *extract.dat*.



```
>
> summary.model()

Iterations = 1:19600
Thinning interval = 1
Number of chains = 1
Sample size per chain = 19600

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

        Mean      SD  Naive SE Time-series SE
V    -0.9932  1.6380 0.0117002      0.0204456
T1 1066.0064 45.8660 0.3276143      0.5564066
K1   17.2540  9.0177 0.0644121      0.3558741
e1    0.3455  0.2273 0.0016233      0.0032398
w1    0.4914  0.3860 0.0027574      0.0053472
X1    0.1216  0.0661 0.0004722      0.0007065
s     5.5143  1.0917 0.0077978      0.0082492

2. Quantiles for each variable:

        2.5%       25%       50%       75%     97.5%
V   -4.17206   -2.0062   -0.9871 5.018e-02    2.2204
T1 974.80410 1035.2120 1066.1112 1.097e+03 1154.6832
K1  11.58791   14.1668   15.6690 1.764e+01   30.2919
e1   0.01861    0.1540    0.3141 5.106e-01    0.8291
w1   0.01865    0.1890    0.3983 7.120e-01    1.4384
X1   0.01064    0.0745    0.1164 1.612e-01    0.2699
s    3.56420    4.7563    5.4444 6.187e+00    7.8476

> █
```

Figure 3: Summary of the estimates of orbital parameters in the radial velocity model.

- **Step 4:** Plot the densities of orbital parameters in the model. Type:

[3] You may have to scroll up a bit to see all of them.

```
densplot.model()
```

The output is shown in Figure[4].

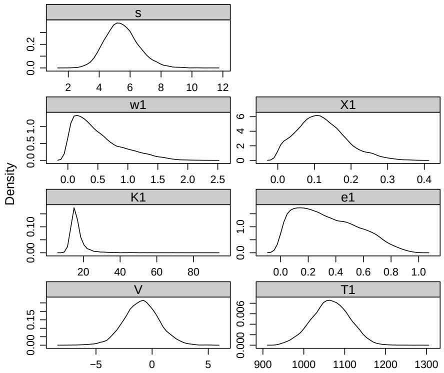


Figure 4: Density plots for orbital parameters of HD187085

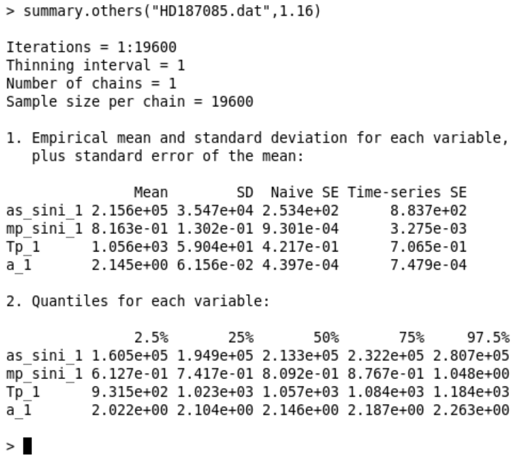
- **Step 5:**Display the estimates of the other useful parameters. Type:

  ```
  summary.others(‘‘RVdata’’,mass_of_the_star)
  ```

  This command needs the name of the data set and the mass of the star as the input. For example, assuming that the mass of HD187085 is $1.16M\odot$, type:

  ```
  summary.others(‘‘HD187085.dat’’,1.16)
  ```

  The output is shown in Figure[5].



```
> summary.others("HD187085.dat",1.16)

Iterations = 1:19600
Thinning interval = 1
Number of chains = 1
Sample size per chain = 19600

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

               Mean        SD  Naive SE Time-series SE
as_sini_1 2.156e+05 3.547e+04 2.534e+02      8.837e+02
mp_sini_1 8.163e-01 1.302e-01 9.301e-04      3.275e-03
Tp_1      1.056e+03 5.904e+01 4.217e-01      7.065e-01
a_1       2.145e+00 6.156e-02 4.397e-04      7.479e-04

2. Quantiles for each variable:

               2.5%       25%       50%       75%     97.5%
as_sini_1 1.605e+05 1.949e+05 2.133e+05 2.322e+05 2.807e+05
mp_sini_1 6.127e-01 7.417e-01 8.092e-01 8.767e-01 1.048e+00
Tp_1      9.315e+02 1.023e+03 1.057e+03 1.084e+03 1.184e+03
a_1       2.022e+00 2.104e+00 2.146e+00 2.187e+00 2.263e+00

> █
```

Figure 5: Summary of the estimates of other useful quantities.

- **Step 6** Plot the densities of other useful quantities as shown in the Figure[6] by

  ```
  densplot.others(‘‘RVdata’’,mass_of_the_star)
  ```

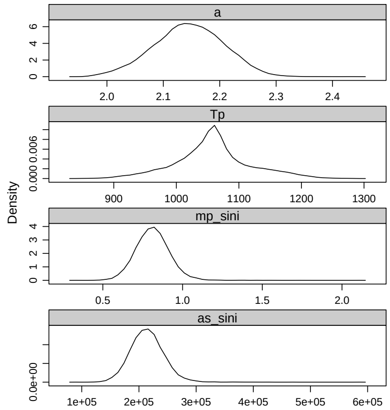


Figure 6: Density plots of other useful astronomical quantities for HD187085

  To plot the posterior distribution of both the model parameters and other useful astronomical quantities together, type:

  ```
  densplot.all(‘‘RVdata’’,mass_of_the_star)
  ```

  The output is show in Figure [7]

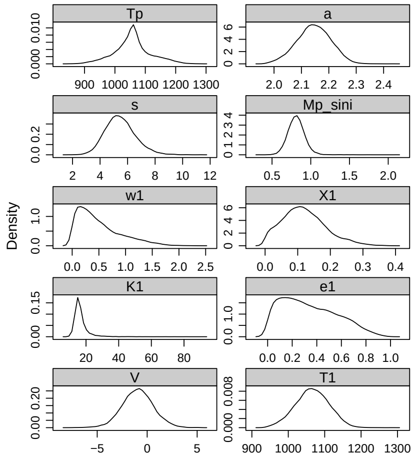


Figure 7: Density plots of model parameters and other useful astronomical quantities for HD187085

- **Step 7:** Finally plot the radial velocity curve along with the radial velocity data. Again you need to specify the radial velocity data file

  ```
  orbit.plot(‘‘HD187085.dat’’)
  ```

  The output is shown in Figure[8]

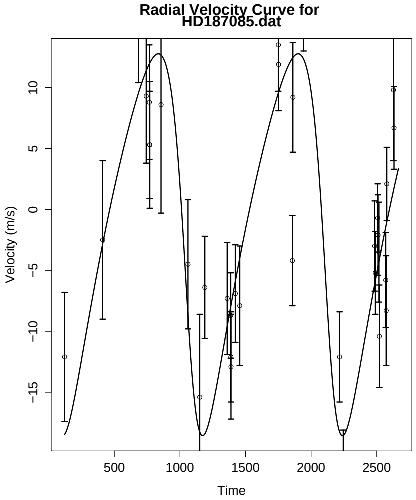


Figure 8: Keplerian orbit fitted onto the radial velocity data of HD187085 using a single planet model

In order to create a postscript of the plots in R use the following command

```
dev.print(device=postscript,‘‘filename’’)
```

The two planet model is automatically picked by the *orbit_plot.R* script by checking for the number of parameters in the *extract.dat* file. Summaries and density plots similar to the 1-planet model are made for 2-planet model by following same steps as above (i.e no new commands are needed).

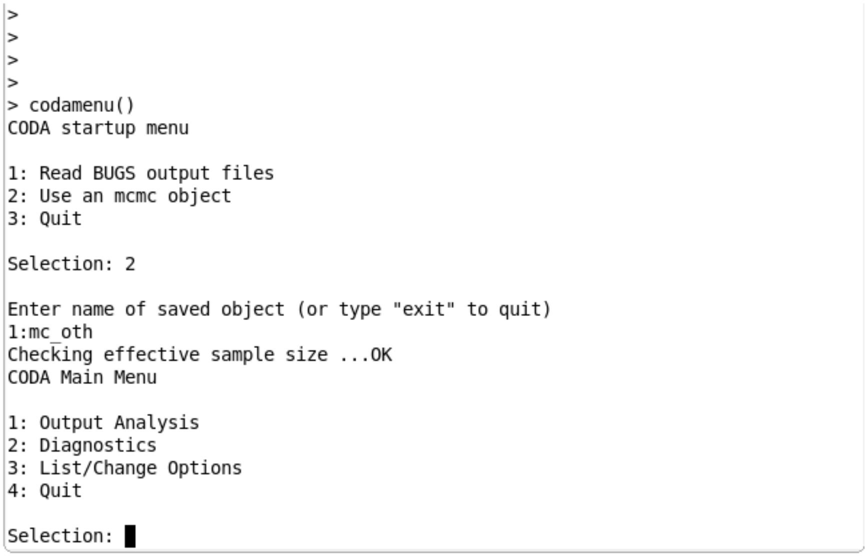
6.2 Convergence Tests

Convergence of Markov Chain is an important aspect of any MCMC method. It is very difficult to tell whether a chain has converged to its stationary distribution. On the other hand it might be reasonably easy to tell whether a chain has not converged. There are packages in R which deals with convergence diagnostics of Markov Chains. These packages offers a variety of tests to check the convergence of Markov Chains. We use *coda* for our analysis. It has an interactive menu to navigate between various tests and other statistical tools available in the package.

In order to analyse the Markov Chain using *coda* one need to create an MCMC object. The R script creates an object called *mc*. This MCMC object corresponds to parameters in the radial velocity model. They are the input to *coda*. To start *coda* in R session type:

```
codamenu()
```

This will start a coda session which looks like Figure[9]. Choose the option 2 in the menu and enter *mc* as the input object. *coda* will check for the *effective step size* first. If it is fine you may proceed further to look at the traceplots and convergence diagnostics.



```
>
>
>
>
> codamenu()
CODA startup menu

1: Read BUGS output files
2: Use an mcmc object
3: Quit

Selection: 2

Enter name of saved object (or type "exit" to quit)
1:mc_oth
Checking effective sample size ...OK
CODA Main Menu

1: Output Analysis
2: Diagnostics
3: List/Change Options
4: Quit

Selection: █
```

Figure 9: Coda menu: Choose option 2

7 Changing Priors

Changing priors involves re-compiling *ExoFit*. The prior densities on each parameter are defined in the source file `ExoFit.vX.XX/src/mcmc.cc` as shown below.

```
//add bonds
mcstate.add_bond(sys_velocity,&uniform);
mcstate.add_bond(period_1,&jeffreys);
mcstate.add_bond(amplitude_1,&mod_jeff);
mcstate.add_bond(eccentricity_1,&uniform);
mcstate.add_bond(long_periastron_1,&uniform);
mcstate.add_bond(periastron_pass_1,&uniform);
mcstate.add_bond(noise_factor,&mod_jeff);
```

Where `jeffreys` represents a Jeffreys prior, `mod_jeff` represents a modified Jeffreys prior with a break at 1.0 and `uniform` represents a uniform prior distribution. For example to change the prior distribution of parameter `period` from

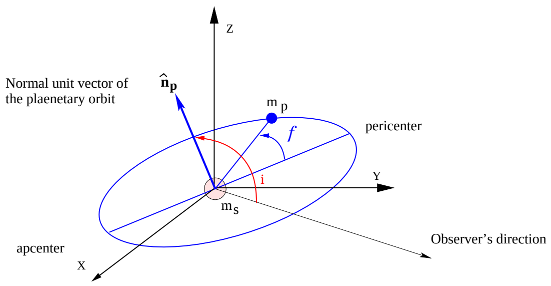


Figure 10: Inclination of the orbital plane with the reference plane. The angle i is defined as the angle between the direction normal to the orbital plane and the observer's line of sight.

`jeffreys` to `uniform`, simply modify the corresponding line to

```
mcstate.add_bond(period,&uniform);
```

Then make the executable binaries using the command `make` as explained in the Section[2].

Appendices

A Modelling of Radial Velocity

A.1 Doppler Spectrography

Planets are many times fainter than their host stars because they shine only by reflecting the star light. This makes their direct imaging extremely difficult. However, the gravitational pull of the planet makes the star wobble and this produces measurable periodic shifts in the apparent speed of the parent star. The motion of the star around the centre of mass causes the observed spectrum of the star to be Doppler shifted according its radial velocity, i.e. the velocity along the line of sight of the observer. This is measured over a course of time to obtain the radial velocity data along with the measurement uncertainties.

A.2 Radial Velocity of Star

A single planet model is assumed here to analyse the radial velocity data. Referring to Fig. 10 the radial velocity of a star can be written as [21, 23]

$$v_i = V - \frac{m_p}{m_s + m_p} \frac{na \sin i}{\sqrt{1 - e^2}} \big(\sin(f_i + \varpi) + e \sin \varpi \big) \,, \qquad (1)$$

where v_i is the ith radial velocity entry corresponding to time coordinate t_i and,
V = the systematic velocity of the system,
m_p = the mass of the planet,
m_s = the mass of the star,
$n = \frac{2\pi}{T}$ the mean motion and T is orbital period of planet,
a = the length of the semi-major axis of the planet,

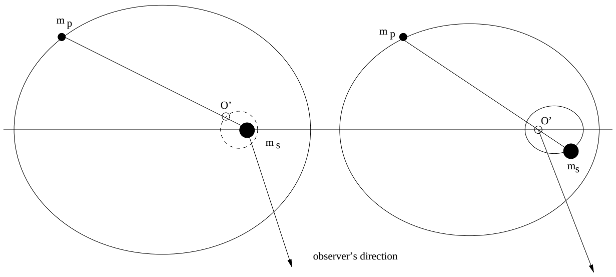


Figure 11: Figure on the left shows the motion of the planet around the star. Figure on the right shows the relative motion of the planet and the star around the centre of mass.

$i =$ the inclination of the orbital plane with the ecliptic,
$e =$ the eccentricity of the planet,
$f_i =$ the true anomaly at time t_i and
$\varpi =$ the longitude of periastron.

Notice that the equation does not contain the time explicitly. On the other hand radial velocity is a function of *true anomaly* which is given by

$$\cos f_i = \frac{\cos E_i - e}{1 - e\cos E_i}\,, \tag{2}$$

where E_i is the *eccentric anomaly* at the instant t_i given by the *Kepler's equation*

$$M_i = E - e\,\sin E\,. \tag{3}$$

Finally we have an equation that contains time explicitly. In the above equation M_i is the *mean anomaly* which can be written as

$$M_i = n(t_i + \tau) = \frac{2\pi}{T}(t_i + \tau)\,, \tag{4}$$

where τ is the time of pericenter passage. For the computational purpose we define *periastron passage factor* χ as the fraction of orbit prior to the start of data-taking that periastron occurred [15]. In other words $\chi T =$ the number of days prior to $t_i = 0$ that the star was at periastron for an orbital period of T days.

Motion of the star as well as the planet can be described by the same equation that describes their relative motion but reduced in scale by a factor of either $m_s/(m_s+m_p)$ or $m_p/(m_s+m_p)$. Let a_s and a_p represent the length of the semi-major axis of the motion star and planet around the centre of mass respectively. Since $a = a_s + a_p$ we have,

$$a_s = \frac{m_p}{m_s + m_p}\,a\,. \tag{5}$$

Substituting this result into equation (1), we obtain

$$v_i = V - K\left(\sin(f_i + \varpi) + e\,\sin\varpi\right), \tag{6}$$

where

$$K = \frac{2\pi}{T}\frac{a_s\,\sin\,i}{\sqrt{1-e^2}}\,. \tag{7}$$

The above relations lead to other parameters of the orbit. The length of the semi-major axis a and mass of the planet $m_p \sin i$ are calculated as follows:

$$
\begin{aligned}
a_s \sin i &= \frac{K\,T\,\sqrt{1-e^2}}{2\pi}\,, && (8) \\
m_p \sin i &\approx \frac{K\,m_s^{\frac{2}{3}}\,T^{\frac{1}{3}}\sqrt{1-e^2}}{2\pi\,G} \quad \text{and} && (9) \\
a &\approx \frac{m_s a_s \sin i}{m_p \sin i}\,. && (10)
\end{aligned}
$$

A.3 Radial velocity data

According to (http://exoplanet.eu) eighteen different radial velocity search programmes are looking for extrasolar planets. Majority of the contributions come from Keck, Lick and Anglo-Australian observatories (the California & Carnegie and Anglo-Australian planet searches) and searches based at l'Observatoire de Haute Provence and La Silla Observatory (the Geneva extrasolar planet search). Radial velocity data for a star consists of time of observation t_i, measured radial velocity v_i and uncertainty associated with each measurement e_i. These uncertainties are a characteristic of the instruments used for measurements. The precision of these instruments have improved from the order of $10ms^{-1}$in 1994 to order of $1ms^{-1}$ [7, 25] at present.[4] This is extremely significant for finding low mass companions as well as planets with large a s.

B Bayesian Retrieval of orbital parameters

The extraction of orbital parameters from the radial velocity data poses considerable statistical challenges. Traditional methods methods first search for periodicity in the observed data using a Lomb-Scargele periodogram and then proceed to fix the other parameters by Levenberg-Marquardt method. Studies by [8] and [9] have identified two cases where these methods become inefficient in accurately characterising the orbital elements:

1. When the orbital period is extremely short and the eccentricity is high.
2. When the duration of observation does not span at least a single orbital phase.

Since the transit probability of the planet increases for short periods, the orbital parameters predicted by the periodogram method can be verified with the help of transit photometry. Incomplete radial velocity data gives rise to a multitude of orbital solutions which is referred to as parameter degeneracy. Higher eccentricities make the radial velocity curve less sinusoidal.[24] makes use of a 2DKLS periodogram to incorporate the effect of eccentricity of the orbits while searching for orbital periods. Recently Bayesian techniques have been employed by [14], [10] and [11] to retrieve the orbital parameters of extrasolar planets. The results show that Bayesian methods tackle the difficulties associated with the traditional methods efficiently and transparently.

[4] The measurement method and its uncertainties are discussed in the corresponding planet discovery papers

B.1 The Bayesian method

The starting point of any Bayesian analysis is Bayes' theorem [2]. Let $\mathbf{y} = (y_1, \ldots, y_i, \ldots, y_n)$ be a vector of n observations whose probability distribution $p(\mathbf{y}|\boldsymbol{\theta}, H)$ is conditional on k parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_i, \ldots, \theta_k)$, where H represents the background information or the hypothesis by which the probability statements are made. Suppose that the parameter $\boldsymbol{\theta}$ has the probability distribution $p(\theta|H)$. Then, Bayes' theorem says

$$p(\boldsymbol{\theta}|\mathbf{y}, H) = \frac{p(\mathbf{y}|\boldsymbol{\theta}, H)}{p(\mathbf{y}|H)}\,. \tag{11}$$

For a continuous $\boldsymbol{\theta}$ we can write

$$p(\mathbf{y}|H) = \int p(\mathbf{y}|\boldsymbol{\theta}, H)\, p(\boldsymbol{\theta}|H)\, d\boldsymbol{\theta}\,, \tag{12}$$

which is constant for given $\mathbf{y}$ and a probability distribution $p(\boldsymbol{\theta}|H)$. Then Equation [11] can be rewritten as

$$p(\boldsymbol{\theta}|\mathbf{y}, H) = C\, p(\mathbf{y}|\boldsymbol{\theta}, H)\, p(\boldsymbol{\theta}|H)\,. \tag{13}$$

In the above equation $p(\boldsymbol{\theta}|H)$ is called *prior distribution* of $\boldsymbol{\theta}$ since it conveys our knowledge about $\boldsymbol{\theta}$ before the data has been observed. Correspondingly, $p(\boldsymbol{\theta}|\mathbf{y}, H)$ is known as the *posterior distribution* of $\boldsymbol{\theta}$ given $\mathbf{y}$. The factor C is a normalising constant which ensures that the posterior distribution integrates to one. We call $p(\mathbf{y}|\boldsymbol{\theta}, H)$ the likelihood function of $\boldsymbol{\theta}$ since $p(\mathbf{y}|\boldsymbol{\theta}, H)$ can be considered as a function of $\boldsymbol{\theta}$ instead of $\mathbf{y}$. Then,

$$p(\boldsymbol{\theta}|\mathbf{y}, H) \propto p(\mathbf{y}|\boldsymbol{\theta}, H)\, p(\boldsymbol{\theta}|H)\,. \tag{14}$$

Statistical inferences regarding $\boldsymbol{\theta}$ are derived from the posterior distribution of $\boldsymbol{\theta}$. The posterior distribution encapsulates all information about unknown quantities $\boldsymbol{\theta}$ following the observation of the data $\mathbf{y}$.

The principal steps in the Bayesian method can be summarised as follows [22].

- **Likelihood:** Find out the Likelihood function $p(\mathbf{y}|\boldsymbol{\theta}, H)$. This is the process of describing the observed data in terms of a chosen set of parameters $\boldsymbol{\theta}$.

- **Prior:** Obtain the prior density $p(\boldsymbol{\theta}|H)$. This is the statement of our knowledge about the unknown parameters before the observation of data.

- **Posterior:** Apply Bayes' theorem to derive the posterior probability distribution $p(\boldsymbol{\theta}|\mathbf{y}, H)$. This describes our knowledge about $\boldsymbol{\theta}$ after observing the data.

- **Inference:** Make Appropriate inference statements. These are derived from the posterior distribution and include point estimates as well as interval estimates.

B.2 Likelihood function

Let d_i represent the measured radial velocity data for the ith instant of time t_i. Observed radial velocity data can be modelled by the equation [14]

$$d_i = v_i + e_i + \epsilon\,, \tag{15}$$

where e_i is the uncertainty component arising from accountable but unequal measurement errors which are assumed to be normally distributed. The term ϵ explains any unknown measurement errors. There can be multiple reasons for the presence of this uncertainty component [7]. For example this could be the result of another planet in the system or caused by the intrinsic anomalies in the star spectrum due to the irregularities on the surface of the star [25, 19, 5]. Thus any noise component that cannot be modelled is described by the term ϵ. The probability distribution of ϵ is chosen to be a Gaussian distribution with finite variance s^2. Therefore the combination of uncertainties $e_i + \epsilon$ has a Gaussian distribution with a variance equal to $\sigma_i^2 + s^2$.

The radial velocity v_i predicted by the mathematical model at an instant t_i is given by the equation (6) as,

$$v_i = V - K\left(\sin(f_i + \varpi) + e\,\sin\varpi\right).$$

Six model parameters namely $T, K, V, e, w,$and χ, as defined in the section (A.2) are used to fit the above equation onto a given radial velocity data.

Each error term e_i in equation (15) is independent. Since they are assumed to follow a Gaussian distribution, the likelihood function is product of N Gaussians [15, 14] where N is the number of observations. Thus

$$p(\mathbf{y}|\boldsymbol{\theta}) = A\,\exp\left[-\sum_{i=1}^{N}\frac{(d_i - v_i)^2}{2(\sigma_i^2 + s^2)}\right], \tag{16}$$

where

$$A = (2\pi)^{-N/2}\left[\prod_{i=1}^{N}\left(\sigma_i^2 + s^2\right)^{-1/2}\right]. \tag{17}$$

and s becomes the seventh parameter in our probability model.

B.3 Choice of Priors

The choice of priors is extremely important in the Bayesian analysis as senseless choice of priors can produce to misleading results. As we have mentioned in the last section a set of priors which can be described as *reference priors* has to be found out. The priors for our problem are chosen in such a way that

1. these reference priors should remain ‘neutral’ or they should ensure that the information gleaned from the data may be allowed to dominate their densities [6].

2. they should convey the known physical aspects of the system unambiguously.

Physical and geometric conditions govern the selection of prior distributions for most of the parameters. Since $\boldsymbol{\theta} = (T, K, V, e, \varpi, \chi, s)$ the prior distribution in our problem can be written as

$$\begin{aligned} p(\boldsymbol{\theta}|H) \quad = \quad & p(T|H)\, p(K|H)\, p(V|H) \\ & p(e|H)\, p(\varpi|H)\, p(\chi|H)\, p(s|H)\,, \end{aligned} \tag{18}$$

on the assumption that they are independent. We will discuss how the above conditions are met for our choice of prior for each parameter in the next few sections.

The sampling of the radial velocity data in most of the cases is highly non-uniform. The sparse sampling makes the retrieval of the orbital period more challenging because a number of orbital periods might yield the same fit. The radial velocity scatter diagram is analysed first for some initial hints on the range of possible orbital periods. For example there are cases in which we observe a clear periodicity in the scatter diagram and hence we can set an upper limit for orbital period. [14] sets the upper limit of the orbital period to be three times the duration of observation. [11] consider 10^3 years as the upper limit for the orbital period. The theoretical limit (Roche limit) for a planet with 10 times the mass of Jupiter orbiting a star with mass equal to that of sun will be 0.2 days. We choose a Jeffreys prior for the orbital period. Therefore

$$p(T) = \frac{1}{T\, \ln\left(\frac{T_{max}}{T_{min}}\right)}\,. \tag{19}$$

From the current mass distribution of extra-solar planets we can set the upper bound $K_{max} = 2129\ ms^{-1}$ which corresponds to a maximum planet to star mass ratio of 0.01. Since the lower bound includes zero we use a modified Jeffreys prior for K [11] given by

$$p(K) = \frac{1}{(K + K_0)\ln\left(\frac{K_0 + K_{max}}{K_0}\right)}\,. \tag{20}$$

For $K \ll K_0$, $p(K)$ behaves like a uniform prior and for $K \gg K_0$ it behaves like a Jeffreys prior. The factor K_0 could be thought as a lower bound in the Jeffreys prior and we choose $K_0 = 1\ ms^{-1}$ which corresponds to the smallest detectable velocity at present. The choice of prior for K becomes significant only when the planet detection is marginal. However, if the posterior distribution has a probability peak near K_0 then we should re-analyse the inference by checking how sensitive is the posterior to the value of K_0.

The systematic velocity is attributed to the absolute motion of the star through the space. Since any value in the range $(0, \infty)$ is possible in this case, we may choose a uniform improper prior for V [15]. Thus,

$$p(V) = \frac{1}{V_{max} - V_{min}}\,. \tag{21}$$

Eccentricity of the orbit can have any value between 0 and 1 excluding 1, since $e = 1$ corresponds to a parabolic orbit. Therefore we choose a uniform prior for e given by

$$p(e) = 1\,. \tag{22}$$

Table 4:
The assumed prior distribution of various parameters and their boundaries. It is similar to choice of priors given by [11], except for the prior distribution of K.

Para.	Prior	Mathematical Form	Min	Max
$T(days)$	Jeffreys	$\frac{1}{T\,\ln\left(\frac{T_{max}}{T_{min}}\right)}$	0.2	15000
$K(ms^{-1})$	Mod. Jeffreys	$\frac{(K+K_0)^{-1}}{\ln\left(\frac{K_0+K_{max}}{K_0}\right)}$	0.0	2000
$V(ms^{-1})$	Uniform	$\frac{1}{V_{max}-V_{min}}$	-2000	2000
e	Uniform	1	0	1
ϖ	Uniform	$\frac{1}{2\pi}$	0	2π
χ	Uniform	1	0	1
$s(ms^{-1})$	Mod. Jeffreys	$\frac{(s+s_0)^{-1}}{\ln\left(\frac{s_0+s_{max}}{s_0}\right)}$	0	2000

Longitude of periastron can have any value in the range $[0, 1]$. We choose a uniform prior for ϖ with a distribution The definition of ϖ in our formulation has a difference of $\pi/2$ with the traditional formalism. Please see the figure [10] to differentiate between these two definitions.

$$p(\varpi) = \frac{1}{2\pi}\,. \tag{23}$$

The periastron passage time is measured as fraction of the given orbital period. It can have any value between 0 and 1. A uniform prior is selected with boundaries 0 and 1. Thus,

$$p(\chi) = 1\,. \tag{24}$$

Our choice is a modified Jeffreys prior. The upper bound for s is taken to be K_{max}. The minimum possible value for s is 0 and therefore following [11], we choose a break s_0 at $1ms^{-1}$. Hence we have,

$$p(s) = \frac{1}{(s+s_0)\,\ln\left(\frac{s_0+s_{max}}{s_0}\right)}\,. \tag{25}$$

Table 4 shows the choice of priors for each parameter and their boundaries.

B.4 2-planet Model

ExoFit has an option to search for two planets in the radial velocity data. We choose the probability model to be similar to that of the single planet model explained in Section B.2. The observed radial velocity data is again modelled by the equation 15. The radial velocity v_i predicted by the mathematical model at an instant t_i is given by

$$\begin{aligned} v_i = V - \Big(& K_1\big(\sin(f_{i1}+\varpi_1) + e_1\,\sin\varpi_1\big) \\ & + K_2\big(\sin(f_{i2}+\varpi_2) + e_2\,\sin\varpi_2\big)\Big)\,. \end{aligned} \tag{26}$$

Para.	Prior	Mathematical Form	Min	Max
$V(ms^{-1})$	Uniform	$\frac{1}{V_max-V_{min}}$	-2000	2000
$T_1(days)$	Jeffreys	$\frac{1}{T_1\ \ln\left(\frac{T_{1\,max}}{T_{1\,min}}\right)}$	0.2	15000
$K_1(ms^{-1})$	Mod. Jeffreys	$\frac{(K_1+K_{1\,0})^{-1}}{\ln\left(\frac{K_{1\,0}+K_{1\,max}}{K_{1\,0}}\right)}$	0.0	2000
e_1	Uniform	1	0	1
ϖ_1	Uniform	$\frac{1}{2\pi}$	0	2π
χ_1	Uniform	1	0	1
$T_2(days)$	Jeffreys	$\frac{1}{T_2\ \ln\left(\frac{T_{2\,max}}{T_{2\,min}}\right)}$	0.2	15000
$K_2(ms^{-1})$	Mod. Jeffreys	$\frac{(K_2+K_{2\,0})^{-1}}{\ln\left(\frac{K_{2\,0}+K_{2\,max}}{K_{2\,0}}\right)}$	0.0	2000
e_2	Uniform	1	0	1
ϖ_2	Uniform	$\frac{1}{2\pi}$	0	2π
χ_2	Uniform	1	0	1
$s(ms^{-1})$	Mod. Jeffreys	$\frac{(s+s_0)^{-1}}{\ln\left(\frac{s_0+s_{max}}{s_0}\right)}$	0	2000

Table 5: The assumed prior distribution of orbital parameters and their boundaries for a 2-planet model. The boundaries for $K2$ can be made smaller in order to speed up the convergence of the Markov Chain.

11 parameters $\{V, T_1, K_1, e_1, w_1, \chi_1, T_2, K_2, e_2, w_2, \chi_2\}$[5], as defined in the Section A.2 are used to fit the above equation onto the radial velocity data. The likelihood function is again given by equations 16 and 17 respectively, and s becomes the 12^{th} parameter in our probability model. The choice of prior distributions for each of these parameters is given in Table 5.

B.5 Posterior Distribution

Posterior distribution is obtained by applying the Bayes' theorem given by the equation (13). This is the output of a Bayesian analysis and it summarises all that we know about the physical system after the observation of data subject our prior believes. Having obtained the posterior density we need to derive suitable inference statements about the quantities. In the present case, we need to state what the posterior distribution has to say about the orbital parameters of the planet. The objective is to extract the information concerning $\boldsymbol{\theta}$ and describe it via effective summary statements. The marginal posterior densities of all parameters present the complete summary in the Bayesian analysis. Useful and interesting features of the posterior distribution should be identified before making summary statements. For example, the posterior distribution may be unimodal but asymmetric or it can be multi-modal with many probability peaks.

Any summary statistic can be expressed in terms of posterior expectations of $\boldsymbol{\theta}$ [12, 4]. The posterior expectation of a function $f(\boldsymbol{\theta})$ can be written as:

$$E[f(\boldsymbol{\theta}|\mathbf{y}, H)] = \frac{\int f(\boldsymbol{\theta}|H)p(\boldsymbol{\theta}|H)\,p(\mathbf{y}|\boldsymbol{\theta}, H)\,d\boldsymbol{\theta}}{\int p(\boldsymbol{\theta}|H)\,p(\mathbf{y}|\boldsymbol{\theta}, H)\,d\boldsymbol{\theta}}\,. \tag{27}$$

[5] Subscripts 1 and 2 indicate planets 1 and 2 respectively.

The multi-dimensional integral in the above equation is one of the key issues in Bayesian inference because the evaluation of such integrals by analytical methods is nearly impossible. Therefore, in most occasions numerical methods are employed.

C MCMC Implementation

Difficulty in evaluating the multi-dimensional integrals is an inherent inability of any Bayesian formulation. Many techniques have been developed in the last 25 years to approximate the integral in equation (27). Simulation methods dominate this area and several computational algorithms were developed to numerically integrate the posterior distribution in order to find out the marginal distributions of each parameter. According to [3] the abundance of computational power has produced a paradigm shift with respect to statistics: Computationally intensive but conceptually simple methods are preferred. Markov Chain Monte Carlo (MCMC) method is one of the most commonly used methods for simulating complex probability distributions. The method is explained in Section C.1 with respect to a general form given by *Metropolis-Hastings* [20, 16] algorithm.

C.1 Metropolis Hastings algorithm

The present section is based on the explanations given by [14] and [12]. Metropolis-Hastings does a random walk through the model parameter space such that the number of samples in particular region is proportional to the posterior density. In order explain the Metropolis Hastings algorithm we define two terms

1. *Target Distribution:* This is the same as the posterior distribution given by equation (13).

2. *Proposal Distribution q:* The samples $\boldsymbol{\theta}_t$ are drawn from this probability distribution

The random walk is achieved through a Markov Chain. At each iteration t the next state $\boldsymbol{\theta}_{t+1} = \boldsymbol{\phi}$ is chosen by first sampling a *candidate* from a proposal distribution $q(\boldsymbol{\phi}|\boldsymbol{\theta})$. The proposal distribution is chosen in such a manner that it is easy to evaluate and is centred on the current state $\boldsymbol{\theta}$. Most widely used choice for a proposal distribution is a multi-variate Gaussian distribution which has the desired property that away from the current sample the probability density decreases [14]. The new sample $\boldsymbol{\phi}$ is accepted with a probability r given by,

$$r = MIN\left[1, \frac{p(\boldsymbol{\phi})\, q(\boldsymbol{\theta}|\boldsymbol{\phi})}{p(\boldsymbol{\theta})\, q(\boldsymbol{\phi}|\boldsymbol{\theta})}\right] \tag{28}$$

where $q(\boldsymbol{\phi}|\theta) = q(\boldsymbol{\theta}|\boldsymbol{\phi})$ for a symmetrical proposal distribution. If the proposal is not accepted the Markov chain remains in the same state. The above process can be summarised as follows:

```
Initialise $\boldsymbol{\theta}_0$; set $t = 0$,
Repeat{
Sample a proposal $\boldsymbol{\phi}$ from $q(\boldsymbol{\phi}|\boldsymbol{\theta})$
Sample a Uniform(0,1) random variable $U$
If $U \leq r$ set $\boldsymbol{\theta}_{t+1} = \boldsymbol{\phi}$
Else set $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t$
Increment $t$
}
```

D The *ExoFit* software package

Bayesian MCMC methods have gained popularity in various areas of astrophysics, for example in multi-parameter estimation from cosmological data sets (e.g. CosmoMC; [18]). From a Bayesian point of view analysis of statistical problems requires an efficient tool for simulating posterior densities and MCMC methods are ideally suited for this purpose. In general the radial velocity of an n-planet model could be approximated as linear combination of n single planet radial velocities. Even though we consider a single planet model for the present analysis, it would be ideal to design the code in such a manner that the extension to multi-planet problem could be achieved without a huge effort.

ExoFit is a step towards achieving the goals mentioned above. It should be considered as a platform to develop MCMC based methods for estimating orbital parameters of a generalised multi-planet model. Object oriented design of *ExoFit* makes it better suited for extending the analysis to multi-planet systems with prior constraints on several orbital parameters such as eccentricity and length of semi-major axis. Following [13], our implementation MCMC consists of the following parts.

1. Data
2. State
3. Bond
4. Update

They are referred to as objects in object oriented analysis. *data* handles the input data into the MCMC analysis. A *state* consists of a set of parameters whose posterior distribution is sought. The parameter values at a particular instant defines the *state* of Markov Chain in the analysis. The parameters in a particular *state* are connected to each other by a *bond*. It consists of prior densities and likelihood. For each state there corresponds a bond strength which is equal to $prior \times likelihood$. In other words it is the posterior density without the normalisation constant in Bayes theorem. An *update* selects the parameters that should be updated at particular iteration. New values for the parameters are proposed according to the *update* defined and the new *bond* strength is then calculated for the proposed state. The new state is accepted or rejected according Metropolis-Hastings method.

The central concept of this approach is that, the MCMC engine remains the same and need not be re-implemented whenever the probability model gets changed. We also take advantage of the commonalities among the different components of MCMC. As an example we notice that, for each MCMC parameter in the Bayesian analysis, we need to specify at least three values.

1. Lower bound

2. Upper bound

3. Step size

Our implementation works for variety or prior distributions and *Update* methods. The only component that requires to be changed is the likelihood function.

References

[1] S. T Balan and O. Lahav. *ExoFit*: Orbital parameters of extra-solar planets from radial velocities. *MNRAS*, 2008.

[2] T. Bayes. An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, f. r. s. communicated by mr. price, in a letter to john canton, a. m. f. r. s. *Philosophical Transactions of the Royal Society*, 53:–370, 1763.

[3] B. A. Berg. *Markov Chain Monte Carlo Simulations And Their Statistical Analysis: With Web-based Fortran Code*. World Scientific, 2004.

[4] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag New York Inc, 175 Fifth Avenue, New York, NY10010, USA, 2 edition, 1980.

[5] F. Bouchy et al. Astroseismology of the planet-hosting star μ arae. *åp*, 440, 2005.

[6] G. P. Box and C. Tiao. *Bayesian Inference in Statistical Analysis*. Addison-Wesley Publishing Company, 1973.

[7] R. P. Butler et al. Catalog of nearby exoplanets. *Atrophysical Jounal*, 646, 2006.

[8] A. Cumming. Detectability of extrasolar planets in radial velocity. *Monthly Notices of the Royal Astronomical Society*, 354, 2004.

[9] A. Cumming, G. W. Marcy, and R. P. Butler. The lick planet search: Detectability and mass thresholds. *Astrophysical Journal*, 526, 1999.

[10] E. B. Ford. Quantifying the uncertainty in the orbits of extrasolar planets. *Astronomical Journal*, 129, 2005.

[11] E. B. Ford and P. C. Gregory. Bayesian model selection and extrasolar planet detection. *ASP Conference Series*, 371, 2007.

[12] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice*. Chapman & Hall London, 1996.

[13] Todd L Graves. Design ideas for markov chain monte carlo software. *Journal of Computational & Graphical Statistics*, 16(1):24–43, 2007.

[14] P. C. Gregory. A bayesian analysis of extrasolar planet data for hd 73526. *Astrophysical Journal*, 631, 2005.

[15] P. C. Gregory. *Bayesian Logical Data Analysis for the Physical sciences: A comparitive Approch with "Mathematica" Support*. Cambridge: Cambridge University Press, 2005.

[16] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[17] H. R. A. Jones, R. P. Butler, C. G. Tinney, G. W. Marcy, B. D. Carter, A. J. Penny, C. McCarthy, and J. Bailey. High-eccentricity planets from the anglo-australian planet search. *Monthly Notices of the Royal Astronomical Society*, 369(1):249–256, 2006.

[18] A. Lewis and S. Bridle. Cosmological parameters from cmb and other data: A monte carlo approach. *Physical Review D*, 66(10):103511, 2002.

[19] M. Mayor, F. Pepe, and D. Queloz et al. Setting new standards with harps. *The Messenger*, 114, 2003.

[20] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, Teller A. H., and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[21] C. D. Murray and S. F. Dermott. *Solar System Dynamics*. Cambridge University Press, 2000.

[22] A. O'Hagan and J. Forster. *Bayesian Inference*, volume 2B of *Kendall's Advanced Theory of Statistics*. Oxford University Press Inc, 198 Madison Avenue, New York, NY10016, 2 edition, 2004.

[23] Y. Ohta, A. Taruya, and Y. Suto. The rossiter-mclaughlin effect and analytic radial velocity curves for transiting extrasolar planetary systems. *Astrophysical Journal*, 622, 2005.

[24] 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. *The Astrophysical Journal*, 660(2):1636–1641, 2007.

[25] F. Pepe, M. Mayor, and D. Queloz et al. The harps search for southern extra-solar planets. i. hd 330075 b: A new "hot jupiter". *åp*, 423, 2004.