Make your own MCMC sampler

Application: Estimating the mass of exoplanets

Jean-Baptiste Delisle (jean-baptiste.delisle@unige.ch) Nathan Hara

November 16, 2021

1 Context

When a star has planetary companions, it describes an epicyclic motion around the center of mass of the system. This motion can be used to indirectly unveil the presence of the companions. In particular, the radial velocity method uses the Doppler effect, which induces a shift of the spectrum of the star depending on its instantaneous velocity in the direction of the line of sight. By acquiring stellar spectra at different times and measuring how the spectral absorption lines shift in wavelength, one can obtain a time series of the star radial (i.e., in the direction of the line of sight) velocity, and search for the periodic signature of planets in this time series.

You have already seen how to search for periodic signals in a time series using a periodogram approach. Once a candidate period is found, one need to determine whether the periodic signal originates from a planet. Model comparison methods (such as the ΔBIC criterion you already experimented) can be used to help answer this complex question.

2 Objectives

In this exercise, we assume that the planetary origin of the considered signal has already been validated, and that an estimate of the period has been found. We now want to determine all the planet's orbital elements, and especially its mass. To simplify the problem, we assume that the planet has a circular orbit (i.e., zero eccentricity).

The exercise is divided in three parts. We first implement the forward model, i.e., given all the physical parameters and the instrument's characteristics, what is the likelihood of the data? Then we compute the maximum likelihood estimates of the parameters, using the known period estimate. Finally, we compute the posterior distribution of the parameters, and interval estimates of the mass. This last part requires to code a Metropolis-Hastings algorithm, which is one of the simplest Markov Chain Monte Carlo (MCMC) algorithms.

3 The forward model

Let us suppose that a planet of mass m orbits a star of mass M at a semi-major axis a and eccentricity 0. We want to express the velocity of the **star** in the direction observer – barycenter (the radial velocity).

To facilitate the calculation of the radial velocity, we define two inertial frame: the observing frame (x, y, z) and the orbital frame (x', y', z'). Both frames have their origin O at the barycenter of the system (star + planet). The observing frame z axis is oriented in the direction observer – star. The plane perpendicular to z is called the sky plane. The orbital frame z' axis is perpendicular to the plane of the orbit and its direction is such that the **planet** evolves in the trigonometric (or counterclockwise) sense. The x and x' axes are pointing at the planet's ascending node, that is the point where the planet's orbit crosses the sky plane from negative to positive z values. The y and y' axes are such that (x, y, z) and (x', y', z') are direct orthonormal bases. The semi-major axis of the relative orbit (planet's orbit with respect to the star) is denoted by a. The angle between the z and z' axes is denoted by i (orbital inclination).

(a) Show that the position of the **star** in the orbital frame at time t is

$$r(t) = -a\frac{m}{M+m}\begin{pmatrix}\cos(n(t-t_0))\\ \sin(n(t-t_0))\\ 0\end{pmatrix}, \tag{1}$$

where t_0 is the time of passage at the ascending node and $n = \sqrt{\frac{G(m+M)}{a^3}} = \frac{2\pi}{P}$ is the mean-motion, and P is the orbital period.

(b) Show that the velocity of the **star** projected onto the z axis at time t is

$$V(t) = V_0 + \frac{(Gn)^{1/3}}{(m+M)^{2/3}} m \sin i \cos(n(t-t_0)), \tag{2}$$

where V_0 is the velocity of the system barycenter with respect to the observer.

(c) We assume that we have N radial velocity measurements taken at times $(t_k)_{k=1..N}$. We also assume that these measurements are affected by Gaussian white (i.e., uncorrelated) noise, and that the instrument errorbar of the k-th measurement is σ_k. Give the expression of the likelihood (or loglikelihood) function.

(d) In the rest of this exercise, we will not use directly the model of Eq. (2), but

$$V(t) = A\cos(nt) + B\sin(nt) + C. \tag{3}$$

What is the advantage of this latter formulation? Give the relations between the parameters A, B, C and the orbital/physical parameters of Eq. (2).

4 Point estimates (maximum likelihood)

We first look for point estimates of the parameters.

(a) We assume that the mean-motion n is known (thanks to a periodogram approach for instance). Give the expression of the maximum likelihood estimate of $\theta = (A, B, C)$.

(b) Load the data in the file rv_data.txt and compute the maximum likelihood values of the parameters θ assuming a period of 6.5 d for the planet.

5 Interval estimates and the Metropolis-Hastings algorithm

A common way to derive uncertainties on the parameters, as well as their correlations, is to compute their posterior distribution $p(\theta|y)$. In the following, we assume that θ belongs to an open subset Θ of $\mathbb{R}^p$ and that it has a prior density distribution $p(\theta)$. The density of the posterior distribution of the parameters is then

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}, \tag{4}$$

where, according to the total probability formula

$$p(y) = \int_{\theta\in\Theta} p(y|\theta)p(\theta)\mathrm{d}\theta. \tag{5}$$

In general, $p(\theta|y)$ does not have an analytical expression. To approximate it, paradoxically enough, it might be simpler to generate samples $(\theta_i)_{i=1}^{N_S}$ from $p(\theta|y)$ and to approximate the true distribution $p(\theta|y)$ from the empirical distribution of the samples. The most common way to generate such a sequence $(\theta_i)_{i=1}^{N_S}$ following the distribution $p(\theta|y)$ is to use a Markov Chain Monte Carlo (MCMC) method. In the following we will implement one of the simplest MCMC algorithm which is the Metropolis-Hastings (MH) algorithm. While more elaborated algorithms have been developed to improve the efficiency and reliability of the sampling, all the MCMC algorithms share the same fundamental principles, and coding your own MH algorithm is a good way to understand the basics of MCMC sampling.

5.1 Principle of the Metropolis-Hastings algorithm

The MH algorithm generates iteratively samples θ_i that follow the target distribution $f(\theta)$. Each new sample θ_{i+1} is generated based on the previous sample θ_i. The sequence of θ_i is called the chain. The MH algorithm requires to define a starting point θ_0 and a proposal distribution q. The proposal distribution (or transition probability) is a function

$$q : \theta \in \Theta \times \theta' \in \Theta \to q_\theta(\theta'), \tag{6}$$

such that for each $\theta \in \Theta$, $q_\theta(\theta')$ is a probability density (over the values of θ'). Then, assuming the state at iteration i is θ_i, the sample θ_{i+1} is generated following the steps:

1. Generate $\theta' \sim q_{\theta_i}$
2. Compute
$$\alpha = \min\left(1, \frac{f(\theta')q_{\theta'}(\theta_i)}{f(\theta_i)q_{\theta_i}(\theta')}\right) \tag{7}$$
3. Generate u with a uniform distribution between 0 and 1.
4. The new sample is then:
$$\theta_{i+1} = \begin{cases} \theta' \text{ if } u \geqslant \alpha \text{ (accept the proposal),} \\ \theta_i \text{ if } u < \alpha \text{ (reject the proposal).} \end{cases} \tag{8}$$

The fraction of iterations where the proposal is accepted is called the acceptance rate.

The principle of this procedure is that if $\theta_i \sim f$, then the new sample θ_{i+1} also follows the distribution f. The key to sampling correctly $f(\theta)$ is to choose an efficient proposal distribution q and a good starting point θ_0. If the proposal distribution is too wide, then the proposed values θ' will most of the time reach bad areas of the parameter space and the proposal will be rejected. If it is too narrow, then the states θ_i and θ_{i+1} will be very close to each other (the samples will be strongly correlated), and the exploration of the parameter space will be slow. A good rule of thumb is to aim at an acceptance rate of about 23%.

In practice, q is often chosen to be symmetric ($q_\theta(\theta') = q_{\theta'}(\theta)$), such that the factor α simplifies to $\alpha = \min(1, f(\theta')/f(\theta))$. More specifically, the distribution q_θ is often chosen to be a (multivariate) Gaussian distribution centered on θ and with a fixed covariance matrix (this automatically ensures the symmetry of the proposal).

5.2 Implementation

Our aim here is to generate samples from the target distribution $f(\theta) = p(\theta|y)$, with the same radial velocity data and model as for the previous point estimation (Sect. 4). However, we do not assume anymore that the period is known. The set of parameters is thus $\theta = (A, B, C, n)$.

(a) Express the criterion of Eq. (7) as a function of the prior probabilities $p(\theta)$, $p(\theta')$ and the likelihoods $p(y|\theta)$, $p(y|\theta')$. What is the crucial advantage of the Metropolis-Hastings algorithm for the generation of samples from a posterior distribution, over a generation of samples with the inverse CDF?

(b) For numerical stability, the factor α is not computed directly, but one computes its logarithm. Express the criterion of Eq. (7) as a function of the logarithms of the different distributions involved.

(c) We assume that the prior distribution on (A, B, C, n) is

$$p(A, B, C, n) = p(A)p(B)p(C)p(n), \tag{9}$$

where A, B and $C \sim G(0, \sigma^2)$ with $\sigma = 100$ m/s, and n follows a uniform distribution over $[0, 4\pi]$ rad/d. We use a proposal distribution of the form

$$q_{(A,B,C,n)}(A', B', C', n') = g(A - A', \sigma_A)g(B - B', \sigma_B)g(C - C', \sigma_C)g(n - n', \sigma_n). \tag{10}$$

where $g(x, \sigma) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}$. In this expression, σ_A, σ_B, σ_C and σ_n are to be tuned to find an efficient proposal distribution.

Write the logprior and loglikelihood functions. Write a generate_proposal function that takes the current state θ and generates a proposal θ'. Write the mh_sampler function that takes the arguments:

- θ_0: starting point,
- logprior: logprior function,
- loglikelihood: loglikelihood function,
- proposal: proposal generating function,
- nsamples: number of samples to generate,

and that returns the chain as well as the acceptance rate. Launch this MH sampler with $n_0 = 2\pi/6.5$ rad/d and A_0, B_0, C_0 given by the maximum likelihood estimates of Sect. 4. The acceptance rate should be between 10 and 35 (adjust the proposal function).

Indication: for σ_n take a fraction of $2\pi/T_{\text{obs}}$ where T_{obs} is the total timespan of observations (the difference between the last and first observation dates).

(d) Bonus question: it is not obvious whether a MCMC chain has converged. Search for MCMC convergence tests in the literature and perform one on your chain.

(e) Use the corner python package to produce a corner plot of the posterior distribution (i.e., histograms of each parameter as well as correlation plots between them). Compute the posterior mean and posterior median of A, B, C and n.

Note: When analyzing the results of an MCMC, it is common practice to ignore the first steps of the chain (typically the first fourth of the samples), to avoid biasing the results toward the starting point θ_0.

(f) The minimum mass of the planet is defined as $m \sin i$. For each sample θ_i of the MCMC, compute the corresponding minimum mass of the planet (you can assume $m \ll M$). Plot the histogram of the minimum mass values and compute its posterior mean and median.

(g) A common way to derive errorbars on physical parameters is to use credible intervals. A credible interval at level α for θ_k is an interval I such that

$$p\{\theta_k \in I|y\} = \alpha \tag{11}$$

Find the smallest credible interval at level $\alpha = 95\%$ for $m \sin i$.

Find the credible interval at 95% such that the probabilities for $m \sin i$ to be on the left and on the right of the interval are the same.