# A brief introduction to the problem of estimating the number of sinusoids, $M$

Vimal Simha
University of Glasgow

August 2004

We are given a set of data consisting of a superposition of $M$ sinusoids and random Gaussian noise. 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.

We begin by considering the given data set $\{D_k\}$ :

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

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 would be

$$F_k = \sum_{i=1}^{m} A_i \sin \omega_i t_k + \phi_i. \tag{2}$$

In order to determine the number of sinusoids $M$, we need to estimate the probability distribution for $M$ or alternatively we need to estimate :

$$\text{prob}(M/\{D_k\}, I) \tag{3}$$

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

$$\text{prob}(M/\{D_k\}, I) = \frac{\text{prob}(\{D_k\}/M, I) \times \text{prob}(M/I)}{\text{prob}(\{D_k\})/I)}. \tag{4}$$

prob$(M/I)$ is called the prior,prob$(\{D_k\}/M, I)$ is called the likelihood, prob$(M/D_k, I)$ is the posterior and 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 $M = M_{\rm max}$. Hence :

$$\mathrm{prob}(M/I) = \frac{1}{M_{\rm max}}. \tag{5}$$

We can determine the evidence $\mathrm{prob}(\{D_k\}/I)$ at a later stage using the normalisation condition

$$\mathrm{prob} \sum (M/\{D_k\}, I) = 1. \tag{6}$$

Next let us consider the likelihood :

$$\mathrm{prob}(\{D_k\}/M, I) = \int \ldots \int \mathrm{prob}(\{D_k\}, \{A_i, \omega_i, \phi_i\}/M, I)\, \mathrm{d}^M A_i\, \mathrm{d}^M \omega_i\, \mathrm{d}^M \phi_i \tag{7}$$

expressing it as a marginal integral over the parameters of the model.

$$\mathrm{prob}(\{D_k\}, \{A_i, \omega_i, \phi_i\}/M, I) = \mathrm{prob}(\{D_k\}/\{A_i, \omega_i, \phi_i\}, M, I) \times prob(\{A_i, \omega_i, \phi_i\}/M, I) \tag{8}$$

from the 'product rule' of probabilities.

Hence :

$$\begin{aligned}&\mathrm{prob}(\{D_k\}/M, I)\\ &\qquad = \int ... \int \mathrm{prob}(\{D_k\}/\{A_i, \omega_i, \phi_i\}, M, I) \ \times \ \mathrm{prob}(\{A_i, \omega_i, \phi_i\}/M, I)\\ &\qquad \mathrm{d}^M A_i \ \mathrm{d}^M \omega_i \ \mathrm{d}^M \phi_i.\end{aligned} \tag{9}$$

Once again, we shall assign a uniform prior to $A_i$, $\omega_i$ and $\phi_i$ while recognising the fact that the phase $\phi$ should lie between 0 and $2\pi$.

$$\mathrm{prob}(\{A_i, \omega_i, \phi_i\}/M, I) = \frac{1}{[2\pi A_{\rm max}\omega_{\rm max}]^M} \tag{10}$$

The likelihood function demands a more involved treatment. Before starting it is worth reiterating that $F_k$ represents the model data and $\{D_k\}$ represents the actual data.

Once again we start with Bayes' theorem :

$$\mathrm{prob}(F_k/\{D_k\}, I) \quad \alpha \quad \mathrm{prob}(\{D_k\}/F_k, I) \ \times \ \mathrm{prob}(F_k/I). \tag{11}$$

Once more we shall assign a uniform prior, thus arriving at the equation

$$\mathrm{prob}(F_k/\{D_k\}, I) \quad \alpha \quad \mathrm{prob}(\{D_k\}/F_k, I). \tag{12}$$

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

$$\mathrm{prob}(\{D_k\}/F_k, I) = \prod_{k=1}^{N} \mathrm{prob}(D_k/F_k, I) \tag{13}$$

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

$$\mathrm{prob}(\{D_k\}/F_k, I) = \frac{1}{\sigma_k\sqrt{2\pi}}\exp[\frac{-(F_k - D_k)^2}{2\sigma^2}]. \tag{14}$$

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

$$\mathrm{prob}(\{D_k\}/\{A_i, \omega_i, \phi_i\}, M, I)\ \alpha\ \exp(\frac{-\chi^2}{2}). \tag{15}$$

where

$$\chi^2 = \sum_{k=1}^{N}(\frac{F_k - D_k}{\sigma_k})^2. \tag{16}$$

Finally we have :

$$\mathrm{prob}(M/\{D_k\}, I)\ \alpha\ \frac{1}{(2\pi A_{\max}\omega_{\max})^M}\int ... \int \exp(\frac{-\chi^2}{2})\,\mathrm{d}^M A_i\ \mathrm{d}^M\omega_i\ \mathrm{d}^M\phi_i. \tag{17}$$