

**BAYESIAN ANALYSIS V: AMPLITUDE ESTIMATION,
MULTIPLE WELL SEPARATED SINUSOIDS**

*G. LARRY BRETTHORST
Washington University
Department of Chemistry
Campus Box 1134
One Brookings Drive
St. Louis, Missouri 63130-4899*

ABSTRACT. Bayesian probability theory is used to estimate the amplitude of a single exponentially decaying sinusoid in NMR free induction decay (FID) data. The posterior probability for the amplitude is derived independent of the phase, frequency, decay rate constant, and variance of the noise. The estimate is shown to be accurate and precise in the sense that as the noise approaches zero, the estimate approaches the true value of the amplitude and the uncertainty in the estimate approaches zero. The uncertainty in the estimate is shown to vary inversely with the square root of the sampling rate. Finally, the calculation is applied in two examples. The first example demonstrates probability theory's ability to estimate frequencies and amplitudes in very low signal-to-noise. The second example illustrates the use of this calculation when the data contain multiple, well-separated sinusoids.

Introduction

In NMR the signal frequencies and the amplitudes are of fundamental importance. The frequencies are important because they convey information about the electronic environment around the nuclei, while the amplitudes are important because they convey information about the number of the nuclei. Recently, Bayesian probability theory has been applied to the problem of estimating the frequency in NMR FID data [1, 2, 3, 4]. In those studies, the joint posterior probability density for the frequencies of the signals was computed independent of the amplitudes, phases, variance of the noise, and decay rate constants. In this paper, the problem of estimating the amplitudes of the exponentially decaying sinusoid signal is addressed. Specifically, the posterior probability for the amplitude of an exponentially decay sinusoid is computed independent of the phase, variance of the noise, frequency, and decay rate constant given the data and a noise sample. When no noise sample is available the generalized results automatically reduce to the simplified case of given just the data. Little tutorial information about probability theory is included. For those needing additional background material, see Refs. [1] through [11].

Traditionally, the amplitudes of the NMR signals (the time $t = 0$ intensity of the time-domain resonance, often referred to as the integrated area of a frequency-domain peak) are estimated using a discrete Fourier transform absorption spectrum. To estimate the amplitudes, the NMR FID data are zero-padded, multiplied by a weighting function, and then a discrete Fourier transform is performed. The resulting spectrum is then phased (often using first and second order phase corrections), baseline corrected (typically the base line is fit to a fourth or fifth order polynomial and subtracted from the absorption spectrum) and blocked off into regions around the peaks. The integrals over these regions are then used as an estimate of the amplitude.

Integral estimates are subject to a number of uncertainties. First, noise in the data corrupts the amplitude estimate in an unknown, random way. Second, the integrated area depends in a systematic way on the phase. Third, baseline correction procedures not only subtract out the baseline, but

also subtract out some of the area associated with the peak. Fourth, when multiple lines are present, integration will estimate the sum of the overlap of the lines, not individual areas. If the resonances are well separated, this overlap effect is small, and the estimated amplitudes are nearly what one would have obtained in the absence of overlap. But even for “noise-free” data, integration yields a complicated weighted average of the resonances, not the true amplitude of the individual resonances. Finally, there is no way to determine how accurately the amplitude has been estimated. Repeating the experiment a number of times and then computing a (mean \pm standard deviation) estimate of the amplitudes does not answer the question of interest. This (mean plus \pm standard deviation) estimate answers a very complicated question about both the long term stability of the spectrometer, and the long run frequency behavior of integration; not how accurately the amplitude has been estimated for the given data set.

The posterior probability for the amplitude of multiple well separated sinusoids can be shown to separate in the multiple single amplitude estimation problems. In the first part of the paper, the joint posterior probability for the amplitude, frequency, and decay rate constant is computed independent of the phase and standard deviation of the noise. Then, using approximations, both the frequency and decay rate constants are removed as nuisance parameters to obtain the posterior probability for the amplitude independent of all of the other parameters. Two approximations are used. In the first approximation, the integrals are approximated by sums. This approximation is useful in the low signal-to-noise cases, i.e., frequency domain signal-to-noise of about 3 or less. In the second approximation, the “sufficient statistic” (a quantity that summarizes all of the information relevant to the estimation of a parameter) is expanded in a Taylor series about its maximum to obtain a Gaussian approximation. This approximation is useful in high signal-to-noise cases (frequency domain signal-to-noise greater than 3).

After deriving the posterior probability for the amplitudes, the question of the uncertainty and accuracy of the estimates is addressed. Two different forms of the uncertainty estimate are derived. The first form assumes the standard deviation of the noise is known. From this calculation the dependence of the parameter estimates on the experimental parameters (sampling rate, acquisition time, etc.) is derived. In the second form, the variance of the noise is removed as a nuisance parameter. From this form, the relationship between the parameter estimates and the noise sample is explored.

Finally, the calculations are applied to both real and simulated NMR FID data to demonstrate the strengths and limitations of the technique. In the first example, the low signal-to-noise approximation is used to estimate the amplitude and frequency on an actual FID with a frequency domain signal-to-noise ratio of approximately 3. In the second example, the posterior probability density for the amplitude of a single exponentially-decaying sinusoid is applied to FID data that contain multiple sinusoids. The limits of applicability are discussed and it is argued that *this probability density function will work for data containing 2 or more frequencies, provided the frequencies are well separated.*

The Posterior Probability Density For The Amplitude

The problem addressed is: given quadrature-detected FID data containing a *single exponentially decaying sinusoid*, calculate the posterior probability for the amplitude of the sinusoid independent of the phase, frequency, decay rate constant, and the variance of the noise. The information available for this calculation includes: (i) the data, (ii) a noise sample, and (iii) the single exponentially decaying sinusoidal model.

In quadrature FID data there are two data sets: the real data (0° phase), and the imaginary data (90° phase). The real data are assumed to be the sum of a signal plus noise:

$$d_R(t_i) = A \cos(\omega t_i + \theta) e^{-\alpha t_i} + e_i \quad (1)$$

where $d_R(t_i)$ denotes a real data sample at time t_i ($1 \leq i \leq N$). In the time domain, A , is the amplitude or intensity of the signal at time $t = 0$, while in the frequency domain it is the integrated area under a peak. The frequency of oscillation is ω , the phase of the sinusoid is θ , the exponential decay rate constant is α , and e_i represents noise. The three parameters ω , θ , α are considered to be nuisance parameters. A nuisance parameter is one that is not of interest in the final probability density function, although it is part of the model [1].

In addition to the signal in the real data, the imaginary data contain the same signal shifted by 90° :

$$d_I(t_i) = -A \sin(\omega t_i + \theta) e^{-\alpha t_i} + e_i \quad (2)$$

where $d_I(t_i)$ denotes an imaginary data sample at time t_i . The actual noise, e_i , present in the imaginary channel is assumed to be different from the noise in the real channel. However, the calculations will be simplified by assuming that the noise carries the same average power in both channels. The real data are denoted by $D_R \equiv \{d_R(t_1), \dots, d_R(t_N)\}$, the imaginary data by $D_I \equiv \{d_I(t_1), \dots, d_I(t_N)\}$, and $D \equiv \{D_R, D_I\}$ denotes both the real and imaginary data.

In previous work [4], it was assumed that a sample of the noise was known. When this sample was present, it placed a scale in the problem against which probability theory could measure small effects. When this sample was not present, the generalized results reduced to those found when no noise sample was gathered [1]. The same strategy will be used here and the more general results will be derived directly. Thus, a noise sample is assumed to be available. This sample is denoted as $D_{\sigma R} \equiv \{d_{\sigma R}(t_{\sigma 1}), \dots, d_{\sigma R}(t_{\sigma N_\sigma})\}$ for the real noise sample, $D_{\sigma I} \equiv \{d_{\sigma I}(t_{\sigma 1}), \dots, d_{\sigma I}(t_{\sigma N_\sigma})\}$ for the imaginary noise sample, and $D_\sigma \equiv \{D_{\sigma R}, D_{\sigma I}\}$ for both samples. The discrete times $t_{\sigma i} \equiv \{t_{\sigma 1}, \dots, t_{\sigma N_\sigma}\}$ are assumed distinct from the sampling times for the data D . In practice the noise sample could be gathered by extending the FID acquisition time, obtaining data after the signal has decayed into the noise, or it could be gathered as a separate data set. When the acquisition time is extended one could incorporate the noise in two different ways, by leaving the noise on the end of the FID (there by calling noise, data) or by making the noise a separate data set. The mathematics will show that these two different sounding procedures have exactly the same effect on the calculation.

The problem is to compute the posterior probability for the amplitude given model Eqs. [1,2], the noise sample D_σ , the data D , and all of the information I . This probability density function is denoted as $P(A|D_\sigma, D, I)$. According to Bayes' theorem [6], it is given by

$$P(A|D_\sigma, D, I) = \frac{P(A|I)P(D_\sigma, D|A, I)}{P(D_\sigma, D|I)} \quad (3)$$

where $P(A|I)$ is the prior probability for the amplitude, the joint marginal probability for the data and the noise sample, $P(D_\sigma, D|I)$ is shown to be a normalization constant, and $P(D_\sigma, D|A, I)$ is the joint marginal probability for the noise and the data samples. A marginal probability density function is one in which one or more nuisance parameters have been removed using the rules of probability theory.

The symbol " I " in the probability density functions represents all of the assumptions that go into the calculation. These assumptions are hypotheses, and like any others appearing in a probability symbol, can be tested using the rules of probability theory. However, in the present calculation they will be taken as given. At present these assumptions include the quadrature nature of the data, the separation of the data into a signal plus additive noise, the single exponentially decaying sinusoidal model, and the simplifying assumption about the noise.

Assigning probabilities

The prior probability for the data and the noise sample (the normalization constant) is addressed first. It may be computed from the joint probability for the noise sample, the data, and the

amplitude:

$$P(D_\sigma, D|I) = \int dA P(D_\sigma, D, A|I) \quad (4)$$

where the sum rule has been applied to the right-hand-side to remove the dependence on the amplitude. The product rule may be used to factor the right-hand-side of Eq. [4] to obtain

$$P(D_\sigma, D|I) = \int dA P(A|I) P(D_\sigma, D|A, I). \quad (5)$$

This is the constant needed to ensure that the total probability is equal to 1.

The prior probability for the amplitude, $P(A|I)$, represents the prior information known about the amplitude before obtaining the data or the noise sample. In most estimation problems, all that is known about the amplitude is that it is bounded and greater than zero. This prior information is vague and uninformative. Any prior which represents this state of knowledge will not have a strong effect on the parameter estimates. Here the prior is taken to be uniform from zero to the upper bound. The upper bound, A_{max} , is the maximum value the digitizer can hold. The prior probability is given by

$$P(A|I) = \begin{cases} \frac{1}{A_{max}} & 0 < A < A_{max} \\ 0 & \text{otherwise} \end{cases}. \quad (6)$$

If the normalization constant (Eq. [5]) is dropped, and the posterior probability is normalized at the end of the calculation, then prior ranges cancel when the posterior probability is normalized. For $P(A|I)$ the prior range is $1/A_{max}$. Assuming normalization will occur at the end of the calculation and dropping all irrelevant constants, the posterior probability for the amplitude is proportional to the joint marginal probability, from Eq. [3] one obtains:

$$P(A|D_\sigma, D, I) \propto P(D_\sigma, D|A, I). \quad (7)$$

The problem has been reduced to finding the joint marginal probability for the data and the noise given the amplitude.

The joint marginal probability, $P(D_\sigma, D|A, I)$, can be computed from the joint probability for the noise sample, the data, and the parameters (the frequency, decay rate constant, and phase):

$$P(D_\sigma, D|A, I) \propto \int d\omega d\alpha d\theta P(D_\sigma, D, \omega, \alpha, \theta|A, I) \quad (8)$$

where the sum rule was applied to remove the functional dependence on the nuisance parameters. Using the product rule, the right-hand-side of this equation may be factored to obtain

$$P(D_\sigma, D|A, I) \propto \int d\omega d\alpha d\theta P(\omega, \alpha, \theta|A, I) P(D_\sigma, D|\omega, \alpha, \theta, A, I) \quad (9)$$

where $P(\omega, \alpha, \theta|A, I)$ is the joint prior probability for the parameters given the amplitude. It represents the prior information available about the parameters before obtaining either the data or noise sample. The joint direct probability for the noise and the data sample, $P(D_\sigma, D|\omega, \alpha, \theta, A, I)$, represents what was learned about the parameters from the data and the noise sample. Combining Eq. [9] and Eq. [7], one obtains

$$P(A|D_\sigma, D, I) \propto \int d\omega d\alpha d\theta P(\omega, \alpha, \theta|A, I) P(D_\sigma, D|\omega, \alpha, \theta, A, I) \quad (10)$$

as the posterior probability for the amplitude. To compute this probability, one must assign both the joint prior probability for the parameters, and the joint direct probability for the data and the noise sample.

To assign the joint prior probability for the parameters, one must state carefully what is known about these parameters. It will be assumed that little prior information is available about these parameters and that they are independent, since there is no physical reason why they should be correlated. This implies the joint prior may be factored to obtain

$$P(\omega, \alpha, \theta|A, I) = P(\omega|I)P(\alpha|I)P(\theta|I). \quad (11)$$

Under this assumption, the joint prior probability for the parameters is the product of the prior probabilities for the individual parameters.

Each of these three priors is assigned separately by stating what is known about the parameters. For uniformly sampled data, frequencies must be in the interval $-\pi \leq \omega \leq \pi$ (in dimensionless units) and a uniform prior will be assigned. This prior is given by

$$P(\omega|I) = \begin{cases} \frac{1}{2\pi} & \text{if } -\pi \leq \omega \leq \pi \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

and indicates ignorance of the frequency in the sense that all values of the frequency within the sampling window are considered equally likely. Note that the prior range ($1/2\pi$) is a constant and will cancel on normalization and can be dropped.

Because sinusoids are periodic under a 2π change in phase, the phase is bounded. No additional prior information is assumed and a uniform prior is assigned to the phase:

$$P(\theta|I) = \begin{cases} \frac{1}{2\pi} & \text{if } -\pi \leq \theta \leq \pi \\ 0 & \text{otherwise} \end{cases}. \quad (13)$$

The prior range ($1/2\pi$) will also cancel when the distribution is normalized, so it will be dropped.

Lastly, the prior for the decay rate constant must be assigned. In this problem, the decay rate constant sets a time scale for the decay. The completely uninformative prior probability for a “scale” parameter is the Jeffreys’ prior [9]. Because this prior is slowly varying compared to the direct probability, the prior will remain nearly constant over the range of values where the direct probability density is sharply peaked as a function of α . The only characteristic of the prior that has a chance of surviving is the value of the prior at the maximum of the likelihood, and this peak value will cancel when the posterior probability density is normalized. Knowing this, the prior for the decay rate constant will be taken as a uniform prior. The difference between using a uniform prior over a Jeffreys’ prior is like the difference in having one more data value, such a difference is negligible, provided that one has enough data to make reasonable estimates [7]. Here the prior will be taken as

$$P(\alpha|I) = \begin{cases} \frac{1}{N} & \text{if } 0 \leq \alpha \leq N \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

which in dimensionless units, indicates that the signal has decayed to $\exp\{-N\}$ when the second data value is acquired and thus easily bounds the decay rate.

With these assumptions, the joint prior probability for the parameters is uniform, and the bounds on these priors will cancel. The problem has been reduced to assigning the joint direct probability for the noise samples and the data given the parameters:

$$P(A|D_\sigma, D, I) \propto \int d\omega d\alpha d\theta P(D_\sigma, D|\omega, \alpha, \theta, A, I). \quad (15)$$

The joint direct probability, $P(D_\sigma, D|\omega, \alpha, \theta, A, I)$, must be further simplified before it can be assigned. The product rule is used to factor the right-hand-side Eq. [15] to obtain

$$P(A|D_\sigma, D, I) \propto \int d\omega d\alpha d\theta P(D_\sigma|\omega, \alpha, \theta, A, I)P(D|D_\sigma, \omega, \alpha, \theta, A, I) \quad (16)$$

where $P(D|D_\sigma, \alpha, \theta, A, I)$ is the probability for the data given the noise sample and the parameters. As a simplifying assumption, the conditional probability for the data given the noise sample will be taken to be independent of the noise sample, i.e., it is assumed that $P(D|D_\sigma, \omega, \alpha, \theta, I) = P(D|\omega, \alpha, \theta, I)$. Additionally, the probability for the noise sample will be taken as $P(D_\sigma|\omega, \alpha, \theta, A, I) = P(D_\sigma|I)$, simply because the noise sample does not contain a signal and cannot depend on these parameters. Thus, the posterior probability density for the amplitude becomes

$$P(A|D_\sigma, D, I) \propto \int d\omega d\alpha d\theta P(D_\sigma|I)P(D|\omega, \alpha, \theta, A, I) \quad (17)$$

where $P(D_\sigma|I)$ is the prior probability for the noise sample, and $P(D|\omega, \alpha, \theta, A, I)$ is the direct probability for the data given the parameters.

The symbol, D , stands for both the real and imaginary data and the noise sample, while the symbol, D_σ , stands for the real and imaginary noise samples. Making this substitution on the right-hand-side of equation [17] yields

$$P(A|D_\sigma, D, I) \propto \int d\omega d\alpha d\theta P(D_{\sigma R}, D_{\sigma I}|I)P(D_R, D_I|\omega, \alpha, \theta, A, I) \quad (18)$$

as the posterior probability for the amplitude.

If the product rule is applied to the first term in the right-hand-side of this equation, $P(D_{\sigma R}, D_{\sigma I}|I)$, one obtains

$$P(D_{\sigma R}, D_{\sigma I}|I) = P(D_{\sigma R}|I)P(D_{\sigma I}|D_{\sigma R}, I). \quad (19)$$

Since the two channels of an NMR spectrometer give projections onto orthogonal functions, knowing the imaginary noise sample $D_{\sigma I}$ will not help in assigning the probability for the real noise sample: $P(D_{\sigma I}|D_{\sigma R}, I) = P(D_{\sigma I}|I)$, and the posterior probability for the amplitude becomes

$$P(A|D_\sigma, D, I) \propto \int d\omega d\alpha d\theta P(D_{\sigma R}|I)P(D_{\sigma I}|I)P(D_R, D_I|\omega, \alpha, \theta, A, I). \quad (20)$$

Similarly,

$$P(D_R, D_I|\omega, \alpha, \theta, A, I) = P(D_R|\omega, \alpha, \theta, A, I)P(D_I|\omega, \alpha, \theta, A, I). \quad (21)$$

With these assumptions the joint probability of the data factors and the posterior probability for the amplitude is given by

$$P(A|D_\sigma, D, I) \propto \int d\omega d\alpha d\theta P(D_{\sigma R}|I)P(D_{\sigma I}|I)P(D_R|\omega, \alpha, \theta, A, I)P(D_I|\omega, \alpha, \theta, A, I) \quad (22)$$

where $P(D_R|\omega, \alpha, \theta, A, I)$ is the direct probability for the real data, and $P(D_I|\omega, \alpha, \theta, A, I)$ is the direct probability for the imaginary data.

The posterior probability for the amplitude has now been sufficiently simplified to permit assignment of the various terms. To do this assignment, notice that Eqs. [1] and [2] constitute a definition of what is meant by noise in this calculation. The direct probability for the data given the parameters *is* the noise prior probability given the parameters. Before these probability density functions can be assigned, a prior probability for the noise must be assigned. To do this, one must state exactly what assumptions are being made about the noise. As in previous works [1, 2, 3, 4, 7] the noise is assumed to carry a finite, but unknown average power. A maximum entropy calculation [5, 10] using this assumption results in the assignment of a Gaussian for the noise prior probability density:

$$P(e_1, \dots, e_N|\sigma, I) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp \left\{ -\sum_{i=1}^N \frac{e_i^2}{2\sigma^2} \right\}. \quad (23)$$

The direct probability for obtaining the real data is then given by

$$P(D_R|\sigma, \omega, \alpha, \theta, A, I) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp \left\{ - \sum_{i=1}^N \frac{[d_R(t_i) - A \cos(\omega t_i + \theta)e^{-\alpha t_i}]^2}{2\sigma^2} \right\} \quad (24)$$

where the notation has been adjusted: first, e_1, \dots, e_N , was replaced by D_R to indicate it is the direct probability for the noise in the real data sample. Second the standard deviation for the noise, σ , was added to the probability density function to indicate that it is a known quantity. Later, the product rule of probability theory will be used to remove the dependence on σ . The direct probability for obtaining the imaginary data is given by

$$P(D_I|\sigma, \omega, \alpha, \theta, A, I) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp \left\{ - \sum_{i=1}^N \frac{[d_I(t_i) + A \sin(\omega t_i + \theta)e^{-\alpha t_i}]^2}{2\sigma^2} \right\}. \quad (25)$$

The prior probability for obtaining the real noise sample is given by

$$P(D_{\sigma R}|\sigma, I) = (2\pi\sigma^2)^{-\frac{N_\sigma}{2}} \exp \left\{ - \sum_{i=1}^{N_\sigma} \frac{d_{\sigma R}(t_{\sigma i})^2}{2\sigma^2} \right\}, \quad (26)$$

and the prior probability for the imaginary noise sample is given by

$$P(D_{\sigma I}|\sigma, I) = (2\pi\sigma^2)^{-\frac{N_\sigma}{2}} \exp \left\{ - \sum_{i=1}^{N_\sigma} \frac{d_{\sigma I}(t_{\sigma i})^2}{2\sigma^2} \right\}. \quad (27)$$

Combining these four terms, the posterior probability for the amplitude is given by

$$\begin{aligned} P(A|\sigma, D_\sigma, D, I) &\propto \int d\omega d\alpha d\theta (2\pi\sigma^2)^{-N-N_\sigma} \exp \left\{ - \frac{2N\overline{d_\sigma^2} + 2N\overline{d^2}}{2\sigma^2} \right\} \\ &\times \exp \left\{ \frac{2Ad_R \cdot \cos(\omega t + \theta)e^{-\alpha t} - 2Ad_I \cdot \sin(\omega t + \theta)e^{-\alpha t}}{2\sigma^2} \right\} \\ &\times \exp \left\{ - \frac{A^2 e^{-\alpha t} \cdot e^{-\alpha t}}{2\sigma^2} \right\} \end{aligned} \quad (28)$$

where the identity $\sin^2(x) + \cos^2(x) = 1$ was used. The mean-square of the noise sample, $\overline{d_\sigma^2}$, is defined as

$$\overline{d_\sigma^2} \equiv \frac{1}{2N_\sigma} \sum_{i=1}^{N_\sigma} d_{\sigma R}(t_{\sigma i})^2 + d_{\sigma I}(t_{\sigma i})^2 \quad (29)$$

and the mean-square of the data value, $\overline{d^2}$, is

$$\overline{d^2} \equiv \frac{1}{2N} \sum_{i=1}^N d_R(t_i)^2 + d_I(t_i)^2. \quad (30)$$

The notation “.” in Eq. [28] means the functions are to be multiplied and a sum over discrete times performed, for example

$$d_R \cdot \cos(\omega t + \theta)e^{-\alpha t} \equiv \sum_{i=1}^N d_R(t_i) \cos(\omega t_i + \theta)e^{-\alpha t_i}. \quad (31)$$

Removing the phase

The next step in the formal derivation of the posterior probability for the amplitude is to perform the integrals over the three nuisance parameters. This is possible for the phase, but not for the frequency and decay rate constants. In the next few paragraphs the integral over the phase will be done; two different approximations will then be derived for the remaining integrals. Removing the phase has much in common with the work of Jaynes [5], which addressed a chirped sinusoidal model. The integral over the phase is evaluated by replacing the sine and cosine terms in Eq. [28] by the appropriate trigonometric relations for the sum of two angles. The posterior probability for the amplitude becomes

$$P(A|\sigma, D_\sigma, D, I) \propto \int d\omega d\alpha d\theta \sigma^{-2N-2N_\sigma} \exp \left\{ -\frac{2N_\sigma \overline{d_\sigma^2} + 2N \overline{d^2}}{2\sigma^2} \right\} \times \exp \left\{ \frac{2A[R(\omega, \alpha) \cos(\theta) - I(\omega, \alpha) \sin(\theta)] - A^2 C(0, 2\alpha)}{2\sigma^2} \right\} \quad (32)$$

where $R(\omega, \alpha)$ and $I(\omega, \alpha)$ are defined as

$$R(\omega, \alpha) \equiv d_R \cdot \cos(\omega t) e^{-\alpha t} - d_I \cdot \sin(\omega t) e^{-\alpha t} \quad (33)$$

and

$$I(\omega, \alpha) \equiv d_R \cdot \sin(\omega t) e^{-\alpha t} + d_I \cdot \cos(\omega t) e^{-\alpha t}. \quad (34)$$

When data are uniformly sampled, ω is taken on a discrete grid, $\omega_i = 2\pi i/N$ ($i = 0, 1, \dots, N-1$), the functions $R(\omega_i, \alpha)$ and $I(\omega_i, \alpha)$ are the real and imaginary parts of the discrete Fourier transform of the complex FID data when the data have been multiplied by a decaying exponential of decay rate α .

The function $C(x, y)$ is defined as

$$C(x, y) \equiv \sum_{i=1}^N \cos(x t_i) e^{-y t_i}. \quad (35)$$

If uniform sampling is used, then $C(x, y)$ may be expressed in closed form [4]. Taking the sampling times to be $t_i = \{0, 1, \dots, N-1\}$, and assuming that the frequencies and decay rate constants are measured in radians, the sum appearing in $C(x, y)$ may be evaluated to obtain

$$C(x, y) = \frac{1 - \cos(x) e^{-y} - \cos(Nx) e^{-Ny} + \cos[(N-1)x] e^{-(N+1)y}}{1 - 2 \cos(x) e^{-y} + e^{-2y}}. \quad (36)$$

If nonuniform sampling is being used, then $C(x, y)$, $R(\omega, \alpha)$, and $I(\omega, \alpha)$ must be computed from their definitions.

The posterior probability for the amplitude, Eq. [32], may be further simplified by combining the terms containing $\sin(\theta)$ and $\cos(\theta)$ into a single term containing a cosine with a phase, one obtains:

$$P(A|\sigma, D_\sigma, D, I) \propto \int d\alpha d\omega \sigma^{-2N-2N_\sigma} \exp \left\{ -\frac{2N_\sigma \overline{d_\sigma^2} + 2N \overline{d^2} + A^2 C(0, 2\alpha)}{2\sigma^2} \right\} \times \int_0^{2\pi} d\theta \exp \left\{ \frac{A \sqrt{R(\omega, \alpha)^2 + I(\omega, \alpha)^2} \cos(\theta + \chi)}{\sigma^2} \right\} \quad (37)$$

where

$$\chi = \tan^{-1} \left[-\frac{I(\omega, \alpha)}{R(\omega, \alpha)} \right]. \quad (38)$$

The integral over the phase may now be recognized as the integral representation for an I_0 Bessel function. Evaluating the integral one obtains

$$\begin{aligned}
P(A|\sigma, D_\sigma, D, I) &\propto \int d\alpha d\omega \sigma^{-2N-2N_\sigma} \exp \left\{ -\frac{2N_\sigma \overline{d_\sigma^2} + 2N \overline{d^2} + A^2 C(0, 2\alpha)}{2\sigma^2} \right\} \\
&\times I_0 \left(\frac{A \sqrt{R(\omega, \alpha)^2 + I(\omega, \alpha)^2}}{\sigma^2} \right).
\end{aligned} \tag{39}$$

Before proceeding with the calculation, note that the argument of the Bessel function is large near the location of the maximum of the joint conditional probability for the amplitude, frequency, and decay rate constant (the integrand in Eq. [39]). Here that maximum (as a function of frequency and decay rate constant) is located essentially at the location of the maximum of a power spectrum. At that maximum of the joint conditional probability, the argument of the Bessel function is given approximately by $A\hat{A}/2\hat{\alpha}\sigma^2$, where \hat{A} and $\hat{\alpha}$ are the “true” amplitude and dimensionless decay rate constant, Eq. [62] below. If the sinusoidal signal decays away to $3T_2^*$ in the run of the data and there are $N = 1000$ complex data values, $\hat{\alpha} \approx 0.003$ and the argument of the Bessel function is approximately $166A\hat{A}/\sigma^2$. If the time domain peak signal-to-noise ratio is as high as 2, the argument of the Bessel function (at the maximum) is approximately $166 \times (2)^2 \approx 666$, so the argument of the Bessel function is large. The Bessel function may then be replaced by its large argument approximation over the region where the integrand is large:

$$I_0(x) \approx \frac{e^{-x}}{\sqrt{2\pi x}} \tag{40}$$

and the posterior probability for the amplitude is given approximately by

$$\begin{aligned}
P(A|\sigma, D_\sigma, D, I) &\approx \int d\alpha d\omega d\sigma \sigma^{-2N-2N_\sigma-1} \exp \left\{ -\frac{2N_\sigma \overline{d_\sigma^2} + 2N \overline{d^2}}{2\sigma^2} \right\} \\
&\times \exp \left\{ \frac{2A \sqrt{R(\omega, \alpha)^2 + I(\omega, \alpha)^2} - A^2 C(0, 2\alpha)}{2\sigma^2} \right\}.
\end{aligned} \tag{41}$$

Notice that a factor, $A^{-\frac{1}{2}} [R(\omega, \alpha)^2 + I(\omega, \alpha)^2]^{-\frac{1}{4}}$, was dropped. Dropping this factor *does not significantly affect* the results of the calculation. At the maximum of the posterior probability, the dropped term is given approximately by $\hat{A}^{-1} C(0, 2\hat{\alpha})^{-\frac{1}{2}}$ where \hat{A} and $\hat{\alpha}$ are the “true” amplitudes and decay rate constants in the data. As has been previously noted, dropping a term of this order is like the difference between using a uniform and Jeffreys’ prior, negligible provided one has sufficient data to make reasonable estimates.

Removing The Standard Deviation Of The Noise

One would like a formulation of this problem that does not depend on knowing the standard deviation of the noise. The standard deviation can be removed by applying the product rule and sum rule. Symbolically the posterior probability for the amplitude, independent of the standard deviation of the noise, is given by

$$P(A|D_\sigma, D, I) = \int d\sigma P(\sigma|I) P(A|\sigma, D_\sigma, D, I) \tag{42}$$

where $P(\sigma|I)$ represents what was known about the standard deviation for the noise before obtaining the noise or the data samples and $P(A|\sigma, D_\sigma, D, I)$ is given by Eq. [41].

The standard deviation is a scale parameter and the completely uninformative prior probability for a scale parameter is the Jeffreys’ prior [9]. As has been noted previously [1, 7], this prior is not

strictly speaking a prior probability at all, rather using it corresponds to using a sequence of proper priors in the limit of infinite uncertainty in the variance. In many problems use of an improper prior must be avoided. However, here it will not affect the results, because the Gaussian cutoff in the exponent is so sharp that one will always have convergent integrals. Multiplying the posterior probability for the amplitude, Eq. [39], by the Jeffreys' prior ($1/\sigma$), and integrating one obtains

$$P(A|D_\sigma, D, I) \propto \int d\alpha d\omega \left[1 + \frac{A^2 C(0, 2\alpha) - 2A \sqrt{R(\omega, \alpha)^2 + I(\omega, \alpha)^2}}{2N_\sigma d_\sigma^2 + 2Nd^2} \right]^{\frac{1-2N-2N_\sigma}{2}}. \quad (43)$$

The integral was evaluated by introducing the change of variables $v = 1/\sigma^2$, which transforms the integral into a Gamma integral. The resulting Gamma function was a constant and was therefore dropped.

Approximations

When the variance of the noise is known, the posterior probability for the amplitude is given exactly by Eq. [39], and approximately by Eq. [41]. When the variance of the noise is unknown it is given approximately by Eq. [43]. These results require that two additional integrals be performed. To date the author has not been able to evaluate these integrals in closed form. However, they can be readily approximated. Here two different approximations are useful. The first approximation replaces the integrals by sums. The second approximation uses a local Gaussian approximation.

Replacement of the integrals by sums is trivial. First, one bounds the region in parameter space where the joint probability for the frequency and decay rate constant is maximum. The only requirement in picking the bounds is that the contribution to the integral of the quantity in brackets (Eq. [43]) should become negligible outside the bounds. Second, one replaces the integrals in Eq. [43] by sums:

$$P(A|D_\sigma, D, I) \propto \sum_{j,k} w_{jk} \left[1 + \frac{A^2 C(0, 2\alpha_j) - 2A \sqrt{R(\omega_k, \alpha_j)^2 + I(\omega_k, \alpha_j)^2}}{2N_\sigma d_\sigma^2 + 2Nd^2} \right]^{\frac{1-2N-2N_\sigma}{2}} \quad (44)$$

where w_{jk} are the weights associated with the numerical quadrature. If there is good evidence for a signal, one expects nearly Gaussian behavior, and a good candidate for the numerical quadrature might be Gauss-Hermit quadrature, which absorbs the Gaussian. The sums are from a lower to upper bound. If Gauss-Hermit quadrature is used, the steps in the decay rate will be nonuniformly spaced. However, if the discrete Fourier transform is used to compute the projections of the data onto the model, the numerical quadrature for the frequency must assume a uniformly sampled abscissa.

The second approximation assumes high signal-to-noise data. In this approximation, the exponent in the posterior probability density (Eq. [41]) is expanded in a Taylor series about the maximum. The Taylor expansion is truncated at second order, and the integrals are then performed. In this calculation the variance of the noise will initially be assumed to be known. To perform the calculation, the two nonlinear parameters will be denoted as $\theta_1 = \omega$ and $\theta_2 = \alpha$. The values of θ_1 and θ_2 that maximize the joint posterior probability for the amplitude and the frequency, will be denoted as $\hat{\theta}_1 = \hat{\omega}$, and $\hat{\theta}_2 = \hat{\alpha}$. Defining the function

$$\overline{h^2} = A^2 C(0, 2\theta_2) - 2A \sqrt{R(\theta_1, \theta_2)^2 + I(\theta_1, \theta_2)^2} \quad (45)$$

and the matrix b_{jk} as

$$b_{jk} = -\frac{1}{2} \left. \frac{\partial^2 \overline{h^2}}{\partial \theta_j \partial \theta_k} \right|_{\hat{\theta}_1, \hat{\theta}_2} \quad j, k = \{1 \text{ or } 2\} \quad (46)$$

the posterior probability for the amplitude is given approximately by

$$P(A|\sigma, D_\sigma, D, I) \propto \sigma^{-2N_\sigma - 2N - 1} \exp \left\{ -\frac{2N_\sigma \overline{d_\sigma^2} + 2N \overline{d^2} - \overline{h^2}}{2\sigma^2} \right\} \Bigg|_{\hat{\theta}_1, \hat{\theta}_2} \quad (47)$$

$$\times \int_{-\infty}^{\infty} d\theta_1 \int_{-\infty}^{\infty} d\theta_2 \exp \left\{ -\sum_{j=1}^2 \sum_{k=1}^2 \frac{(\hat{\theta}_j - \theta_j)(\hat{\theta}_k - \theta_k) b_{jk}}{2\sigma^2} \right\}.$$

Evaluating the integrals one has

$$P(A|\sigma, D_\sigma, D, I) \propto \sigma^{-2N_\sigma - 2N - 3} [b_{11}b_{22} - b_{12}]^{-\frac{1}{2}} \exp \left\{ -\frac{2N_\sigma \overline{d_\sigma^2} + 2N \overline{d^2} - \overline{h^2}}{2\sigma^2} \right\} \Bigg|_{\hat{\theta}_1, \hat{\theta}_2}. \quad (48)$$

The functions b_{11} , b_{22} and b_{12} are given by

$$b_{11} = B \left[R(\theta_1, \theta_2) \frac{\partial^2 R(\theta_1, \theta_2)}{\partial \theta_1^2} + I(\theta_1, \theta_2) \frac{\partial^2 I(\theta_1, \theta_2)}{\partial \theta_2^2} \right], \quad (49)$$

$$b_{22} = B \left[R(\theta_1, \theta_2) \frac{\partial^2 R(\theta_1, \theta_2)}{\partial \theta_2^2} + I(\theta_1, \theta_2) \frac{\partial^2 I(\theta_1, \theta_2)}{\partial \theta_1^2} \right] - \frac{A^2}{2} \frac{\partial^2 C(0, 2\theta_2)}{\partial \theta_2^2}, \quad (50)$$

$$b_{12} = b_{21} = B \left[R(\theta_1, \theta_2) \frac{\partial^2 R(\theta_1, \theta_2)}{\partial \theta_1 \partial \theta_2} + I(\theta_1, \theta_2) \frac{\partial^2 I(\theta_1, \theta_2)}{\partial \theta_1 \partial \theta_2} \right], \quad (51)$$

and

$$B \equiv \frac{A}{\sqrt{R(\theta_1, \theta_2)^2 + I(\theta_1, \theta_2)^2}} \quad (52)$$

where the mixed partial of $R(\theta_1, \theta_2)$ and $I(\theta_1, \theta_2)$ may be computed from Eqs. [33,34]. Since the quantity $b_{11}b_{22} - b_{12}$ in Eq. [48] contains the amplitude to a power of two, the term $[b_{11}b_{22} - b_{12}]^{-\frac{1}{2}}$ varies like $1/A$. But two other terms of this same order were dropped earlier (one in Eqs. [14] and one in [41]), because they make negligible contributions to the results. Dropping these terms, yields

$$P(A|\sigma, D_\sigma, D, I) \propto \sigma^{-2N_\sigma - 2N - 3} \exp \left\{ -\frac{2N_\sigma \overline{d_\sigma^2} + 2N \overline{d^2} - \overline{h^2}}{2\sigma^2} \right\} \Bigg|_{\hat{\omega}, \hat{\alpha}} \quad (53)$$

as the posterior probability for the amplitude when the variance of the noise is known and

$$P(A|D_\sigma, D, I) \propto \left[1 + \frac{A^2 C(0, 2\alpha) - 2A \sqrt{R(\omega, \alpha)^2 + I(\omega, \alpha)^2}}{2N_\sigma \overline{d_\sigma^2} + 2N \overline{d^2}} \right]^{\frac{3-2N-2N_\sigma}{2}} \Bigg|_{\hat{\omega}, \hat{\alpha}} \quad (54)$$

when the variance of the noise is unknown. Note that this approximation has restored a single-term form of the low signal-to-noise approximation, Eq. [44]. In doing the Gaussian approximation the integrals over the frequency and decay rate constants behaved as if they were delta functions. The frequency delta function constrains the frequency ω to the peak value of the power spectrum $\hat{\omega}$; while the delta function for the decay rate constant constrains δ to the ‘‘matched filter’’ value $\hat{\alpha}$.

Precision And Accuracy Estimates

In the previous sections, the posterior probability for the amplitude independent of the phase, decay rate constant, frequency, and variance of the noise was computed. Note that this is a continuous

function of A – not a point estimate. Here those results are applied to derive an estimate of the expected value of the amplitude and its uncertainty. The results will be expressed as a (mean \pm standard deviation) approximation. This calculation is done assuming high signal-to-noise data. Two forms of the calculation will be done: in the first form, the standard deviation of the noise will be assumed known; and in the second, it will be removed as a nuisance parameter.

The (mean \pm standard deviation) estimate of the amplitude requires one to compute both the expected amplitude and expected mean-square of the amplitude. The expected value of a parameter is a weighted average of all possible values of the parameter. The weights for each value of A are the posterior probability associated with this hypothesis, so the expected value of the amplitude is a weighted average over hypotheses. It is computed by multiplying the posterior probability by the amplitude and integrating:

$$\langle A \rangle = \frac{\int dA A P(A|\sigma, D, D_\sigma, I)}{\int dA P(A|\sigma, D, D_\sigma, I)}. \quad (55)$$

The extra factor (the denominator) is needed to normalize the posterior probability. The posterior probability for the amplitude is replaced by Eq. [53] and the indicated integrals are evaluated. This gives

$$\langle A \rangle = \frac{\sqrt{R(\hat{\omega}, \hat{\alpha})^2 + I(\hat{\omega}, \hat{\alpha})^2}}{C(0, 2\hat{\alpha})} \quad (56)$$

as the estimated amplitude. The expected square of the amplitude is given by

$$\langle A^2 \rangle = \frac{\int dA A^2 P(A|\sigma, D, D_\sigma, I)}{\int dA P(A|\sigma, D, D_\sigma, I)}. \quad (57)$$

Evaluating this integral gives

$$\langle A^2 \rangle = \frac{R(\hat{\omega}, \hat{\alpha})^2 + I(\hat{\omega}, \hat{\alpha})^2}{C(0, 2\hat{\alpha})^2} + \frac{\sigma^2}{C(0, 2\hat{\alpha})}. \quad (58)$$

The standard deviation for the width of the posterior is

$$\sqrt{\langle A^2 \rangle - \langle A \rangle^2} = \frac{\sigma}{\sqrt{C(0, 2\hat{\alpha})}}. \quad (59)$$

And the amplitude estimate is given by

$$(A)_{est} = \frac{\sqrt{R(\hat{\omega}, \hat{\alpha})^2 + I(\hat{\omega}, \hat{\alpha})^2}}{C(0, 2\hat{\alpha})} \pm \frac{\sigma}{\sqrt{C(0, 2\hat{\alpha})}}. \quad (60)$$

The frequencies and decay rate constants will be expressed in dimensional units, but first note that the function $C(0, 2\alpha)$ is given approximately by

$$\begin{aligned} C(0, 2\alpha) &= \sum_{i=1}^N e^{-2\alpha t} \\ &\approx \int_0^N dx e^{-2\alpha x} \\ &= \frac{1 - e^{-2\alpha N}}{2\alpha} \\ &\approx \frac{1}{2\alpha} \end{aligned} \quad (61)$$

provided that $2\alpha N$ is large, i.e., the signal decays away in the sampling time. So the accuracy estimate is given approximately by

$$(A)_{est} \approx 2\hat{\alpha} \sqrt{R(\hat{\omega}, \hat{\alpha})^2 + I(\hat{\omega}, \hat{\alpha})^2} \pm \sigma \sqrt{2\hat{\alpha}}. \quad (62)$$

The conversion to the more familiar dimensional quantities is given by

$$f = \frac{\omega N}{2\pi A_T} \quad \text{and} \quad R = \frac{\alpha N}{\pi A_T} \quad (63)$$

where f is the frequency in Hertz, A_T is the total acquisition time, and R is the observed decay rate constant in Sec.⁻¹. The precision estimate is given by

$$(A)_{est} = \frac{2\pi A_T R_2^*}{N} \sqrt{R(\hat{f}, \hat{R}_2^*)^2 + I(\hat{f}, \hat{R}_2^*)^2} \pm \sigma \sqrt{\frac{2\pi A_T \hat{R}_2^*}{N}} \quad (64)$$

where \hat{f} and \hat{R}_2^* are the dimensional frequency and decay rate constant.

This (mean \pm standard deviation) estimate has three interesting features. First, it demonstrates that for a fixed amplitude the precision estimate of the amplitude becomes larger (more uncertain) as the decay rate constant increases. This agrees well with one's intuition that the more rapidly a line decays, the less accurately one should be able determine its value. Second, if the noise is uncorrelated and *if one can hold the time-domain signal-to-noise level constant*, the precision estimate can be halved by sampling four times faster. Note, however, that automatic adjustment of audio filters to allow high frequencies to be measured increases the noise level [12, 13] and cancels the effect. Third, the precision estimate implies that the frequency domain peak signal-to-RMS-noise ratio changes as a function of sampling rate. Sampling four times faster does not change the amplitude of the resonance – so the amplitude estimate must be constant to within the noise. But 4 times as many data values have been gathered, $N \rightarrow 4N$. For the amplitude to remain constant, the absolute value spectrum (the radical in the amplitude estimate) must be four times larger. This strongly suggest that if the noise is uncorrelated and *if the time domain signal-to-noise level can be held constant*, the frequency domain signal-to-noise ratio can be increased by increasing the sampling rate.

This effect can be easily verified. A straight-forward calculation shows that the discrete Fourier transform of a perfectly phased exponentially decaying sinusoid is given by

$$\mathcal{F}\{\hat{A}[\cos(\hat{\omega}t_i) - i \sin(\hat{\omega}t_i)]e^{-\hat{\alpha}t_i}\} = \hat{A}C(\hat{\omega} - \omega, \hat{\alpha}) \quad (65)$$

where \mathcal{F} signifies the discrete Fourier transform. To obtain the peak signal-to-RMS-noise this must be evaluated at $\omega = \hat{\omega}$ and the peak value in the discrete Fourier transform is approximately given by

$$\mathcal{F}\{\hat{A}[\cos(\hat{\omega}t_i) - i \sin(\hat{\omega}t_i)]e^{-\hat{\alpha}t_i}\} \Big|_{\omega=\hat{\omega}} \approx \frac{\hat{A}}{\hat{\alpha}}, \quad (66)$$

where the same integral approximation was used in Eq. [61]. Converting to dimensional quantities one has

$$\begin{aligned} \mathcal{F}\{\hat{A}[\cos(\hat{\omega}t_i) - i \sin(\hat{\omega}t_i)]e^{-\hat{\alpha}t_i}\} \Big|_{\omega=\hat{\omega}} &\approx \frac{\hat{A}N}{\pi R_2^* A_T} \\ &= \frac{\hat{A}S_w}{\pi R_2^*} \end{aligned} \quad (67)$$

where S_w is the sampling rate (the number of complex points sampled per second).

To verify the effect one also needs the RMS noise level in the frequency domain. This is simply related to the RMS noise level in the time domain. If σ represent the RMS noise level in the time

domain, then

$$\begin{aligned} \text{RMS noise level Frequency domain} &= \sqrt{N}\sigma \\ &= \sqrt{S_w A_T}\sigma \end{aligned} \quad (68)$$

and the RMS peak signal-to-noise ratio in the frequency domain is given by

$$\text{RMS S/N Ratio Frequency domain} \approx \frac{\hat{A}}{\pi R_2^* \sigma} \sqrt{\frac{S_w}{A_T}} \quad (69)$$

which demonstrates that the frequency domain signal-to-noise level scales like the square root of the sampling rate. So even if one's only analysis tool is the discrete Fourier transform, this effect is still present; *provided one can keep the time domain signal-to-noise constant.*

In the previous discussion, the precision of the Bayesian estimate was derived and it was shown how this estimate depends on the experimental parameters. In next few paragraphs it will be demonstrated that the Bayesian estimate is accurate in the sense that as the noise goes to zero the parameters estimated go the “true” values. To see this, suppose the real data contains a signal of the form

$$d_R(t_i) = [\hat{A}_1 \cos(\hat{f}t) + \hat{A}_2 \sin(\hat{f}t)] \exp\{-\hat{R}_2^* t\} + e_i, \quad (70)$$

then the imaginary data will then be given by

$$d_I(t_i) = [-\hat{A}_1 \sin(\hat{f}t) + \hat{A}_2 \cos(\hat{f}t)] \exp\{-\hat{R}_2^* t\} + e_i \quad (71)$$

where \hat{A}_1 , and \hat{A}_2 are the true amplitudes of the sine and cosine components of the signal, \hat{f} , and \hat{R}_2^* are the frequency and decay rate constants of the signal. The functions $R(f, R_2^*)$ and $I(f, R_2^*)$ are given approximately by

$$R(f, R_2^*) \approx \hat{A}_1 C(\hat{f} - f, \hat{R}_2^* + R_2^*) \quad \text{and} \quad I(f, R_2^*) \approx \hat{A}_2 C(\hat{f} - f, \hat{R}_2^* + R_2^*), \quad (72)$$

provided the standard deviation of the noise is small. The estimated amplitude becomes

$$(A)_{est} = \sqrt{\hat{A}_1^2 + \hat{A}_2^2} \pm \sigma \sqrt{\frac{2\pi A_T \hat{R}_2^*}{N}}. \quad (73)$$

This shows that the estimate is precise and accurate, in the sense that as the noise goes to zero, the amplitude estimate goes to the true amplitude of the sinusoid and the uncertainty in the estimate goes to zero.

Two assumptions were made in doing these calculations: first, it was assumed that the frequency and the decay rate constant were well determined by the data, so the joint posterior probability was very sharply peaked. This made it possible to use the Gaussian approximation and simplify the results. Second, it was assumed that the standard deviation of the noise was known from prior information. By making this assumption, the presence of the noise sample did not affect either the amplitude estimate or the precision estimate. This is expected; if the noise standard deviation is known, gathering a noise sample cannot help one define it better.

Suppose the noise standard deviation was not known. How would this affect the estimates? The uncertainty estimate was computed in two basic steps first the expected amplitude was computed. This expected amplitude consists of two terms: (i) the projection of the model onto the signal, and (ii) the projection of the model onto the noise. But the average projection of the model onto the noise is zero, so the expected amplitude does not depend on the variance of the noise. But the expected mean-square of the amplitude, Eq. [57], does depend on the noise standard deviation because the noise carried a finite power, and thus the uncertainty in the estimated amplitude does depend on the noise standard deviation.

Repeating the calculation for the mean-square and removing the standard deviation of the noise as a nuisance parameter (using a Jeffreys' prior), one obtains

$$\langle A^2 \rangle = \frac{R(\hat{\omega}, \hat{\alpha})^2 + I(\hat{\omega}, \hat{\alpha})^2}{C(0, 2\hat{\alpha})} + \frac{\langle \sigma^2 \rangle}{C(0, 2\hat{\alpha})} \quad (74)$$

where the expected value of the variance is given by

$$\langle \sigma^2 \rangle = \frac{N\overline{d^2} + N_\sigma\overline{d_\sigma^2} - [R(\hat{\omega}, \hat{\alpha})^2 + I(\hat{\omega}, \hat{\alpha})^2]/2C(0, 2\hat{\alpha})}{N + N_\sigma - 2}. \quad (75)$$

The (mean \pm standard deviation) approximation for the amplitude is given by

$$(A)_{est} = \frac{\sqrt{R(\hat{\omega}, \hat{\alpha})^2 + I(\hat{\omega}, \hat{\alpha})^2}}{C(0, 2\hat{\alpha})} \pm \sqrt{\frac{N\overline{d^2} + N_\sigma\overline{d_\sigma^2} - [R(\hat{\omega}, \hat{\alpha})^2 + I(\hat{\omega}, \hat{\alpha})^2]/2C(0, 2\hat{\alpha})}{C(0, 2\hat{\alpha})[N + N_\sigma - 2]}}. \quad (76)$$

This estimate has a number of interesting features. First, as was previously noted, the estimated amplitude does not depend on whether or not a noise sample was gathered. Second, if no noise sample is gathered, $N_\sigma = 0$, the noise variance is estimated to be the mean-square residual. Thus everything probability theory cannot place into the signal is placed into the noise: the accuracy estimates are as wide as is consistent with the data and the model. Second, it does not matter how the noise sample is gathered: by either extending acquisition of the data D or by gathering a separate noise sample D_σ . Both procedures gives exactly the same results. Although an extended signal is more likely to have the same noise characteristics as the signal than a sample gathered at a different time (instrument drifts). To see this observe that when extending data acquisition, after the signal has decayed into the noise, the projection of the model onto the signal is zero so the sufficient statistic is unaffected by extending the noise sample. But the number of complex data values and the total length of the data sample have both increased just as if a noise sample had been gathered. So the two ways of incorporating noise yield the same results. Third, gathering a noise sample will help improve the precision estimates *if simple models are used to analyze complex spectra*. Simple models do not fit all the features in the data and $\langle \sigma^2 \rangle$ is estimated to be large, if $N_\sigma = 0$. But if one gathers a noise sample, the uncertainty in the estimate will become smaller because the estimated noise variance will asymptotically approach the case when the variance is known from prior information. The asymptotic value is approached when $N_\sigma\overline{d_\sigma^2} \gg N\overline{d^2}$ and the precision estimate becomes

$$\sqrt{\frac{N\overline{d^2} + N_\sigma\overline{d_\sigma^2} - [R(\hat{\omega}, \hat{\alpha})^2 + I(\hat{\omega}, \hat{\alpha})^2]/2C(0, 2\hat{\alpha})}{C(0, 2\hat{\alpha})[N + N_\sigma - 2]}} \approx \sqrt{\frac{\overline{d_\sigma^2}}{C(0, 2\hat{\alpha})}}. \quad (77)$$

The amplitude estimate becomes

$$(A)_{est} = \frac{\sqrt{R(\hat{\omega}, \hat{\alpha})^2 + I(\hat{\omega}, \hat{\alpha})^2}}{C(0, 2\hat{\alpha})} \pm \sqrt{\frac{\overline{d_\sigma^2}}{C(0, 2\hat{\alpha})}} \quad (78)$$

and the expected standard deviation for the noise is determined from the prior information, $\langle \sigma \rangle = \overline{d_\sigma}$. The amplitude estimate reduces to that found when the standard deviation was known from prior information, Eq. [73].

Frequency and Amplitude Estimation – Low Signal-To-Noise

Two examples will be given. In the first example, noisy, experimental NMR FID data will be analyzed to demonstrates probability theory's ability to estimate both frequencies and amplitudes

under low signal-to-noise conditions. In the second example, it is demonstrated that the posterior probability for the amplitude of an exponentially decaying sinusoid may be used even when the data contain multiple frequencies, provided that the frequencies are well separated in the discrete Fourier transform.

The data are a ^1H FID of an HOD sample containing a small amount of GdCl_3 (a relaxation agent). The linewidth of the single HOD resonance is approximately 830 Hz. The data were sampled for 1.2368 seconds with a sweepwidth of 20,000 Hz. There are a total of 24736 complex data values in the FID. The ^1H FID was collected approximately -1000 Hz off resonance. Repetitive signal averaging was used to give an initial RMS peak time-domain signal-to-noise ratio of greater than 4000 (allowing very precise parameter estimates to be obtained). The FID was then scaled so that the amplitude of the signal was 10. Finally, noise was added to make an extremely noisy FID. The first 50 data values from the real channel of the FID data are shown in Fig. 1A. The peak *time domain* signal-to-RMS-noise ratio of these data is approximately 2. The absorption spectrum, using a matched filter, is shown in Fig. 1B. The RMS peak signal-to-noise ratio in the frequency domain is approximately 3. The true decay rate constant as well as the phase for the absorption spectrum were determined before adding noise to the FID. The dotted line represents the zero point on the plot.

One of the problems with using the absorption spectrum is that the Y-axis does not have any specific meaning in the sense that it does not directly answer any questions about the frequencies or the amplitudes. Experience using the absorption spectrum teaches people how to interpret it in a qualitative way. In general terms, peaks well above the noise are indicative of resonances. The height of the peak times the linewidth is roughly equal to the intensity (area) of the resonance. In Fig. 1B there are a number of small peaks, and one larger one near -1000, but are any of them significant? There is simply no way to tell. The posterior probability for the frequency of a single exponentially decaying sinusoidal model is plotted in Fig. 1D. The frequency is estimated to be -959 ± 200 Hz at one standard deviation. All of the noise oscillations around the main peak were suppressed; they were more than 3 orders of magnitude down in probability, and were completely negligible for frequency estimation. Note that the frequency estimate has an uncertainty that is about one fourth the linewidth even in this extremely noisy data.

To quantify the intensity of a peak in the frequency domain, one integrates the absorption spectrum over the region of interest. The integral must be done over a number of linewidths to ensure that most of the intensity in the peak is included in the integral. In Fig. 1B the resonance linewidth is 830 Hz and a matched filter was used. This broadens the peak to over 1660 Hz. To obtain an amplitude estimate that includes 90% of the total intensity, one must integrate over approximately 6 linewidths (9960 Hz) on either side of the peak, or approximately the total spectral window. Integrating over ± 6 linewidths, one obtains 4.8 as the estimate of the amplitude, while the amplitude is known to be 10.

To quantify the intensity using probability theory, one computes the posterior probability density for the amplitude, independent of the frequency, decay rate constant, phase, and variance of the noise. In this paper, two different versions of this probability density function have been derived: one usable in high signal-to-noise, and one in low signal-to-noise. The low signal-to-noise calculation is used here. This amplitude probability density function is shown in Fig. 1C. Note that this probability density function is smooth and has none of the jagged appearance of the absorption spectrum, that is because probability theory has estimated the amplitude from a small region around the peak at -1000 Hz. Even for low signal-to-noise data, the probability density for the frequency peaks sharply and smoothly (integration is a smoothing operator). When the frequency and decay rate constant are removed as nuisance parameters, this smooths the resulting probability density function for the amplitude, even in low signal-to-noise. From Fig. 1C the amplitude is estimated to be 11 ± 1.9 at one standard deviation, the true amplitude is 10.

Figure 1: Frequency and Amplitude Estimation – Low Signal-To-Noise

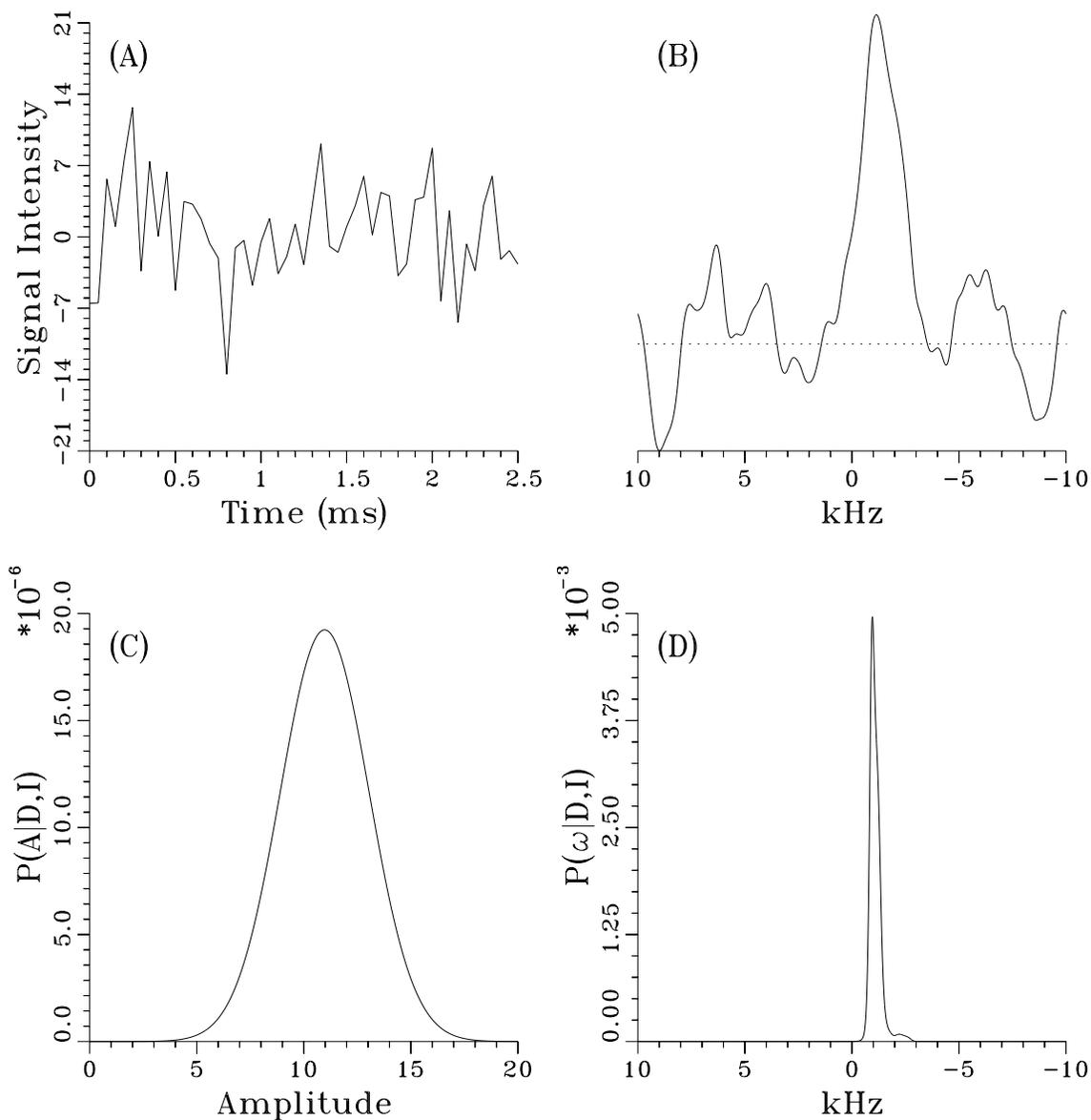


Fig. 1. The real channel of a ^1H NMR FID of HOD (A). There are $N = 50$ samples shown. The data span approximately 2.5 ms. Gaussian white noise was added to the data to make an extremely noisy data set. The peak signal-to-RMS-noise ratio in these time domain data is approximately 2.0. The absorption spectrum (B) has a number of small peaks. The largest is near -1000, but are any of them significant? The posterior probability density for the frequency (C) estimates the value of the frequency to be -959 ± 200 Hz at one standard deviation. The probability density function (D) has a well defined peak, and estimates the amplitude to be 11 ± 1.9 at one standard deviation.

Amplitude Estimation – Multiple Sinusoids

In the second example, the posterior probability for a single exponentially decaying sinusoid is applied to data that contain multiple sinusoids. This is in direct conflict to the assumption made in the model; why would one expect this calculation to work? To understand why, one need only look at Eq. [54]. The data appears in the posterior probability only in the $R(\omega, \alpha)$ and $I(\omega, \alpha)$ functions – the real and imaginary parts of the discrete Fourier transform. The radical in the posterior probability is an absolute value spectrum when a matched filter is used. The discrete Fourier transform is a linear operator so the transform of a sum of multiple exponentially decaying sinusoids is just the sum of the transforms of the individual sinusoids. The discrete Fourier transform of an exponentially decaying sinusoid is nearly a Lorentzian and the peak heights of the Lorentzians are nearly unchanged provided the peaks are well-separated. Because these heights are unchanged, the posterior probability for the amplitude will work just as if the other peaks were not present. But *one must match the filter to the peak of interest*.

To demonstrate how to use the posterior probability for the amplitude, on data with multiple sinusoids a computer simulated FID will be used. This computer-simulated FID has three resonances. The real FID data used was generated from

$$d_R(t) = 5 \cos(\omega_1 t) e^{-\alpha_1 t} + 10 \cos(\omega_2 t) e^{-\alpha_2 t} + 50 \cos(\omega_3 t) e^{-\alpha_3 t} + \epsilon_i \quad (79)$$

where the noise has standard deviation one. The imaginary data were generated from

$$d_I(t) = -5 \sin(\omega_1 t) e^{-\alpha_1 t} - 10 \sin(\omega_2 t) e^{-\alpha_2 t} - 50 \sin(\omega_3 t) e^{-\alpha_3 t} + \epsilon_i. \quad (80)$$

The sampling time is taken as 1ms and 1024 complex data points are collected. The frequencies and decay rate constants used in generating the data, are given by

$$\begin{aligned} \omega_1 &= 79.577 \text{ Hz} & \alpha_1 &= 1.59 \text{ Sec.}^{-1} \\ \omega_2 &= 63.667 \text{ Hz} & \alpha_2 &= 3.18 \text{ Sec.}^{-1} \\ \omega_3 &= -159.1 \text{ Hz} & \alpha_3 &= 31.8 \text{ Sec.}^{-1} \end{aligned} \quad (81)$$

The absorptions spectrum of this FID is shown in Fig. 2A. All of the lines are well-separated in the sense that none of the lines are riding on the shoulder of the other.

To apply probability theory to this example, one computes the posterior probability for the amplitude using the high signal-to-noise version of the posterior probability, Eq. [54]. The posterior probability for the amplitude evaluated for each of the three resonances is shown in Fig. 2 panel B through D. Panel B is for the resonance at -159 Hz, C resonance at 63.6 Hz and, D the resonance at 79.5 Hz.

Using traditional frequency domain procedures, one simply integrates over the three peaks, but before the discrete Fourier transform can be done, one must halve the first data value. Otherwise, absolute quantification of the amplitudes (the frequency domain areas) cannot be done. This is mentioned only because the $R(\omega, \alpha)$ and $I(\omega, \alpha)$ functions that appear in the posterior probability for the amplitude can be computed from a fast discrete Fourier transform. However, one must never halve the first data value when using probability theory. Additionally, for the traditional frequency domain analysis one must specify an arbitrary region over which the integrals must be performed. In this example, the limits integration will be taken to be over ± 6 linewidths, i.e., the integrals will cover approximately 90 percent of the total area. The results from both the traditional analysis and the Bayesian analysis are

Frequency	FT	Bayes
Hz	Amplitude Estimate	Amplitude Estimate
-159.1	45.1	51.5 ± 1.5
63.667	4.97	6.1 ± 0.7
79.577	9.41	10.5 ± 0.4

Figure 2: Amplitude Estimation – Multiple Sinusoids

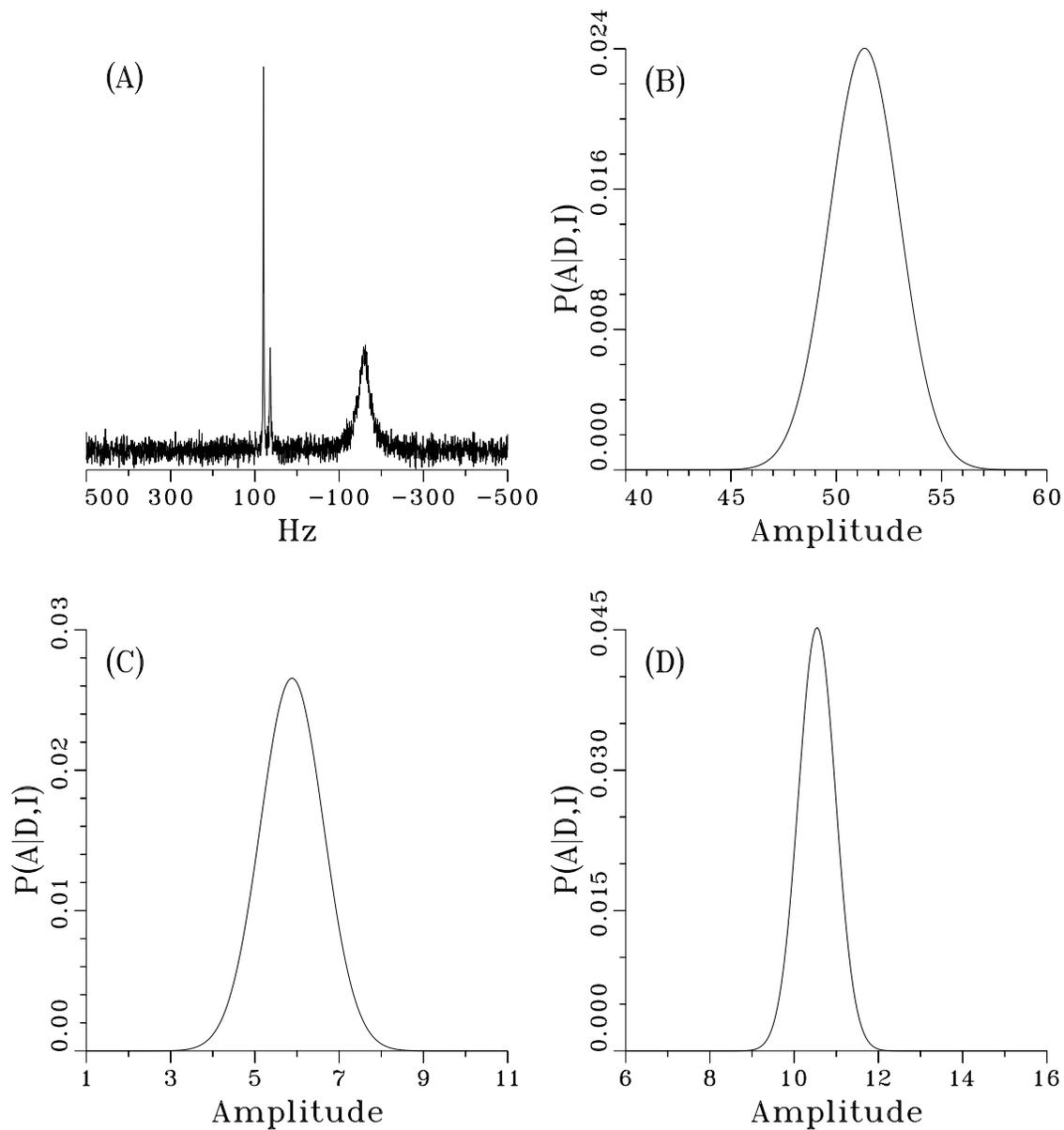


Fig. 2. If there are multiple sinusoids in the FID, and if the sinusoids are well-separated, then the posterior probability for the amplitude can be applied to each line. This procedure will work under the same conditions the discrete Fourier transform will work, and it fails for the same reasons. In panel (A) a simulated absorption spectrum is shown. This spectrum has three well-separated lines. Panel (B) contains the posterior probability for the amplitude of the rapidly decaying component on the right. Panel (C) contains the posterior probability for the small center line and panel (D) contains the posterior probability for the line on the left. The three amplitudes are estimated to be 51.5 ± 1.5 , 6.1 ± 0.7 , and 10.5 ± 0.4 respectively.

Both techniques have done reasonably well at determining the amplitudes in this high signal-to-noise example. A major difference is that probability theory tells one how accurately the amplitude has been determined. In a recent study these accuracy estimates have been analyzed to determine if they are reliable estimates of the uncertainty in a parameter estimate [14]. In that study the number of times the deviation (the difference between the “true” value of the parameter minus the estimated value) was less than the one, two, three times the width of the posterior probability was counted and it was found that this deviation was consistent with a Gaussian error distribution, indicating that the width of the posterior probability is a good indication of the uncertainty in the estimate.

Summary And Conclusions¹

In probability theory, the quantities appearing inside a probability symbol represent hypotheses. In parameter estimation problems, these hypotheses are indexed by a continuous parameter. As this parameter is varied, an infinite number of hypothesis are being tested. In the amplitude estimation calculation given in this paper, the hypothesis is of the form “the amplitude has value A given the data and the prior information.” When the index A is varied one is testing to see which values of the amplitude are supported by the data. In the regions of high probability, one should have strong belief in the hypotheses. Thus the spread in a probability density function represents how strongly one should believe a particular hypotheses, or parameter value. In this manner, probability theory carries with it both an estimate of the parameter and how uncertain one is of the parameter value.

The precision of the amplitude estimates depends not only on the time domain signal-to-noise ratio, but also inversely on the square root of the number of data values gathered. This square root of N effect can be used to improve amplitude and frequency estimates, just as signal averaging can. However, the calculation done in this paper implicitly assumes the noise is uncorrelated. This assumption is not true for oversampled data when the audio filters are turned on and the spectral window is wider than the filter bandwidth. To verify the effect, make sure that the audio filters have been turned off.

The situation is more complex than this simple explanation implies. When turned on, audio filters introduce correlations into the noise. When these correlations are incorporated into the probability theory calculation, one finds the effect is present, but it has been tempered. The increase in frequency domain signal-to-noise level will be observed for increasing sweepwidths, until all of the previously aliased noise is in the spectral window. Further increasing the sampling rate does not improve the signal-to-noise ratio. This work, to be published later, implies that spectrometers with audio filters that cut off sharply just outside the spectral window will not realize the effect; while spectrometers with audio filters that cut off gradually will realize some gain from oversampling.

But even here, care must be taken. Typically, when the sampling rate is increased, data acquisition begins earlier. This may not be desirable due to insufficient dead time. If data acquisition begins too soon, the resulting spectrum will be corrupted by baseline distortions. The solution is to begin sampling at the time appropriate for the spectrometer dead time and filter settings and then increase the sampling rate and the total complex points gathered. The resulting spectrum will have an increased frequency domain signal-to-noise ratio proportional to the square root of the N , if the noise is uncorrelated, if the noise is correlated the effect will be smaller. So if the sampling rate can be doubled, one obtains a square root of 2 increase in frequency domain signal-to-noise ratio. The frequency domain signal-to-noise ratio will continue to improve with increasing sampling rate, until all of the noise is inside the sampling window. Beyond this point, no increase in signal-to-noise ratio will be observed.

The calculation given in this paper assumes that only a single frequency is present. However, when the frequencies are well separated, this calculation can be applied to the individual peaks just as if the other frequencies were not present. When the frequencies are not well separated

¹ Copies of all computer code used in the examples are available from the author on $5\frac{1}{4}$ inch diskettes

and significant overlap occurs, the peak values of the absolute value spectrum will change and the calculation presented here no longer gives accurate answers. To estimate the amplitudes when significant overlap occurs, more complex models must be used. An initial, exploration of this problem is given in Refs. [15], and a more thorough investigation of this problem is underway.

Although not shown here in, a companion paper [14] demonstrates that probability theory can be used to estimate amplitudes and frequencies to a precision that is about three times better than using the discrete Fourier transform absorption spectrum. Additionally, probability theory can estimate both the frequencies and amplitudes in FID data whose signal-to-noise that is about three times lower than is possible using traditional methods. Thus the use of probability theory offers both greater sensitivity and accuracy in estimating the frequencies and amplitudes.

Acknowledgments

The encouragement of Professor J. J. H. Ackerman is greatly appreciated as are extensive conversations with Professor E. T. Jaynes. This work was partially supported by a gift from the Monsanto Company and NIH grant GM30331, J. J. H. Ackerman principal investigator.

References

- [1] Bretthorst, G. Larry, "Bayesian Analysis. I. Parameter Estimation Using Quadrature NMR Models," *J. Magn. Reson.* **88**, pp. 533-551 (1990).
- [2] Bretthorst, G. Larry, "Bayesian Analysis. II. Model Selection," *J. Magn. Reson.* **88**, pp. 552-570 (1990).
- [3] Bretthorst, G. Larry, "Bayesian Analysis. III. Applications to NMR Signal Detection, Model Selection and Parameter Estimation," *J. Magn. Reson.* **88**, pp. 571-595 (1990).
- [4] Bretthorst, G. Larry, "Bayesian Analysis. IV. Noise and Computing Time Considerations," *J. Magn. Reson.* **93**, pp. 369-394 (1991).
- [5] Jaynes, E. T., *Bayesian Spectrum and Chirp Analysis*, in "Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems," (C. R. Smith and G. J. Erickson *Eds.*), pp. 1-37, Reidel, Dordrecht, The Netherlands, 1987.
- [6] Bayes, Rev. T., *Philos. Trans. R. Soc. London* **53**, 370 (1763); reprinted in *Biometrika* **45**, 293 (1958), and "Facsimiles of Two Papers by Bayes," with commentary by W. Edwards Deming Hafner, New York, 1963.
- [7] Bretthorst, G. Larry, *Bayesian Spectrum Analysis and Parameter Estimation*, in "Lecture Notes in Statistics" **48**, Springer-Verlag, New York, 1988.
- [8] Jeffreys, H., "Theory of Probability," Oxford Univ. Press, London, 1939; Later editions, 1948, 1961.
- [9] Jaynes, E. T., "Papers on Probability, Statistics and Statistical Physics" (R. D. Rosenkrantz, *Ed.*), D. Reidel, Dordrecht, The Netherlands, 1983.
- [10] Bretthorst, G. Larry, *An Introduction To Parameter Estimation Using Bayesian Probability Theory*, in "Maximum-Entropy and Bayesian Methods," (P. F. Fougère *Ed.*), Kluwer, Dordrecht, The Netherlands, 1990.
- [11] Lee, Peter M., "Bayesian Statistics: An Introduction," Oxford Univ. Press, New York, 1989.

- [12] Ernst, R. Richard, "Sensitivity Enhancement in Magnetic Resonance," in *Advances in Magnetic Resonance*, Ed. John S. Waugh, bf 2, pp. 1-135, Academic Press, New York, 1966.
- [13] Lindon, J. C. and A. G. Ferrige, "Digitization And Data Processing In Fourier Transform NMR," in *Progress in NMR Spectroscopy*, **14**, pp. 27-66, Pergamon Press Ltd, Great Britten, 1980.
- [14] Kotyk, John J., N. G. Hoffman, W. C. Hutton, G. L. Bretthorst, and J. J. H. Ackerman, *J. Magn. Reson.*, in press.
- [15] Bretthorst G. Larry, *Estimating The Ratio Of Two Amplitudes In Nuclear Magnetic Resonance Data*, in "Maximum-Entropy and Bayesian Methods," (C. R. Smith and G. J. Erickson *Eds.*), Reidel, Dordrecht, The Netherlands, 1991, in press.