Formulation of the discrete problem

Vimal Simha
University of Glasgow

August 2004

Earlier, 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

$$F_k = \sum_{i=1}^{M} A_i \sin(\omega_i t_k) \tag{1}$$

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)

Having generated the simulated actual 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:

$$\text{prob}(M/\{D_k\}, I) = \frac{\text{prob}(\{D_k\}/M, I) \ \times \ \text{prob}(M/I)}{\text{prob}\{D_k\}/I} \tag{2}$$

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

$$\text{prob}(M/I) = \frac{1}{6} \tag{3}$$

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

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

$$\text{prob}(\{D_k\}/M, I) = \int ... \int \text{prob}(\{D_k\}/\{A_i, \omega_i\}, M, I) \times \text{prob}(\{A_i, \omega_i, \}/M, I)\, \text{d}^M A_i \ \text{d}^M \omega_i. \tag{4}$$

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 continous 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 would be

$$\text{prob}(\{A_i, f_i\}/M, I) = \frac{1}{(25)^M}. \tag{5}$$

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.

Since the term $\frac{1}{(25)^M}$ appears in the expression for $prob(M/\{D_k\}, I)$, it penalises the model with the larger M i.e. the model with the larger number of free parameters recognising the fact that it would be easier to obtain a better fit with a larger number of parameters. It acts in some sense as an 'Occam factor'.

From the previous section we have

$$\text{prob}(\{D_k\}/\{A_i, f_i\}, M, I) \ \alpha \ \exp(\frac{-\chi^2}{2}) \tag{6}$$

Hence

$$\text{prob}(M/\{D_k\}, I) \ \alpha \ \frac{1}{(25)^M} \ \times \int \ldots \int \exp(\frac{-\chi^2}{2}) \\ \text{d}^M A_i \ \text{d}^M f_i. \tag{7}$$

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

$$\text{prob}(M/\{D_k\}, I) \ \alpha \ \frac{1}{(25)^M} \ \times \sum \ldots \sum \exp(\frac{-\chi^2}{2}) \tag{8}$$

All other terms are constants and can be obtained by the normalisation condition

$$\sum \text{prob}(M/\{D_k\}, I) = 1. \tag{9}$$