Bayesian evidence as a diagnostic for testing models of gravitational wave data

Vimal Simha
University of Glasgow

August 2004

Contents

Abstract

The Laser Interferometer Space Antenna (LISA) is expected to detect a large number of sources in the frequency range 0.0001Hz to 0.1Hz. A large number of sources such as compact galactic binaries and supermassive black hole binaries are thought to emit gravitational radiation in this range. In this study we examine how parameters such as the number of sources M, the amplitude A, the frequency f and the phase ϕ can be extracted from the data stream. Firstly, we set up a simplified discretised problem and estimate the number of sinusoids M with a comprehensive integration. Following, this we set up a more complex continuous problem and attempt to extract the parameters using Markov-Chain Monte-Carlo techniques.

Chapter 1

Introduction to the problem of estimating the number of sinusoids

A data set consisting of a superposition of M sinusoids and random Gaussian noise is observed. The problem is to determine the number of sinusoids M, for which there is the greatest evidence or in essence the most probable value of M. We approach the problem using the techniques of Bayesian hypothesis testing.

1.1 Formulation of the general problem of parameter estimation

We begin by considering the given data set $\{D_k\}$. The data set $\{D_k\}$ is modeled as a sum of M sinusoids of amplitudes $\{A_i\}$, frequencies $\{\omega_i\}$, phases $\{\phi_i\}$ and random Gaussian noise $\{n_i\}$.

$$\{D_k\} = \Sigma_{i=1}^{m} A_i \sin \omega_i t_k + \phi_i + n_i \tag{1.1}$$

where $\{A_i\}$ is the set of amplitudes, $\{\omega_i\}$ is the set of frequencies and $\{\phi_i\}$ is the set of phases. The Gaussian noise is $\{n_i\}$ and t_k refers to the k time samples which give the data set $\{D_k\}$.

Our model data for a particular amplitude, frequency and phase under consideration is given by $\{F_k\}$ which again is a summation of M sinusoids but without the Gaussian noise.

$$F_k = \sum_{i=1}^{M} A_i \sin \omega_i t_k + \phi_i. \tag{1.2}$$

In this problem we need to determine the most probable value of M. To do so we need to estimate the probability distribution of M or the probability of a particular M given the data set.

$$\text{prob}(M/\{D_k\}, I) \tag{1.3}$$

where the notation means that we are considering the probability of M given the data set $\{D_k\}$ and any prior information that we may have I.

We begin with Bayes' theorem

$$\text{prob}(M/\{D_k\}, I) = \frac{\text{prob}(\{D_k\}/M, I) \times \text{prob}(M/I)}{\text{prob}(\{D_k\})/I)}. \tag{1.4}$$

$\text{prob}(M/I)$ is called the prior i.e. the probability of M before we have the data, $\text{prob}(\{D_k\}/M, I)$ is called the likelihood i.e. the likelihood of a particular data set given a set of parameter values, $\text{prob}(M/$
$D_k, I)$ is called the posterior probability and $\text{prob}(\{D_k\})/I)$ is called the evidence.

At the outset, we assign a uniform prior from $M = 0$ up to a maximum value of $M = M_{\max}$. By assigning a uniform prior, we are not expressing any preference for a particular value of M or even a range of values. If it were possible to obtain constraints on the M values from astrophysical considerations, then a suitably modified prior would be appropriate. Our uniform prior serves as an expression of our ignorance of possible M values.

$$\text{prob}(M/I) = \frac{1}{M_{\max}}. \tag{1.5}$$

We can determine the evidence $\text{prob}(\{D_k\}/I)$ at a later stage using the normalisation condition that the sum of probabilities for all values of M should be one.

$$\text{prob} \sum (M/\{D_k\}, I) = 1. \tag{1.6}$$

Next let us consider the likelihood, which is the probability of obtaining a particular data set given the parameters for the model. First, we express the likelihood as a marginal integral over the parameters.

$$\text{prob}(\{D_k\}/M, I) = \int \dots \int \text{prob}(\{D_k\}, \{A_i, \omega_i, \phi_i\}/M, I)\, \mathrm{d}^M A_i\, \mathrm{d}^M \omega_i\, \mathrm{d}^M \phi_i \tag{1.7}$$

This expression can be further simplified using the product rule of probabilities.

$$\text{prob}(\{D_k\}, \{A_i, \omega_i, \phi_i\}/M, I) = \text{prob}(\{D_k\}/\{A_i, \omega_i, \phi_i\}, M, I) \times \mathit{prob}(\{A_i, \omega_i, \phi_i\}/M, I) \tag{1.8}$$

Putting both of these together we get :

$$\text{prob}(\{D_k\}/M, I) = \int ... \int \text{prob}(\{D_k\}/\{A_i, \omega_i, \phi_i\}, M, I) \times \text{prob}(\{A_i, \omega_i, \phi_i\}/M, I)\, \mathrm{d}^M A_i\; \mathrm{d}^M \omega_i\; \mathrm{d}^M \phi_i. \tag{1.9}$$

In the above expression, note that $\text{prob}(\{A_i, \omega_i, \phi_i\}/M, I)$ plays the role of a prior on the parameters. Once again, we shall assign a uniform prior as an expression of our ignorance about the possible parameter values. Note that the phase ϕ should lie between 0 and 2π.

$$\text{prob}(\{A_i, \omega_i, \phi_i\}/M, I) = \frac{1}{[2\pi A_{\text{max}}\omega_{\text{max}}]^M} \tag{1.10}$$

The likelihood function $\text{prob}(\{D_k\}/\{A_i, \omega_i, \phi_i\}, M, I)$ demands a more involved treatment. Before starting it is worth reiterating that $\{F_k\}$ represents the model data which is a function of $\{A_i, \omega_i, \phi_i\}$ and $\{D_k\}$ represents the actual data which includes random Gaussian noise.

Once again we start with Bayes' theorem :

$$\text{prob}(F_k/\{D_k\}, I) \quad \alpha \quad \text{prob}(\{D_k\}/F_k, I) \ \times \ \text{prob}(F_k/I). \tag{1.11}$$

Once more we shall assign a uniform prior as an expression of our ignorance, thus arriving at the equation

$$\text{prob}(F_k/\{D_k\}, I) \quad \alpha \quad \text{prob}(\{D_k\}/F_k, I). \tag{1.12}$$

We assume that each individual datum D_k is independent. Therefore we can say :

$$\text{prob}(\{D_k\}/F_k, I) = \prod_{k=1}^{N} \text{prob}(D_k/F_k, I) \tag{1.13}$$

We also assume that the noise associated with each datum can be represented by a Gaussian distribution.

$$\text{prob}(\{D_k\}/F_k, I) = \frac{1}{\sigma_k\sqrt{2\pi}} \exp[\frac{-(F_k - D_k)^2}{2\sigma^2}]. \tag{1.14}$$

Using the above calculation, in the context of our problem :

$$\text{prob}(\{D_k\}/\{A_i, \omega_i, \phi_i\}, M, I) \ \alpha \ \exp(\frac{-\chi^2}{2}). \tag{1.15}$$

where

$$\chi^2 = \sum_{k=1}^{N} (\frac{F_k - D_k}{\sigma_k})^2. \tag{1.16}$$

Finally we have :

$$\text{prob}(M/\{D_k\}, I) \ \alpha \ \frac{1}{(2\pi A_{\text{max}}\omega_{\text{max}})^M} \int ... \int \exp(\frac{-\chi^2}{2}) \, \text{d}^M A_i \ \text{d}^M \omega_i \ \text{d}^M \phi_i. \tag{1.17}$$

1.2 Formulation of a simplified discrete problem

In the previous section we saw how we can determine the most probable value of M in the most general case. In this section we generate simulated data and attempt to recover the value of M by marginalising over the other parameters.

In order to simplify the problem slightly (by M dimensions!) we shall set the phase $\phi = 0$. Also, henceforth we shall consider frequencies f rather than angular frequencies ω as we have done so far.

Our model data is now represented by

$$F_k = \sum_{i=1}^{M} A_i \sin(2\pi f_i t_k) \tag{1.18}$$

where k is the number of time samples of the data.

We generate a simulated data set consisting of the model data and artificially generated noise sampled randomly from a Gaussian distribution. For our simulated data set $\{D_k\}$ we take $M = 3$ and set the amplitudes A_i equal to 1, 3 and 5 and the frequencies f_i equal to 0.01, 0.03 and 0.05. We generate one hundred time samples evenly sampled in time with a time gap of one second between observations. Noise of mean zero and standard deviation $\sigma = 1$ is added to all the time samples. (Note that σ is no longer σ_k as the noise is assumed not to depend on each observation)

Having generated the simulated data, we need to formulate the problem in a suitable way so as to be able to recover the most probable value of M from the data set $\{D_k\}$. We need to find $\text{prob}(M/\{D_k\}, I)$.

The starting point as always must be Bayes' theorem:

$$\text{prob}(M/\{D_k\}, I) = \frac{\text{prob}(\{D_k\}/M, I) \ \times \ \text{prob}(M/I)}{\text{prob}\{D_k\}/I} \tag{1.19}$$

For this problem let us consider six possible values of M namely $0, 1, 2, 3, 4, 5$. So, that

$$\text{prob}(M/I) = \frac{1}{6} \tag{1.20}$$

for each of the possible values of M as we shall assume that we have no prior knowledge regarding M.

Following the same procedure as in the previous section, we arrive at the equation :

$$\text{prob}(\{D_k\}/M, I) = \int ... \int \text{prob}(\{D_k\}/\{A_i, \omega_i\}, M, I) \times \text{prob}(\{A_i, \omega_i, \}/M, I) \, \mathrm{d}^M A_i \ \mathrm{d}^M \omega_i. \tag{1.21}$$

Our next step is to assign the prior namely $\text{prob}(\{A_i, \omega_i, \}/M, I)$. At, this point we shall make a further simplification. We shall formulate the problem as a discrete one rather than a continuous one in that we shall only allow certain amplitudes and frequencies and shall forbid intermediate values of them.We consider five values of the amplitude $A_i = 1, 2, 3, 4, 5$ and five values of the frequency $f_i = 0.01, 0.02, 0.03, 0.04, 0.05$. Because of the way we are formulating the problem, we are disallowing any other values of the amplitude and frequency than those stated above.

As we are allowing five possible values for each A_i and f_i. Our prior is

$$\text{prob}(\{A_i, f_i\}/M, I) = \frac{1}{(25)^M}. \tag{1.22}$$

Since, M represents the number of sinusoids, this prior tells us that there are 25 possible combinations of frequency and amplitude for each sinusoid. For a larger value of M, say n this would tell us that there were $(25)^n$ possible combinations of frequency and amplitude, 25 for each of the n sinusoids.

From the previous section we have

$$\text{prob}(\{D_k\}/\{A_i, f_i\}, M, I) \ \alpha \ \exp(\frac{-\chi^2}{2}) \tag{1.23}$$

Hence

$$\text{prob}(M/\{D_k\}, I) \ \alpha \ \frac{1}{(25)^M} \ \times \int \ldots \int \exp(\frac{-\chi^2}{2}) \quad \text{d}^M A_i \ \text{d}^M f_i. \tag{1.24}$$

As we have formulated the problem as a discrete one we replace the integral by a sum over all the $2M$ parameters.

$$\text{prob}(M/\{D_k\}, I) \ \alpha \ \frac{1}{(25)^M} \ \times \sum \ldots \sum \exp(\frac{-\chi^2}{2}) \tag{1.25}$$

We need some more information in order to obtain the exact value. This information is present in the form of the normalisation condition i.e. the fact that the sum of the probabilities for all values of M is one.

$$\sum \text{prob}(M/\{D_k\}, I) = 1. \tag{1.26}$$

Chapter 2

The Occam factor

In the last section we found an expression for the probability of a particular value of the number of sinusoids, M which would allow us to construct the probability distribution for M. The expression requires closer study before progress can be made.

$$\text{prob}(M/\{D_k\}, I) \;\alpha\; \frac{1}{(25)^M} \;\times\; \sum \ldots \sum \exp(\frac{-\chi^2}{2}). \tag{2.1}$$

M is the number of sinusoids. Therefore, a larger M value implies a larger number of amplitudes and frequencies i.e. the sets $\{A_i\}$ and $\{f_i\}$ have more members. As a result of this, for a larger value of M, we have a larger number of parameters to fit the data. Usually, a better fit can be obtained with more parameters. Therefore, if we only consider the fit of the model with the data, we would tend to prefer models with more parameters against models with fewer parameters. This leads us to a counterintuitive result, i.e. it would be possible to fit data as accurately as desired by introducing an arbitrary number of parameters. Clearly, we require a more robust method for identifying the optimum number of parameters.

William of Occam (1287-1347), a fourteenth century English logician and friar first formulated what is sometimes called the principle of parsimony. These days we know it as Occam's Razor. It is typically stated as, *"entia non sunt multiplicanda preaeter necessitatem"* in Latin, which is translated as "entities are not to be multiplied beyond necessity" in English. There are a number of variations of this, one of which is most appropriate for our purpose here *"frustra fit per plura quod potest fieri per pauciora"* (in vain we do by many which can be done by means of fewer).

Notice that in the equation above, besides the least squares term, there is the other term $\frac{1}{25^M}$ in this specific case, or $\frac{1}{(A_{max}-A_{min})(f_{max}-f_{min})^M}$ in general. For a larger value of M this reduces the $prob(M/\{D_k\}, I)$ (the probability of that particular value of M). Thus, the Bayesian approach adopted by us introduces a natural Occam factor which penalises the model with the larger number of parameters. Hence, although a better fit may be obtained with a larger number of parameters, the Occam factor which is intrinsic to the approach adopted here compensates for this.

Chapter 3

Solving the discrete problem

3.1 Simulated Data

In the previous sections we obtained an expression for the posterior probability distribution of M in an immediately usable form. Our aim here is to set up and implement a suitable procedure for finding the posterior probability distribution for M. In order to do that we need to find the individual probabilities for each value of M. We consider six possible values of M (as the prior) namely 0,1,2,3,4 and 5 and find the probability for each of them.

As, we said in the previous section we have allowed only 5 values for the amplitude and frequency. For a given value of M we calculate the value of χ^2 for every possible combination of frequency (frequencies if $M > 1$) and amplitude(s). This gives a $2M$ dimensional array of χ^2's. We go on to calculate $\exp \frac{(-\chi^2)}{2}$ of the elements of the array and find the sum of all its elements. This sum divided by $(25)^M$ is the (unnormalised) value of $\text{prob}(M/\{D_k\}, I)$. Once we have obtained the probabilities for all the possible values of M we go on to normalise it using the fact that the sum of all the probabilities must be one.

Just to reiterate, our model consists of three sinusoids with amplitudes equal to 1, 3 and 5 with frequencies of 0.01, 0.03 and 0.05 respectively. Random Gaussian noise of mean equal to zero and $\sigma = 1$ has been added to it.(Figure 1). It is also instructive to look at the power spectrum which is the Fourier transform of the simulated time series data (Figure 2).

3.2 Frequencies and amplitudes

Having generated the simulated data, the next logical step is to calculate a χ^2 surface consisting of the χ^2 values for each combination of parameters. Such a computation yields a $3M$ dimensional hypersurface. It might be interesting as well as instructive to explore this before going further.

Let us first explore one-dimensional 'slices' through the χ^2 surface. Let us set the value of M equal to 3 which is its actual value at this time. We shall explore variations in M at a later stage. For example let us set the values of all frequencies and all but one amplitudes to their actual values to explore the χ^2 distribution with respect to each of the amplitudes. For the purpose of this

Figure 3.1: Simulated data for amplitude equal to 1,3 and 5 and frequency equal to 0.01,0.03 and 0.05 respectively

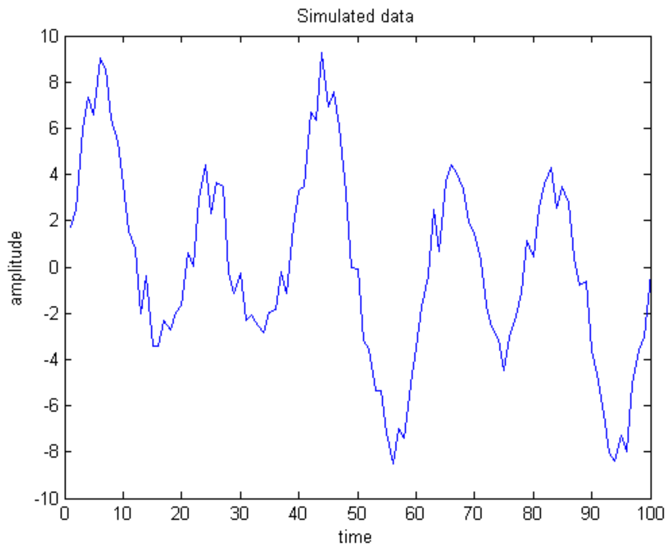


Figure 3.2: Power spectrum for the simulated data for amplitude equal to 1,3 and 5 and frequency equal to 0.01,0.03 and 0.05 respectively

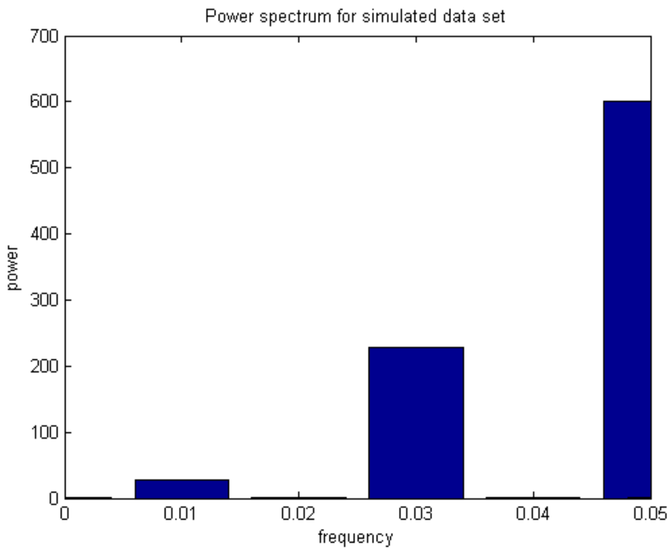
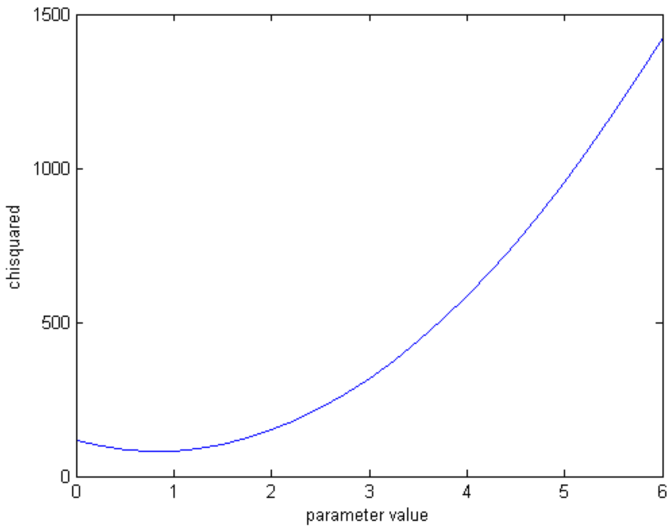


Figure 3.3: χ^2 distribution for amplitude $A1$ with all other parameters set to their actual values



investigation, let us call the amplitudes $A1$, $A2$ and $A3$ and the frequencies $f1$, $f2$ and $f3$ where their actual values are 1,3,5 and 0.01,0.03 and 0.05 respectively.

The amplitude χ^2 plots and hence the probability distributions are smooth with respect to changes in the amplitudes. Now, let us set the values of the amplitudes and all but one frequency to their correct values and explore the change in χ^2 as one of the frequencies is changed.

We see that the χ^2 distribution is much more sensitive to changes in frequency than amplitude. This implies that it might be possible to obtain a better constraint on the frequency than the amplitude.

Besides these it is also possible to look at contour plots by holding all but two parameters constant. We have produced two such χ^2 contour plots for $A1$ and $A2$, one with the correct frequencies and the other with erroneous frequencies.

3.3 The number of sinusoids

We now go on to use these χ^2 hypersurfaces to compute the posterior probability distribution for the number of sinusoids M. In the previous chapter, we have derived an expression for the probability for each particular M value. This essentially involves taking the exponential of $\frac{\chi^2}{2}$ and dividing it by the Occam factor.

As we said earlier, we consider only six possible values of M. For each of these we compute the probability by marginalising over the amplitudes and frequencies. The normalised posterior probability distribution for M is given in the figure below.

We can see that $M = 3$, is clearly preferred over the alternatives. However,

Figure 3.4: χ^2 distribution for amplitude $A2$ with all other parameters set to their actual values

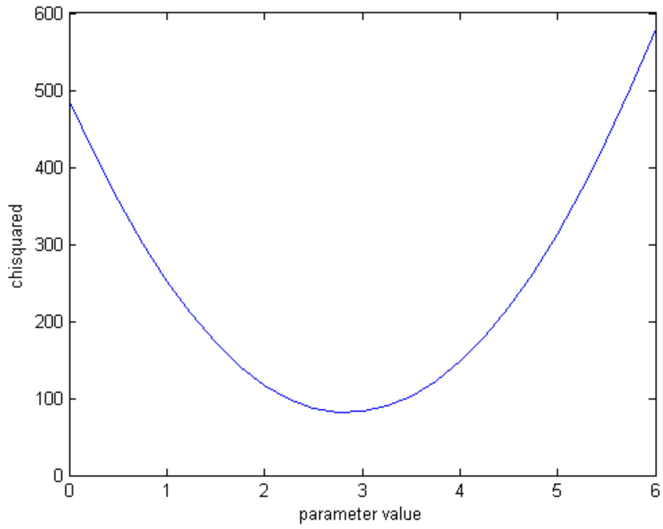


Figure 3.5: χ^2 distribution for amplitude $A3$ with all other parameters set to their actual values

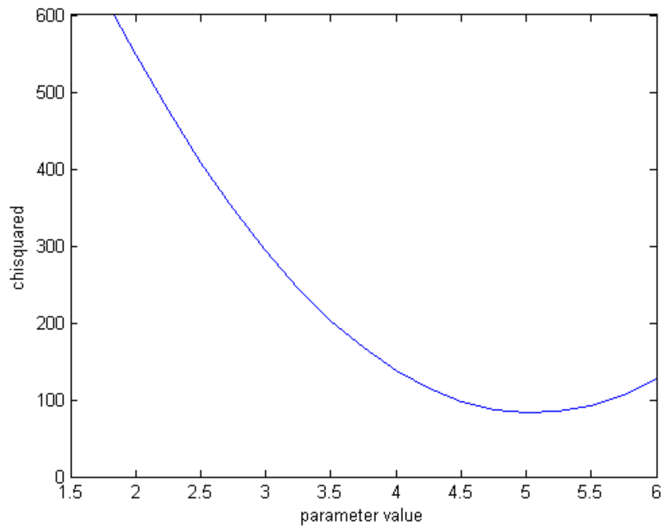


Figure 3.6: χ^2 distribution for frequency $f1$ with all other parameters set to their actual values

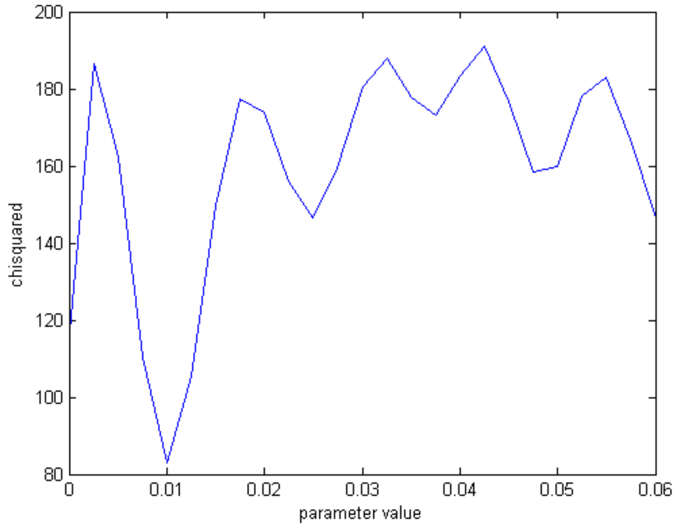


Figure 3.7: χ^2 distribution for frequency $f2$ with all other parameters set to their actual values

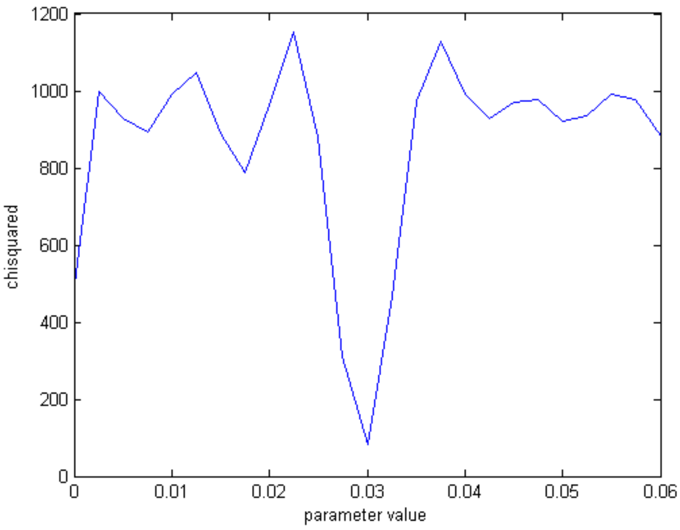


Figure 3.8: χ^2 distribution for frequency $f3$ with all other parameters set to their actual values

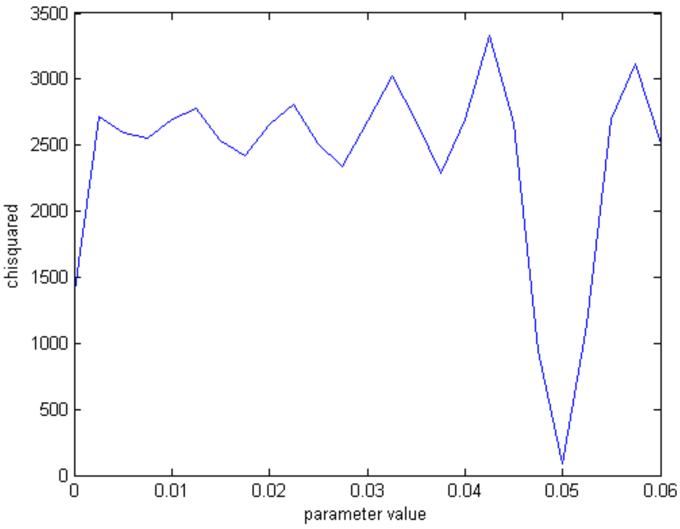


Figure 3.9: χ^2 contour plot for $A1$ and $A2$ with all other parameters fixed to correct values

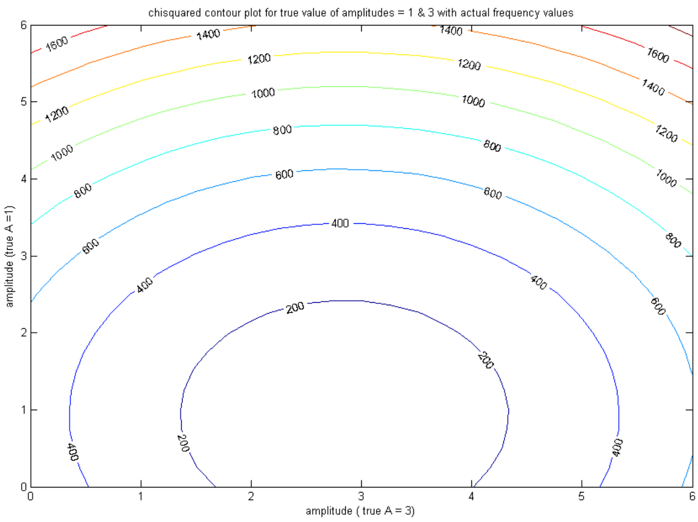


Figure 3.10: χ^2 contour plot for $A1$ and $A2$ with slightly erroneous frequencies

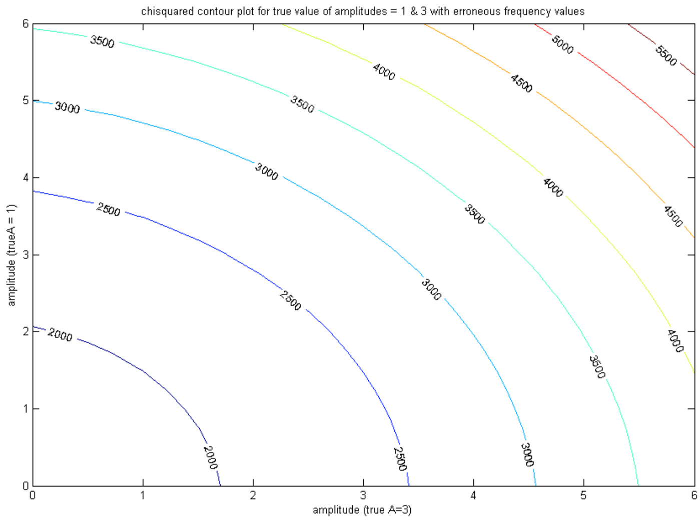


Figure 3.11: Normalised posterior probability distribution for M

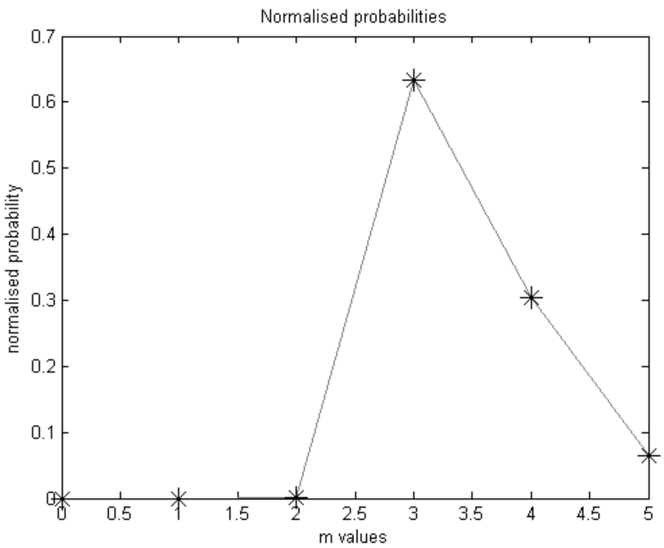
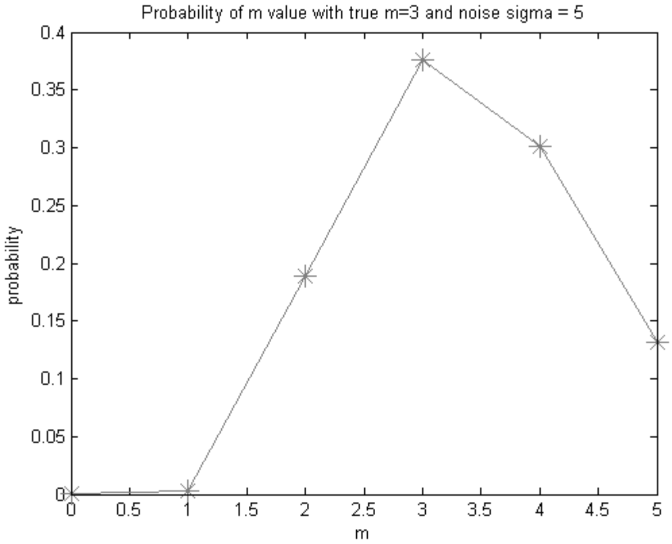


Figure 3.12: Posterior pdf for M with $\sigma = 5$



$M = 4$ gives a reasonable fit with certain combinations of parameters which is reflected in the figure. But, it has to contend with the Occam factor!

3.4 Other sets of simulated data

Next, we investigate the behaviour of the posterior probability distribution of M for different sets of simulated data. First of all we investigate the effect of increasing the level of noise. We use the same set of parameters but increase the value of the noise, σ to 5.

We are still able to delineate $M = 3$ as the most probable value. Let us now increase σ to 15.

Now, even for a very low signal to noise ratio we are able to identify the number of sinusoids.

We also investigate probability distributions for M for actual values of M equal to other values than 3.

Perhaps, it is also worth looking at what happens if there is no signal whatsoever and we have just random Gaussian noise.

Figure 3.13: Posterior pdf for M with $\sigma = 15$

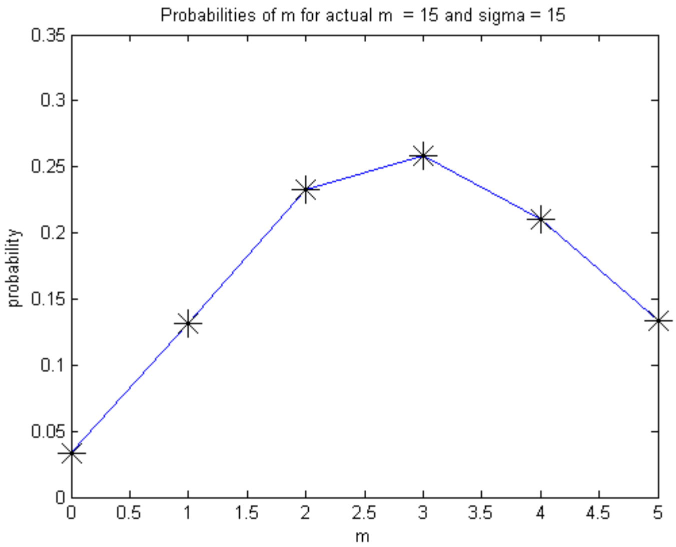


Figure 3.14: Posterior pdf for M with actual $M = 1$

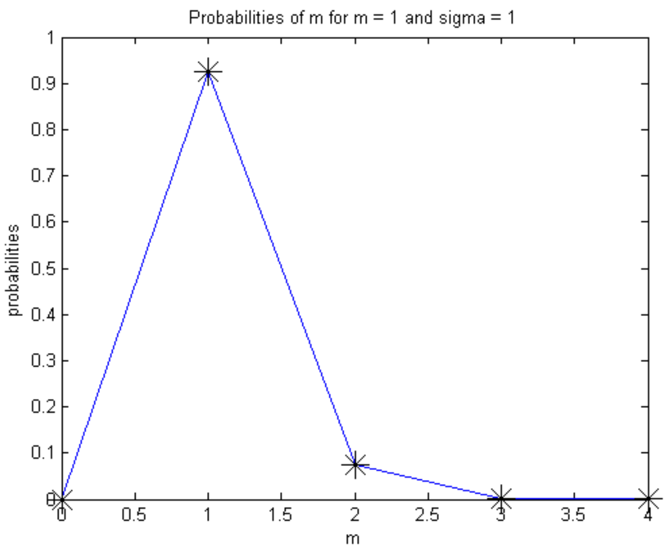


Figure 3.15: Posterior pdf for M with actual $M = 2$

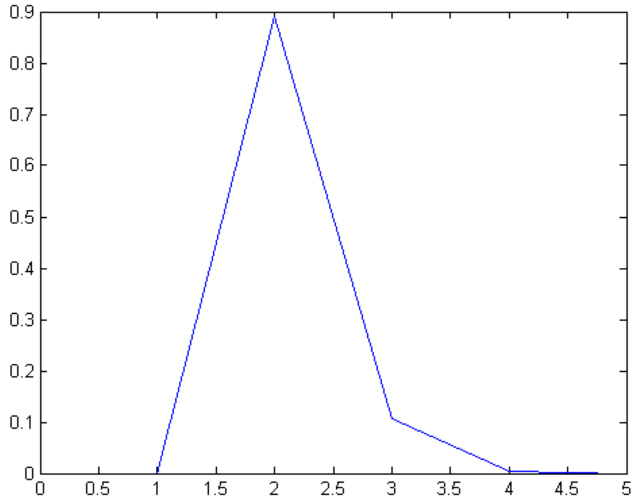


Figure 3.16: Posterior pdf for $M = 0$ and $\sigma = 1$ (unnormalised)

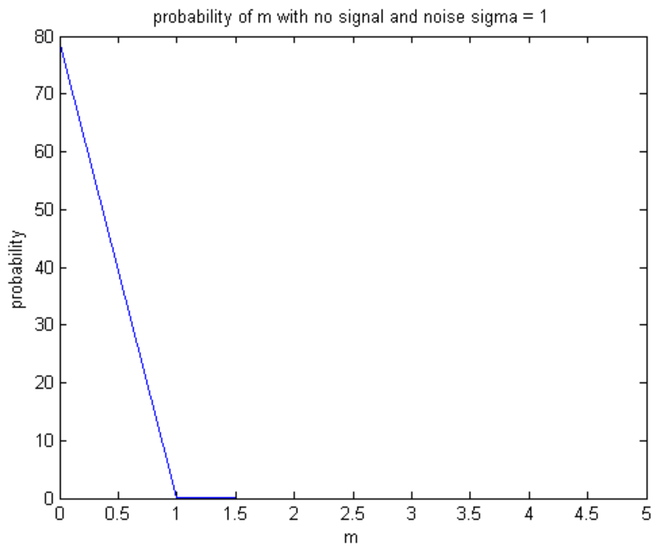
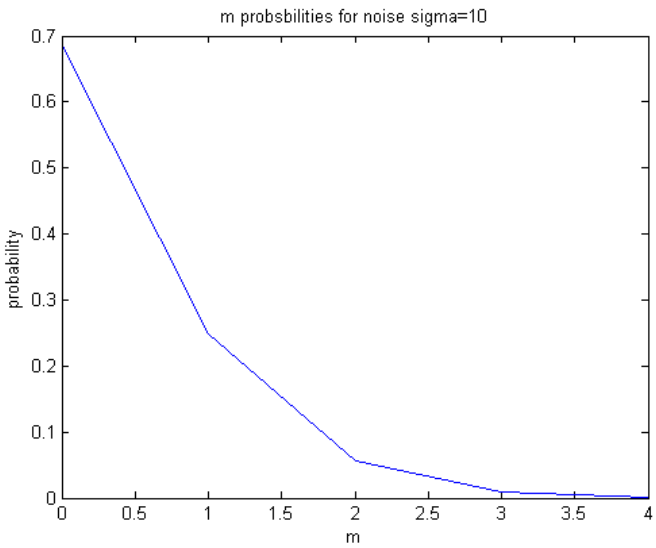


Figure 3.17: Posterior pdf for $M = 0$ and $\sigma = 10$



Chapter 4

Determination of frequencies and amplitudes by marginalisation

Once we have the probability distribution for M and the most probable value of M, the next step would be to determine the posterior probability distribution for the amplitudes and the frequencies. This is done by marginalising over all other parameters including M. Thus it is different from what was done previously. Previously, we computed the probability distribution for a particular parameter given the values of the other parameters. Now, we compute the probability distribution for a particular parameter by marginalising over the other parameters i.e. by integrating over all other parameters.

Now, let us take a particular parameter say $A1$. We compute the posterior probability distribution by marginalising over all the other parameters.

We do the same thing for frequency $f1$.

As can be seen by looking at the graphs, we have come across a problem. The probability distributions for $A2$ and $A3$ are exactly identical to the one for $A1$ and the probability distributions for $f2$ and $f3$ are exactly identical to the one for $f1$. This anomalous situation has arisen as a consequence of the symmetry between $A1$, $A2$,... (and $f1$,$f2$,...). There is no way of discriminating between the elements of the set of amplitudes or between the elements of the set of frequencies. It is absolutely necessary to find a way of discriminating between them before any progress can be made.

4.1 The Hierarchical Prior

The resolution of this anomaly lies in the prior! If instead of using a symmetric prior for each amplitude and frequency, we use an asymmetric prior. This asymmetric prior is of a hierarchical nature. We stipulate, for example $A1 \geq A2 \geq A3$ or $f1 > f2 > f3$ as a prior. This means that the parameter space is explored with the hierarchical orientation of the parameters always maintained and the probabilities are computed with this constraint. This means that it is now possible to produce probability distributions for any of the parameters such

Figure 4.1: Posterior pdf for amplitude $A1$, computed by marginalising over all other parameters.

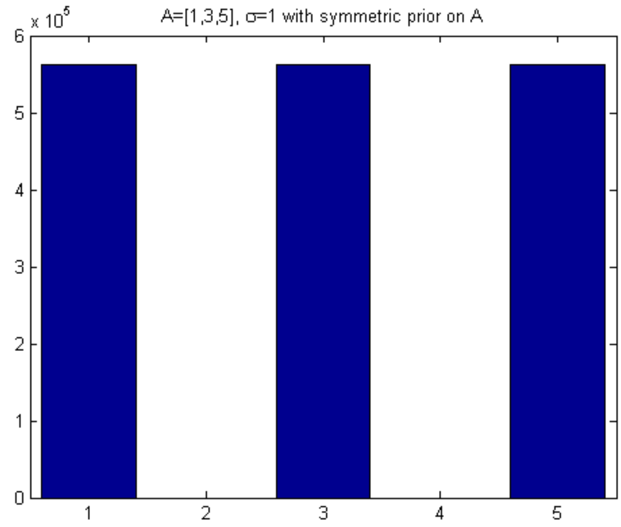


Figure 4.2: Posterior pdf for amplitude $f1$, computed by marginalising over all other parameters.

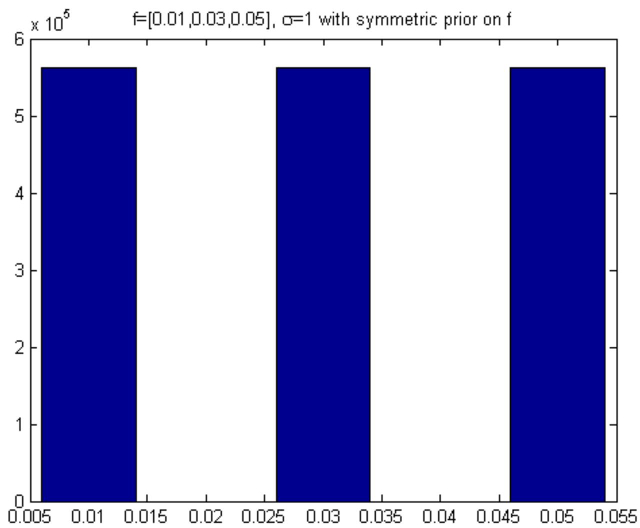
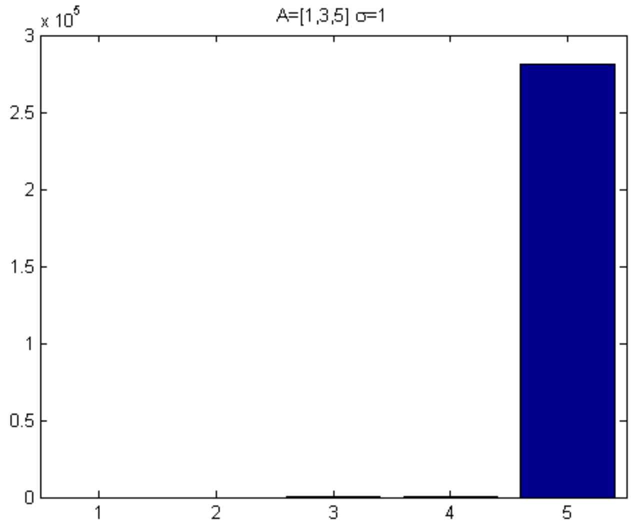


Figure 4.3: Posterior pdf for $A1$ with hierarchical prior on amplitude.



as $A1$ or $A2$ and they would be different this time! It is important to clarify that while same amplitudes are allowed, same frequencies cannot be allowed as that would be equivalent to having the same signal, whereas it is possible to have same amplitudes but different frequencies to qualify as different signals. A cautionary note is in order at this stage. One must take care to impose the hierarchical prior on either the amplitude or the frequency but not both as each amplitude has a frequency associated with it!

4.2 Hierarchical prior on amplitude

With the hierarchical prior, which delineates the different amplitudes or frequencies it is possible to obtain posterior probability distributions for each of the parameters. We proceed to do so. First let us impose the hierarchical prior on the amplitude. Let us say that $A1 \geq A2 \geq A3$ and investigate the posterior probability distributions for $A1$, $A2$ and $A3$.

There are frequencies associated with each of these amplitudes i.e. $f1$ associated with $A1$, $f2$ with $A2$ and so on. We now investigate the probability distributions for $f1$, $f2$ and $f3$.

4.3 Hierarchical prior on frequency

Now, we impose the hierarchical prior on the frequency and investigate the posterior probability distributions for the amplitudes and frequencies.

Figure 4.4: Posterior pdf for $A2$ with hierarchical prior on amplitude.

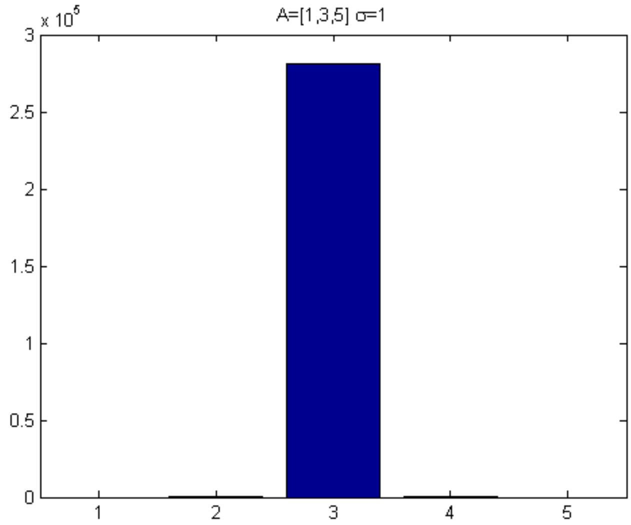


Figure 4.5: Posterior pdf for $A3$ with hierarchical prior on amplitude.

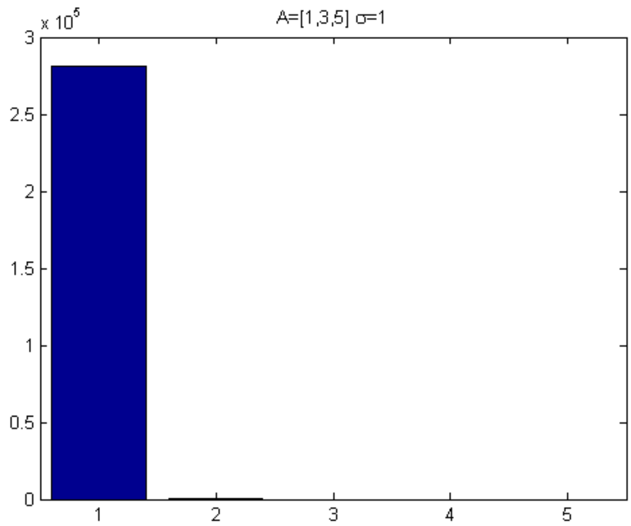


Figure 4.6: Posterior pdf for $f1$ with hierarchical prior on amplitude.

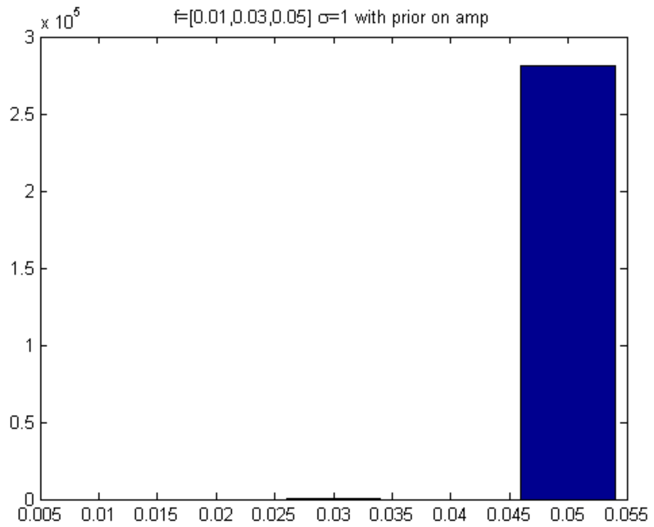


Figure 4.7: Posterior pdf for $f2$ with hierarchical prior on amplitude.

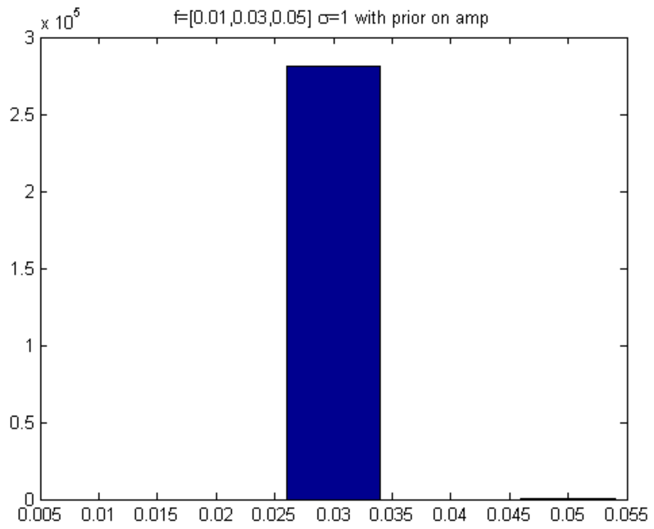


Figure 4.8: Posterior pdf for $f3$ with hierarchical prior on amplitude.

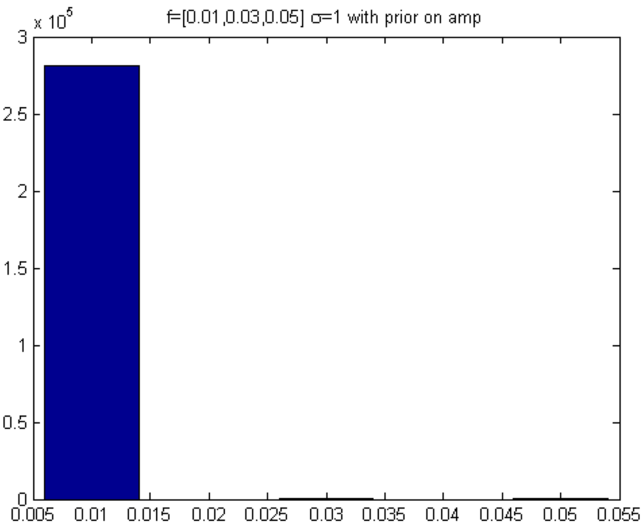


Figure 4.9: Posterior pdf for $A1$ with hierarchical prior on frequency.

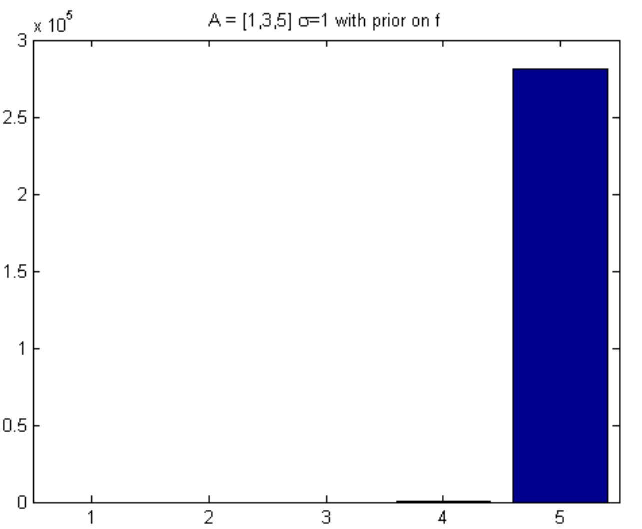


Figure 4.10: Posterior pdf for $A2$ with hierarchical prior on frequency.

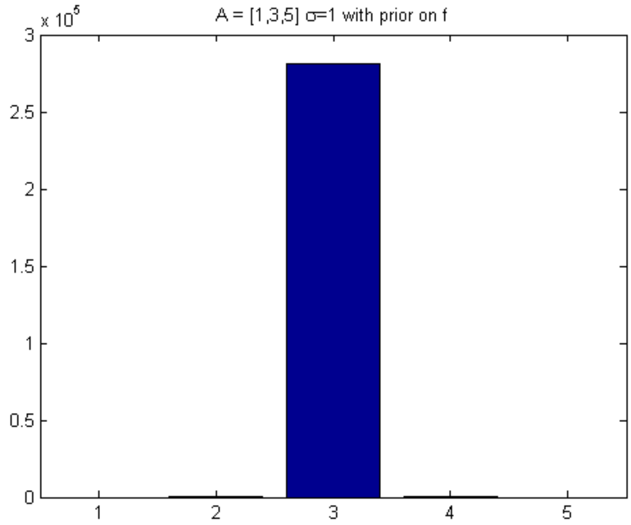


Figure 4.11: Posterior pdf for $A3$ with hierarchical prior on frequency.

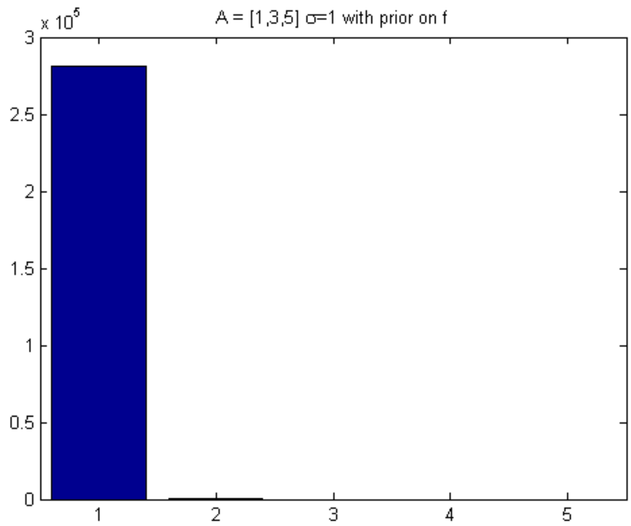


Figure 4.12: Posterior pdf for $f1$ with hierarchical prior on frequency.

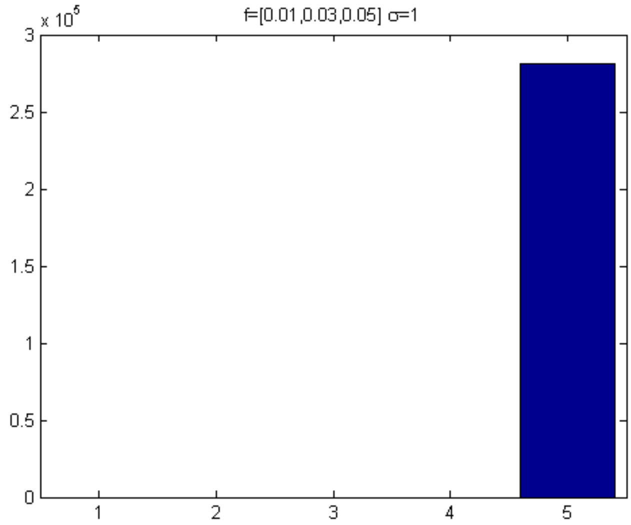


Figure 4.13: Posterior pdf for $f2$ with hierarchical prior on frequency.

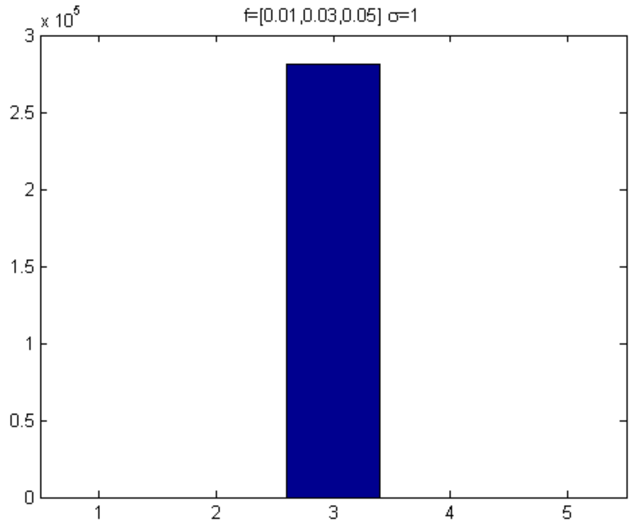
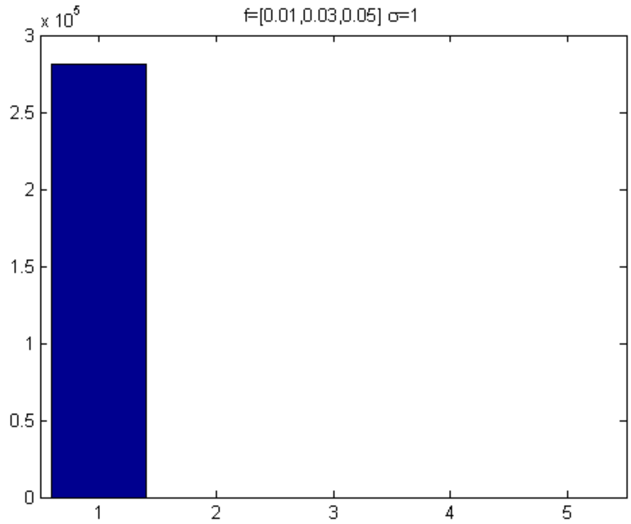


Figure 4.14: Posterior pdf for $f3$ with hierarchical prior on frequency.



4.4 Hierarchical priors with low signal to noise ratios

We see that the above probability distributions are extremely clear. They show a strong preference for the actual values of the amplitudes and frequencies. We would also like to know if there is a difference between imposing the hierarchical priors on the frequency or amplitude. One way of doing this is by increasing the noise level thereby decreasing the signal to noise ratio. This also allows us to find the limits of the signal to noise ratio for which this procedure works.

In order to achieve these twin aims, let us consider the same set of amplitudes and frequencies but increase the noise σ to 5. With this set of simulated data we now impose a hierarchical prior on amplitude and investigate the pdf's for the amplitudes and frequencies.

Next we impose the hierarchical prior on frequencies and explore the probability distributions for frequency and amplitude.

4.5 Amplitude or frequency

The results obtained from the previous section allow us to make some observations regarding the suitability of using hierarchical priors on the amplitude or the frequency. We see that in most cases we are able to identify the most probable amplitude and frequency. In cases where we are unable to do so, such as the pdf for $A1$ with $\sigma = 5$ and the hierarchical prior on amplitude, we feel that this is due to the fact that we have imposed rather artificial limits such as only allowing certain amplitude values and also limiting the range of amplitudes. These effects are expected to dissapear once a continous problem with a larger range of amplitude values are allowed.

Figure 4.15: Posterior pdf for $A1$ with $\sigma = 5$ and the hierarchical prior on amplitudes.

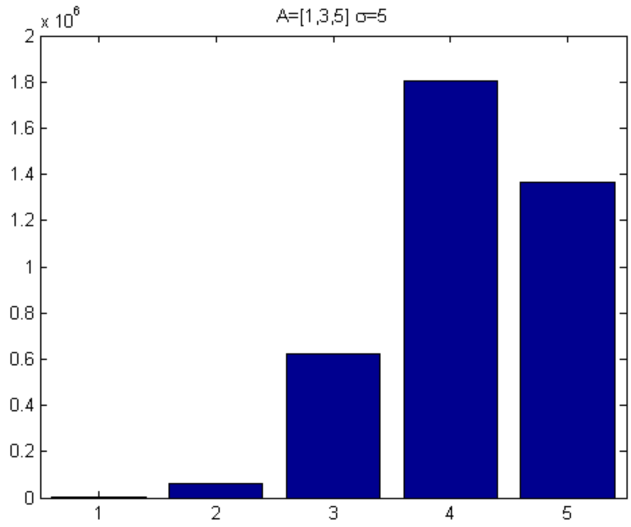


Figure 4.16: Posterior pdf for $A2$ with $\sigma = 5$ and the hierarchical prior on amplitudes.

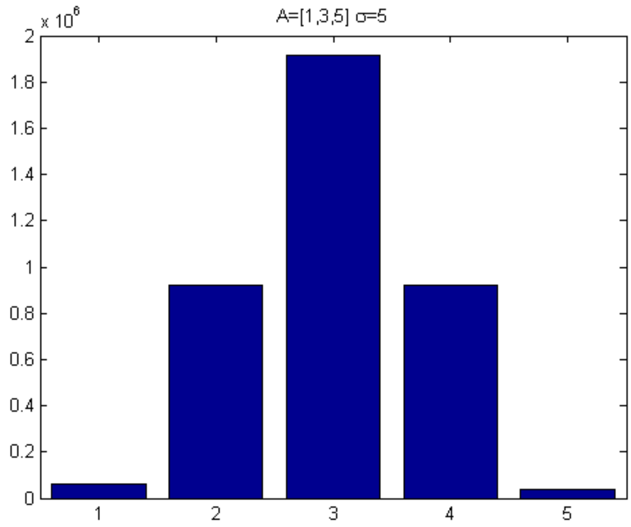


Figure 4.17: Posterior pdf for $A3$ with $\sigma = 5$ and the hierarchical prior on amplitudes.

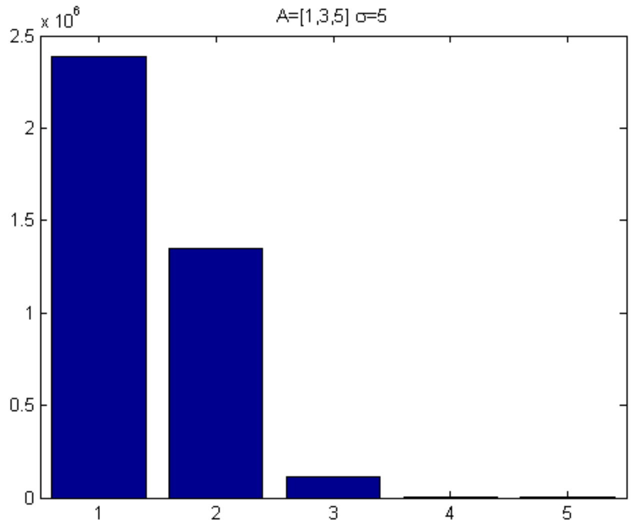


Figure 4.18: Posterior pdf for $f1$ with $\sigma = 5$ and the hierarchical prior on amplitudes.

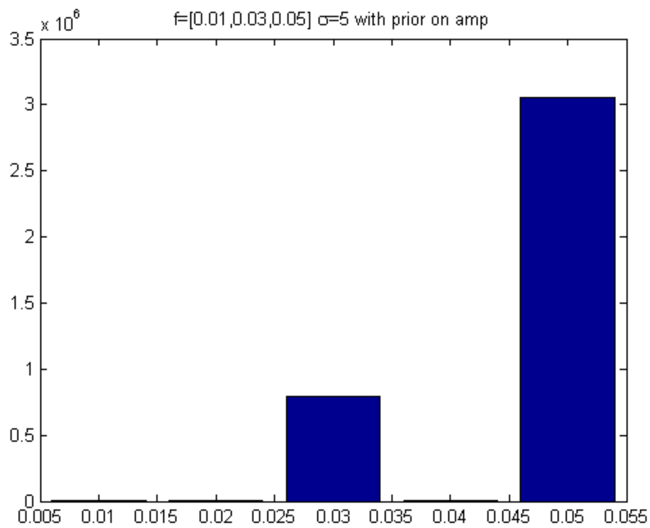


Figure 4.19: Posterior pdf for $f2$ with $\sigma = 5$ and the hierarchical prior on amplitudes.

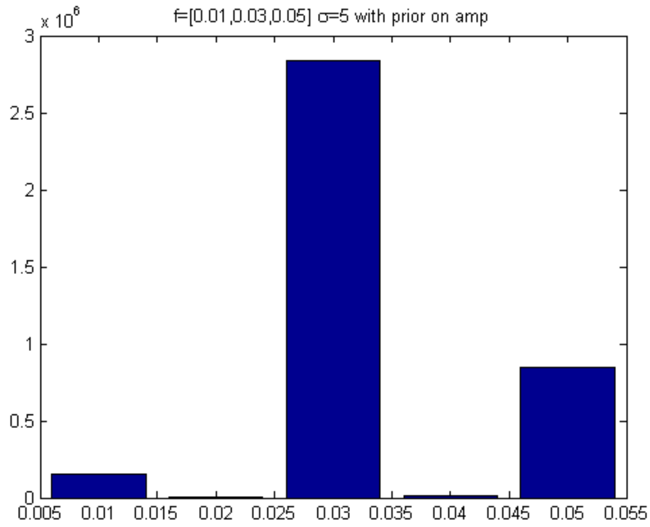


Figure 4.20: Posterior pdf for $f3$ with $\sigma = 5$ and the hierarchical prior on amplitudes.

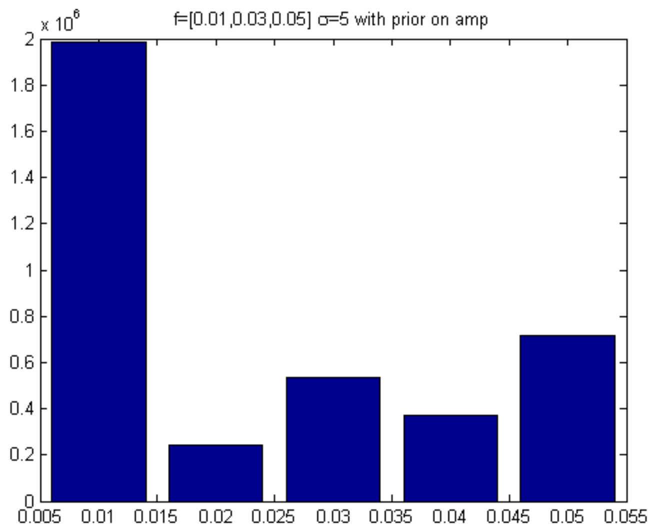


Figure 4.21: Posterior pdf for $A1$ with $\sigma = 5$ and the hierarchical prior on frequencies.

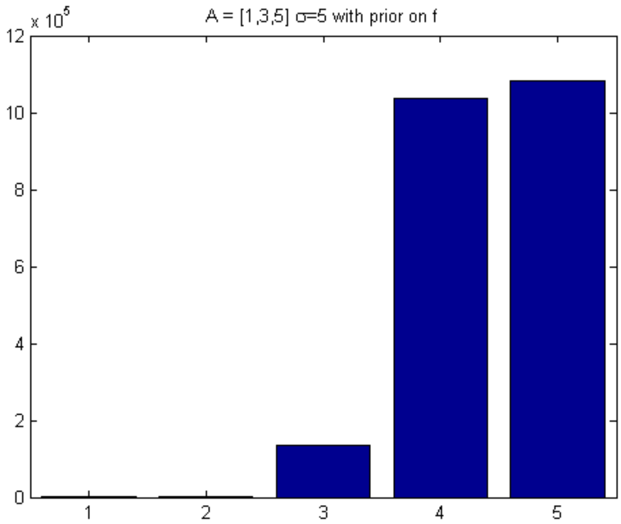


Figure 4.22: Posterior pdf for $A2$ with $\sigma = 5$ and the hierarchical prior on frequencies.

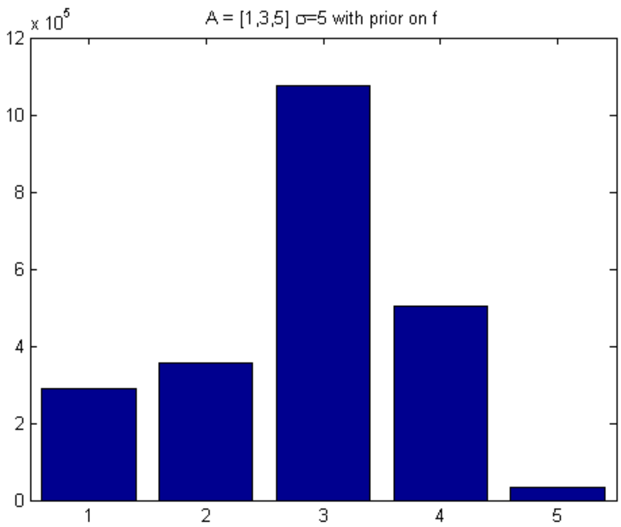


Figure 4.23: Posterior pdf for $A3$ with $\sigma = 5$ and the hierarchical prior on frequencies.

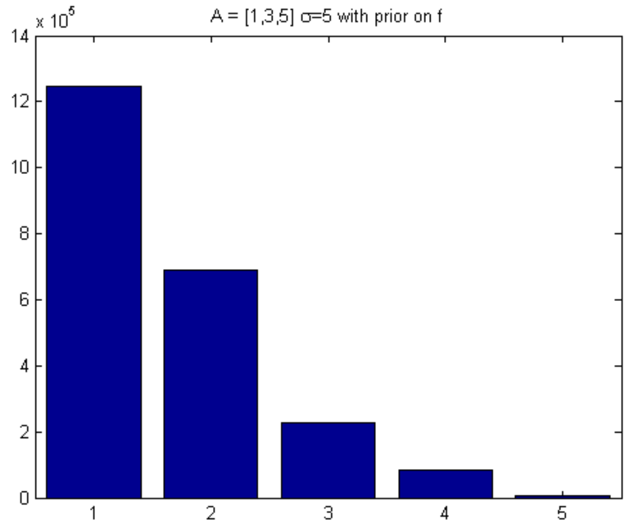


Figure 4.24: Posterior pdf for $f1$ with $\sigma = 5$ and the hierarchical prior on frequencies.

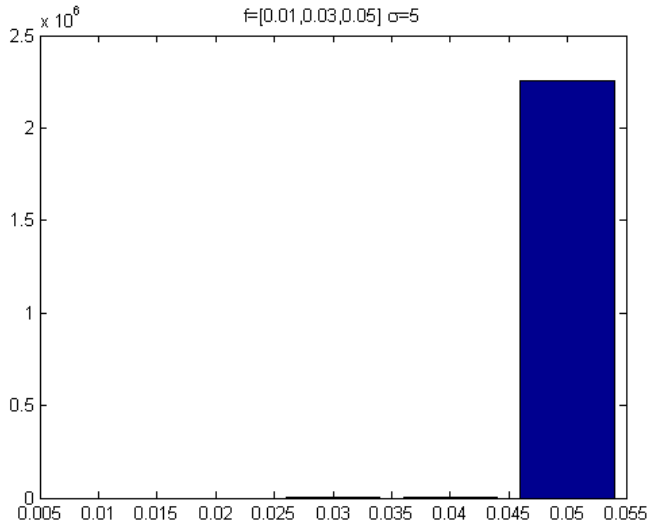


Figure 4.25: Posterior pdf for $f2$ with $\sigma = 5$ and the hierarchical prior on frequencies.

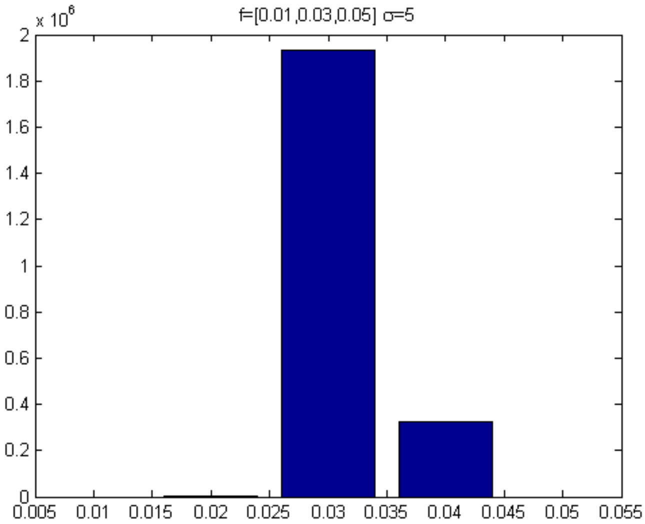
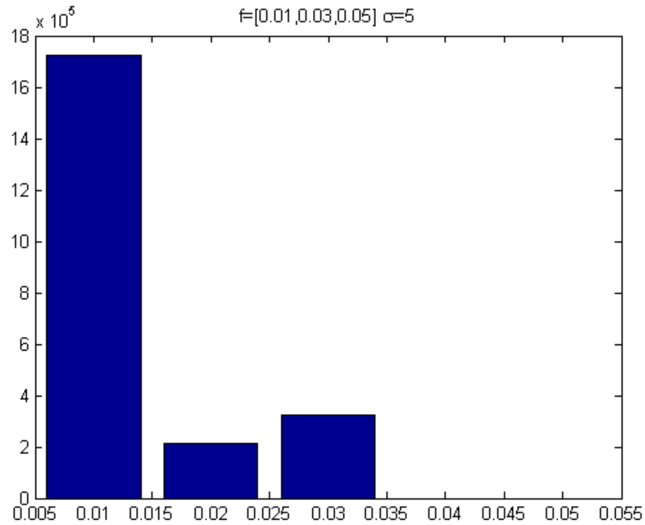


Figure 4.26: Posterior pdf for $f3$ with $\sigma = 5$ and the hierarchical prior on frequencies.



As before, we see that a stronger constraint can be obtained on the frequency than on the amplitude. We also see that the hierarchical prior on the