

"NONUNIFORM SAMPLING: BANDWIDTH AND ALIASING"

G. LARRY BRETTHORST
Washington University
Dept. of Chemistry
St. Louis, MO 63130

Abstract. For spectroscopic measurements there are good reasons why one might consider using nonuniformly nonsimultaneously sampled complex data. The primary one is that the effective bandwidth, the largest spectral window free of aliases, can be much wider than with uniformly sampled data. In this paper we discuss nonuniformly nonsimultaneously sampled data, describe how these data are traditionally analyzed, analyze them using probability theory and show how probability theory generalizes the discrete Fourier transform: first for uniformly sampled data, then for nonuniformly sampled data and finally for nonuniformly nonsimultaneously sampled data. These generalizations demonstrate that aliases are not so much removed by nonuniform nonsimultaneous sampling as they are moved to much higher frequencies.

1. Introduction

The problem of estimating the frequency of a sinusoid occurs in many different areas of science and engineering. Such data may be sampled in time, space, angle, or a host of other ways. In the discussions which follow, we will speak of the data as if they were sampled in time with the understanding that everything we say is equally applicable to data sampled in space, etc. Usually the data associated with frequency estimation are uniformly sampled; that is to say, the time interval between data samples, the dwell time, is a constant. From the standpoint of estimating a frequency, uniform samples are not more informative than nonuniform samples; they are more convenient because of the almost universal use of the fast discrete Fourier transform¹ in the frequency estimation procedure. However, in many cases, it is simply not possible to gather uniformly sampled data. For example, astronomic observations might be interrupted by clouds or the sun coming up. Additionally, as we shall see, there are good reasons why one might want to gather nonuniformly sampled data. If one has nonuniform samples, for

¹In this paper when we say "discrete Fourier transform" we mean the discrete Fourier transform of uniformly sampled data.

whatever reason, they cannot be analyzed using the discrete Fourier transform. Other analysis techniques must be used.

One such technique is the use of the Lomb-Scargle periodogram [1–3]. The Lomb-Scargle periodogram was derived using a stationary sinusoidal model with a frequency dependent phase shift. Nonlinear least-squares was used by Lomb to constrain the sine and cosine amplitudes to the values that minimized Chi-Squared. The resulting statistic turns out to be the sufficient statistic that one would derive using Bayesian probability theory for estimating the frequency given the Lomb model, see Appendix A for the details of this derivation. A sufficient statistic is a function of the data that summarizes all of the information in the data about the hypothesis of interest. Lomb’s derivation was based on an intuitive device rather than probability theory, consequently one could never be sure that the statistic uses all of the information in the data, nor can intuition alone tell one what function of the sufficient statistic is appropriate for frequency estimation.

In the case of the Lomb-Scargle periodogram, the periodogram is typically normalized by one half the expected mean-square data value. This amounts to estimating the variance of the noise and then using the log-likelihood as the statistic for frequency estimation. However, its clear from the derivations given in Appendix A, that this normalization will result in parameter estimates that are much too conservative; because it fails to account for how well the sinusoid fits the data, and so in high signal-to-noise will misestimate the uncertainty in the parameters estimates. Additionally, when calculating the Lomb-Scargle periodogram the data are often modified prior to computing the periodogram, for example as in Numerical Recipes [14]. The modification consists of computing the average data value and then subtracting that average from the data prior to computing the periodogram. This device is commonly used to account for a constant offset in the data. However, it guarantees that the data can show no evidence of low frequencies. If one does the probability theory calculations using a constant plus a stationary sinusoid as the model, probability theory will never lead one to subtract the average from the data. Rather, other statistics will appear that account for the presence of the constant in a consistent manner.

The Bayesian calculations presented in this paper will be for quadrature NMR data that has been sampled at differing times and with differing numbers of data values in each channel. We will derive the solution to the frequency estimation problem given an exponentially decaying sinusoidal model. Then, through a series of simplifications, we will reduce this calculation to the case of frequency estimation for a stationary sinusoid given real (nonquadrature) data. In the process of making these simplifications, we will encounter the Lomb-Scargle periodogram, the Schuster periodogram [4] and a weighted power spectrum as the sufficient statistic for frequency estimation. Because each will have been derived from the rules of probability theory we will see the exact conditions under which each is an optimal frequency estimator. Thus, by making these simplifications, we will see how probability theory generalizes the discrete Fourier transform to handle nonuniformly nonsimultaneously sampled data and what is happening to the aliasing phenomenon in these data.

2. The Discrete Fourier Transform

When the data consist of uniformly sampled time domain data containing some type of harmonic oscillations, the discrete Fourier transform is almost universally used as the frequency estimation technique. This is done for a number of reasons, but primarily because the technique is fast and experience has shown that the frequency estimates obtained from it are often very good. The discrete Fourier transform, $\mathcal{F}(f_k)$, is defined as:

$$\mathcal{F}(f_k) = \sum_{j=0}^{N-1} \mathbf{d}(t_j) \exp\{2\pi f_k t_j \mathbf{i}\} \quad (1)$$

where $\mathbf{i} = \sqrt{-1}$, $\mathbf{d}(t_j)$ is the complex discretely sampled data,

$$\mathbf{d}(t_j) = d_R(t_j) + \mathbf{i}d_I(t_j), \quad (2)$$

and is composed of real data samples, $d_R(t_j)$, and imaginary data samples, $d_I(t_j)$, N is the total number of complex data samples and f_k is the frequency. For uniformly sampled data, the times are given by

$$t_j = j\Delta T, \quad j \in \{0, 1 \dots N - 1\}, \quad (3)$$

where ΔT is the dwell time, the time interval between data samples, and the frequencies are given by

$$f_k = \frac{k}{N\Delta T} \quad k \in \left\{ -\frac{N}{2}, -\frac{N}{2} + 1, \dots, \frac{N}{2} \right\}. \quad (4)$$

These frequencies are the ones at which a discrete Fourier transform is exactly equal to the continuous Fourier transform of a bandlimited function [5,6]. The largest frequency interval free of aliases for a bandlimited function is given by

$$-f_{Nc} \leq f \leq f_{Nc} \quad (5)$$

and is called the bandwidth. The frequency f_{Nc} is called the Nyquist critical frequency and is given by

$$f_{Nc} = \frac{1}{2\Delta T}. \quad (6)$$

Nothing would prohibit one from taking f_k as a continuous variable and evaluating Eq. (1) at different frequencies. Indeed, this is exactly what the common practice of zero-padding² does. After all, adding zero to a sum does not change that sum, so for any given frequency zero padding has no effect in Eq. (1). The only effect is in Eq. (4): changing N changes the frequencies at which Eq. (1) is evaluated.

²To zero pad a data set, one adds zeros to the end of a data set, sets N to the length of this new zero padded data set and then runs a fast discrete Fourier transform on the zero padded data.

If we expand the right-hand side of Eq. (1), we obtain

$$\mathcal{F}(f_k) = R(f_k) + \mathbf{i}I(f_k) \quad (7)$$

where

$$R(f_k) = \sum_{j=0}^{N-1} [d_R(t_j) \cos(2\pi f_k t_j) - d_I \sin(2\pi f_k t_j)] \quad (8)$$

and

$$I(f_k) = \sum_{j=0}^{N-1} [d_R(t_j) \sin(2\pi f_k t_j) + d_I \cos(2\pi f_k t_j)] \quad (9)$$

are the real and imaginary parts of the discrete Fourier transform.

Three different ways of viewing the results of the discrete Fourier transform are common: the absorption spectrum, the power spectrum and the absolute-value spectrum. The absorption spectrum, the real part of an appropriately phased discrete Fourier transform, is commonly used in NMR. In NMR the sinusoids usually have the same phase; consequently if one multiplies Eq. (7) by $\exp\{\mathbf{i}(\theta + T_0 f_k)\}$, the phase of the sinusoids can be made to cancel from the discrete Fourier transform. The two parameters, θ and T_0 , are the zero- and first-order phase corrections, and must be estimated from the discrete Fourier transform. An absorption spectrum's usefulness is limited to problems in which the sinusoids have the same phase parameters. This is common in NMR, but not with other physical phenomena, consequently, we will not discuss the absorption spectrum further.

The power spectrum is defined as

$$\text{Power}(f_k) = \frac{R(f_k)^2 + I(f_k)^2}{N} \quad (10)$$

and is the square of the absolute-value spectrum. It has been shown by Bretthorst [7–9] and Woodward [10], and as we will demonstrate shortly, the power spectrum is the sufficient statistic in a Bayesian calculation for the posterior probability for the frequency given a single stationary sinusoidal model.

In this paper we will make several plots of the discrete Fourier transform and its generalizations to nonuniformly nonsimultaneously sampled data. To allow direct comparison of these plots we will always plot the same function of the data, the base 10 logarithm of the posterior probability for the frequency of a stationary sinusoid independent of the phase, amplitude and variance of the noise, Eq. (36) below. For uniformly sampled quadrature data, this probability is a simple function of the power spectrum, see Bretthorst [11] for a more extended discussion of the relationship between the discrete Fourier transform and the posterior probability for a stationary frequency.

To illustrate the use of the discrete Fourier transform as a frequency estimation tool, suppose we had the data shown in Fig. 1(A). The signal in this simulated data is an exponentially decaying sinusoid of amplitude 10 plus Gaussian noise of zero mean and standard deviation one. These data were generated with a dwell time of $\Delta T = 0.01$ Sec. One hundred complex data values were generated at times

Figure 1. A Uniformly Sampled Exponentially Decaying Sinusoid

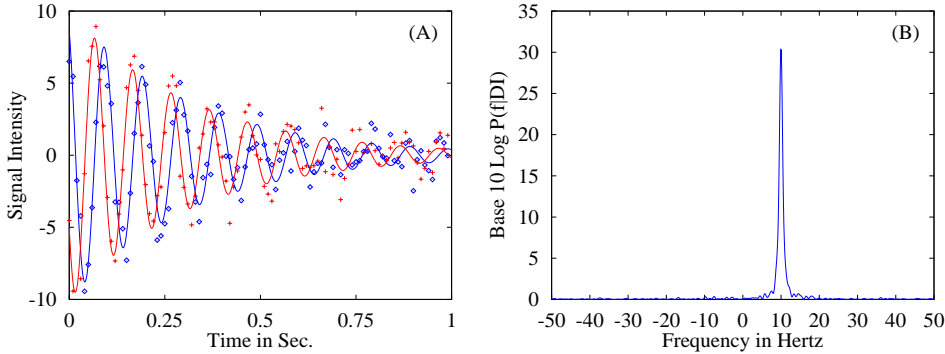


Fig. 1. Panel (A) is computer simulated data. It contains a single exponentially decaying sinusoidal signal plus noise. The lines represent the real and imaginary parts of the sinusoid. The locations of the data values are denoted by the isolated characters. The Nyquist critical frequency for this data set is $f_{Nc} = 50$ Hz. Panel (B) is a plot of the base 10 logarithm of the posterior probability for the frequency of a stationary sinusoid given these data.

ranging from 0 to 0.99 seconds. The frequency is 10 Hz, and the decay rate constant is 3 Sec.^{-1} . The real and imaginary data values are represented by the isolated characters in Fig. 1(A). The lines represent the real and imaginary parts of the true sinusoid. The Nyquist critical frequency for these data is one-half the inverse of the dwell time:

$$f_{Nc} = \frac{1}{2\Delta T} = \frac{1}{2(0.01 \text{ Sec.})} = 50 \text{ Hz.} \quad (11)$$

Figure 1(B) is a plot of the base 10 logarithm of the posterior probability over the bandwidth ($-50 \text{ Hz} \leq f \leq 50 \text{ Hz}$). This base 10 logarithm starts at essentially zero and then increases some 30 orders of magnitude, coming to a very sharp peak at 10 Hz.

We would like to investigate the aliasing phenomenon. To do this we must evaluate the discrete Fourier transform outside the bandwidth and this cannot be done using the fast discrete Fourier transform. Because we are plotting the base 10 logarithm of the posterior probability for the frequency of a stationary sinusoid, we can simply evaluate the posterior probability for any frequency and the sufficient statistic will be the power spectrum evaluated at that frequency; we are not restricted to the frequencies f_k specified by Eq. (4). The resulting plot is shown in Fig. 2. Outside of the interval ($-f_{Nc} \leq f \leq f_{Nc}$), the base 10 logarithm of the posterior probability, and thus the power spectrum and the discrete Fourier transform, are periodic functions of f with a period equal to the frequency interval spanned by the bandwidth. In Fig. 2, the frequency interval plotted is ($-10f_{Nc} \leq f \leq 10f_{Nc}$), so there should be 10 peaks in this range as Fig. 2 shows.

Figure 2. The $\text{Log}_{10}P(f|DI)$ At Frequencies Greater Than f_{Nc}

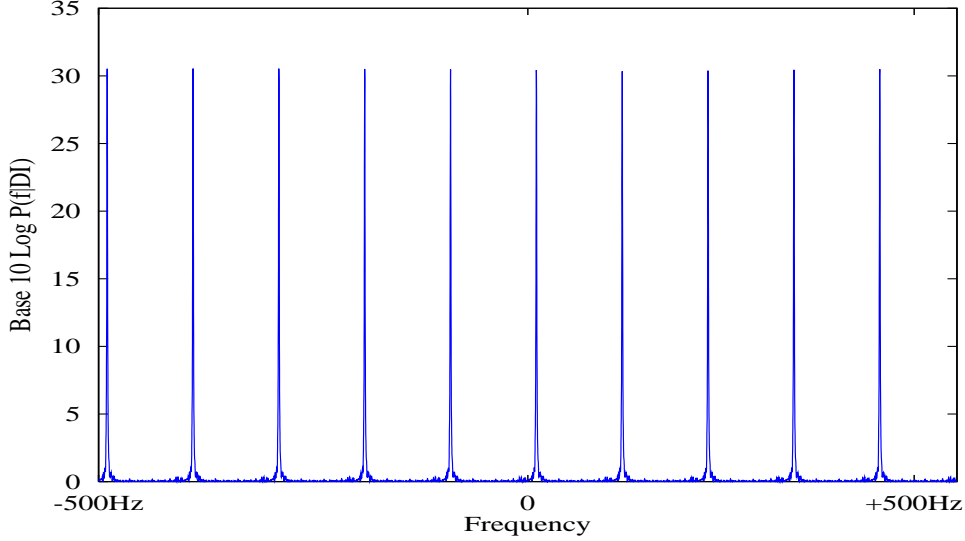


Fig. 2. The reason that the discrete Fourier transform of uniformly sampled data is rarely computed at frequencies greater than the Nyquist critical frequency is simply that the discrete Fourier transform is a periodic function with a period equal to the bandwidth $2f_{Nc}$.

To understand why the discrete Fourier transform is a periodic function of frequency, suppose we wish to evaluate the discrete Fourier transform at the frequencies

$$f_k = \frac{k}{N\Delta T}, \quad k = mN + k', \quad k' \in \left\{ -\frac{N}{2}, -\frac{N}{2} + 1, \dots, \frac{N}{2} \right\}. \quad (12)$$

By itself the index k' would specify the normal frequency interval, Eq. (4), of a discrete Fourier transform. However, the integer m shifts this frequency interval up or down by an integer multiple of the total bandwidth. If $m = 0$, we are in the interval $(-f_{Nc} \leq f_k \leq f_{Nc})$; if $m = 1$, we are in the interval $(f_{Nc} \leq f_k \leq 3f_{Nc})$, etc. If we now substitute Eqs. (12) and (3) into the discrete Fourier transform, Eq. (1), the reason the discrete Fourier transform is periodic becomes readily apparent

$$\mathcal{F}(f_{k'}) \equiv \sum_{j=0}^{N-1} \mathbf{d}(t_j) \exp \left\{ \frac{2\pi i(mN + k')j}{N} \right\}, \quad (13)$$

$$= \sum_{j=0}^{N-1} \mathbf{d}(t_j) \exp \{2\pi i m j\} \exp \left\{ \frac{2\pi i k' j}{N} \right\}, \quad (14)$$

$$= \sum_{j=0}^{N-1} \mathbf{d}(t_j) \exp \left\{ \frac{2\pi \mathbf{i} k' j}{N} \right\}, \quad (15)$$

$$= \sum_{j=0}^{N-1} \mathbf{d}(t_j) \exp \{2\pi \mathbf{i} f_k t_j\}. \quad (16)$$

In going from Eq. (14) to (15) a factor, $\exp\{\mathbf{i}(2\pi m j)\}$, was dropped because both m and j are integers, so $(2\pi m j)$ is an integer multiple of 2π , and the complex exponential is one. Aliases occur because the complex exponential canceled leaving behind a discrete Fourier transform on the interval $(-f_{Nc} \leq f_k \leq f_{Nc})$. The integer m specifies which integer multiple of the bandwidth is being evaluated and will always be an integer no matter how the data are collected. However, the integer j came about because the data were uniformly sampled. If the data had not been uniformly sampled the relationship, $t_j = j\Delta T$, would not hold, the complex exponential would not have cancelled, and aliases would not have been present.

Because frequency estimation using the discrete Fourier transform was not derived using the rules of probability theory, there is no way to be certain that the estimates we obtain are the best we could do. What is needed is the solution to the single frequency estimation problem using Bayesian probability theory. Consequently, in the next section we analyze this problem, then in the following sections we will derive the conditions under which the discrete Fourier transform power spectrum is a sufficient statistic for single frequency estimation and we will see how probability theory generalizes the discrete Fourier transform to nonuniformly nonsimultaneously sampled data, and we will see the effect of these generalizations on the aliasing phenomenon.

3. Single-Frequency Estimation

The problem to be addressed is the estimation of the frequency of a single exponentially decaying sinusoid independent of the amplitude and phase of the sinusoid, given nonuniformly nonsimultaneously sampled quadrature data. First, what do we mean by nonuniformly nonsimultaneously sampled quadrature data? “Quadrature” simply means we have a measurement of the real and imaginary parts of a complex signal. So nonuniformly nonsimultaneously sampled quadrature data are measurements of the real and imaginary parts of a complex signal for which the measurements of the real and imaginary parts of the signal occur at different times. But if we have differing numbers of data samples with differing sample times, we really have two data sets: a real and an imaginary data set.³ The real data set will be designated as $D_R \equiv \{d_R(t_1) \cdots d_R(t_{N_R})\}$, where d_R means a real data sample, t_i is the time the data sample was acquired, and N_R is the total number of data samples in the real data set. Similarly, the imaginary data set will be denoted by $D_I \equiv \{d_I(t_1) \cdots d_I(t_{N_I})\}$. We impose no restrictions on the number of data samples or the acquisition times in either channel. They could be the same or different,

³Of course, when we say an “imaginary data set” we mean only that the data are a measurement of the imaginary part of the signal; not that the data are imaginary numbers.

depending on the limit we investigate. Note that the times do not carry a channel indication so the context within the equations will have to establish which times we are referring to. If the equation context fails, we will clarify it in the text.

To perform any calculation using probability theory, the hypothesis of interest must be related to the information we actually possess. For the problem of estimating the frequency of a complex sinusoid, this means relating the frequency to the quadrature data through a model. If the complex data are given by Eq. (2), then the data and the sinusoid are related by

$$\mathbf{d}(t_j) = \mathbf{A} \exp \{-\mathbf{f}t_j\} + \mathbf{n}(t_j). \quad (17)$$

The complex amplitude is given by $\mathbf{A} = A_1 - \mathbf{i}A_2$, and is equivalent to the amplitude and phase of the sinusoid. The complex frequency, $\mathbf{f} = \alpha + 2\pi\mathbf{i}f$, contains two parameters: the decay rate constant, α , and the frequency, f . Note the minus signs in the definition of the complex amplitude and the one in Eq. (17). These signs correspond to a convention establishing what is meant by a positive frequency. The signs were chosen to model the data produced by a Varian NMR spectrometer. Other vendors use different conventions. Changing these conventions will change none of the conclusions that follow and very few of the actual details of the calculations. The decay rate constant, α , has units of inverse seconds, the frequency, f , Hertz, and the times, t_j , seconds. The quantity $\mathbf{n}(t_j)$ represents the complex noise at time t_j . Note that in this equation the times, t_j , simply designate the times at which we actually have data. If the datum happen to be a measurement of the real part of the signal, then the time would be associated with the real data set, and similarly for the imaginary part of the signal.

If we separate Eq. (17) into its real and imaginary parts, we have for the real part

$$d_R(t_i) = M_R(t_i) + n_R(t_i) \quad (18)$$

$$M_R(t_i) \equiv [A_1 \cos(2\pi f t_i) - A_2 \sin(2\pi f t_i)] \exp \{-\alpha t_i\} \quad (19)$$

and for the imaginary part we have

$$d_I(t_j) = M_I(t_j) + n_I(t_j) \quad (20)$$

$$M_I(t_j) \equiv -[A_1 \sin(2\pi f t_j) + A_2 \cos(2\pi f t_j)] \exp \{-\alpha t_j\}, \quad (21)$$

where $n_R(t_i)$ and $n_I(t_j)$ represent the noise in the real and imaginary data at times t_i and t_j . The quantity that we would like to estimate is the frequency, f , and we would like to estimate it independent of the amplitudes and variance of the noise. The decay rate constant, α , will sometimes be treated as a nuisance parameter, sometimes estimated, and sometimes taken as a given, depending of our purpose at the time. For now we will estimate it.

In probability theory as logic, all of the information about a hypothesis is summarized in a probability density function. For this problem, the probability density function is designated as $P(f\alpha|D_R D_I I)$, which should be read as the joint posterior probability for the frequency and decay rate constant given the real and

imaginary data sets and the information I . The information I is all of the other information we have about the parameters appearing in the problem. The joint posterior probability for the frequency and decay rate constant is computed from the joint posterior probability for all of the parameters, $P(f\alpha A_1 A_2 \sigma | D_R D_I I)$, by application of the sum and product rules of probability theory:

$$P(f\alpha | D_R D_I I) = \int dA_1 dA_2 d\sigma P(f\alpha A_1 A_2 \sigma | D_R D_I I). \quad (22)$$

Using the product rule, the right-hand side of this equation may be factored to obtain

$$P(f\alpha | D_R D_I I) = \int dA_1 dA_2 d\sigma P(f\alpha A_1 A_2 \sigma | I) P(D_R D_I | f\alpha A_1 A_2 \sigma I). \quad (23)$$

We will assume that the prior, $P(f\alpha A_1 A_2 \sigma | I)$, can be factored into independent priors for each parameter, and these priors will be assigned an appropriately uninformative prior probability (uniform priors for the amplitudes, and a Jeffreys' prior for the standard deviation of the noise). The uninformative prior for the frequency would normally be taken as a uniform prior. The uninformative prior for the decay rate constant would typically be taken as a Jeffreys' prior. However, in what follows we will sometimes wish to estimate the decay rate constant, sometimes eliminate it as a nuisance parameter and sometimes treat it as a given. Consequently, we will specify which prior we are using for the decay rate constant at the appropriate time.

If the two data sets are logically independent, the joint direct probability for the data, $P(D_R D_I | f\alpha A_1 A_2 \sigma I)$, will factor as

$$P(D_R D_I | f\alpha A_1 A_2 \sigma I) = P(D_R | f\alpha A_1 A_2 \sigma I) P(D_I | f\alpha A_1 A_2 \sigma I), \quad (24)$$

and the joint posterior probability for the frequency and decay rate constant becomes

$$P(f\alpha | D_R D_I I) = \int dA_1 dA_2 \frac{d\sigma}{\sigma} P(D_R | f\alpha A_1 A_2 \sigma I) P(D_I | f\alpha A_1 A_2 \sigma I), \quad (25)$$

where, for now, we are using a uniform prior for the decay rate constant. The two direct probabilities for the data are to be assigned given that one knows all of the parameters appearing in the model. But if one knows all of these parameters, then from Eqs. (18)-(20) one can simply compute the noise in the real and imaginary data sets. If one can assign a probability for the noise one can then assign these two direct probabilities. For the reasons explained in Jaynes [12], and further elaborated by Bretthorst [13], the probability for the noise will be assigned as a Gaussian and the joint posterior probability for the frequency and decay rate

constant is given by

$$\begin{aligned}
P(f\alpha|D_R D_I I) &\propto \int dA_1 dA_2 \frac{d\sigma}{\sigma} \\
&\times \sigma^{-N_R} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=0}^{N_R-1} [d_R(t_i) - M_R(t_i)]^2\right\} \\
&\times \sigma^{-N_I} \exp\left\{-\frac{1}{2\sigma^2} \sum_{j=0}^{N_I-1} [d_I(t_j) - M_I(t_j)]^2\right\}.
\end{aligned} \tag{26}$$

The proportionality sign comes about because a number of constants have been ignored. Substituting Eqs. (19) and (21) for $M_R(t_i)$ and $M_I(t_j)$ into Eq. (26), the joint posterior probability for the frequency and decay rate constant is given by

$$P(f\alpha|D_R D_I I) \propto \int dA_1 dA_2 \frac{d\sigma}{\sigma} \sigma^{-(N_R+N_I)} \exp\left\{-\frac{Q}{2\sigma^2}\right\} \tag{27}$$

where

$$Q \equiv (N_R + N_I)\bar{d}^2 - 2\sum_{l=1}^2 A_l T_l + \sum_{k,l=1}^2 g_{kl} A_k A_l. \tag{28}$$

The mean-squared data value is defined as

$$\bar{d}^2 = \frac{1}{N_R + N_I} \left[\sum_{i=0}^{N_R-1} d_R(t_i)^2 + \sum_{j=0}^{N_I-1} d_I(t_j)^2 \right]. \tag{29}$$

The projection of the data onto the model, the vector T , is given by

$$\begin{aligned}
T_1 &\equiv \sum_{i=0}^{N_R-1} d_R(t_i) \cos(2\pi f t_i) \exp\{-\alpha t_i\} \\
&- \sum_{j=0}^{N_I-1} d_I(t_j) \sin(2\pi f t_j) \exp\{-\alpha t_j\}
\end{aligned} \tag{30}$$

and

$$\begin{aligned}
T_2 &\equiv -\sum_{i=0}^{N_R-1} d_R(t_i) \sin(2\pi f t_i) \exp\{-\alpha t_i\} \\
&- \sum_{j=0}^{N_I-1} d_I(t_j) \cos(2\pi f t_j) \exp\{-\alpha t_j\}.
\end{aligned} \tag{31}$$

The matrix g_{kl} is defined as

$$g_{kl} \equiv \begin{pmatrix} a & c \\ c & b \end{pmatrix}, \tag{32}$$

with

$$a = \sum_{i=0}^{N_R-1} \cos^2(2\pi f t_i) \exp\{-2\alpha t_i\} + \sum_{j=0}^{N_I-1} \sin^2(2\pi f t_j) \exp\{-2\alpha t_j\} \quad (33)$$

where the sum over the cosine uses the t_i associated with the real data set, while the sum over the sine uses the t_j associated with the imaginary data set. Similarly, b is defined as

$$b = \sum_{i=0}^{N_R-1} \sin^2(2\pi f t_i) \exp\{-2\alpha t_i\} + \sum_{j=0}^{N_I-1} \cos^2(2\pi f t_j) \exp\{-2\alpha t_j\}, \quad (34)$$

where the sum over the sine uses the t_i associated with the real data set, while the sum over the cosine uses the t_j associated with the imaginary data set. Finally, c is defined as

$$\begin{aligned} c = & - \sum_{i=0}^{N_R-1} \cos(2\pi f t_i) \sin(2\pi f t_i) \exp\{-2\alpha t_i\} \\ & + \sum_{j=0}^{N_I-1} \sin(2\pi f t_j) \cos(2\pi f t_j) \exp\{-2\alpha t_j\}, \end{aligned} \quad (35)$$

where the sum over the cosine-sine product uses the t_i associated with the real data set, while the sum over the sine-cosine product uses the t_j associated with the imaginary data set.

The integrals over A_1 and A_2 are both Gaussian integrals and are easily evaluated. The integral over σ is a gamma integral and is also easily evaluated. Evaluating this triple integral, one obtains

$$P(f\alpha|D_R D_I I) \propto \frac{1}{\sqrt{ab - c^2}} \left[(N_R + N_I) \overline{d^2} - \overline{h^2} \right]^{(2-N_R-N_I)/2} \quad (36)$$

as the joint posterior probability for the frequency and decay rate constant. In Fig. 1(B) and Fig. 2 it is the base 10 logarithm of this posterior probability density function that is plotted as a function of frequency. In those plots we wished to illustrate the relationship of the sufficient statistic, $\overline{h^2}$, to the discrete Fourier transform power spectrum, and for reasons that will become apparent shortly, we set $\alpha = 0$, thereby plotting the base 10 logarithm of the posterior probability for the frequency of a stationary sinusoid—in spite of the fact that the resonance in the simulated data was exponentially decaying. The sufficient statistic, $\overline{h^2}$, is given by

$$\overline{h^2} \equiv \frac{bT_1^2 + aT_2^2 - 2cT_1T_2}{ab - c^2}. \quad (37)$$

While not obvious, it is this statistic that generalizes the discrete Fourier transform to nonuniformly nonsimultaneously sampled data. This statistic will reduce to the Lomb-Scargle periodogram, a weighted normalized power spectrum and the Schuster periodogram under appropriate conditions.

If we are interested in estimating both the frequency and decay rate constant, Eq. (36) should be used. If we are interested in only the frequency, then the decay rate constant is a nuisance parameter and should be removed by application of the sum rule. A Jeffreys' prior would be the appropriate uninformative prior probability for the decay rate constant. The posterior probability for the frequency is thus given by

$$P(f|D_R D_I I) \propto \int d\alpha \frac{P(\alpha|I)}{\sqrt{ab - c^2}} \left[(N_R + N_I) \overline{d^2} - \overline{h^2} \right]^{(2-N_R-N_I)/2}. \quad (38)$$

If the decay rate constant is a given, the prior, $P(\alpha|I)$, should be taken as a delta function and the integral in Eq. (38) evaluated analytically. Otherwise, the integral must be evaluated numerically. Finally, if we want the best estimate of the decay rate constant, a bounded uniform prior probability for the frequency would be appropriate, and one would compute

$$P(\alpha|D_R D_I I) \propto \int df \frac{1}{\alpha \sqrt{ab - c^2}} \left[(N_R + N_I) \overline{d^2} - \overline{h^2} \right]^{(2-N_R-N_I)/2}. \quad (39)$$

This integral must also be evaluated numerically, but because the integrand is a very sharply peaked function of frequency, a Gaussian approximation works well.

4. How Probability Generalizes The Discrete Fourier Transform

We are now in a position to demonstrate how probability theory generalizes the discrete Fourier transform and what the effect of this generalization is on the aliasing phenomenon. We mentioned earlier that the sufficient statistic for frequency estimation is related to a power spectrum, and we demonstrate that next. Several simplifications must be made to reduce Eq. (37) to a power spectrum. First, we must be estimating the frequency of a stationary sinusoid. A stationary sinusoid has no decay, so $\alpha = 0$. Second, the data must be uniformly sampled. But uniform sampling is not quite enough; the data must also be simultaneously sampled. With these assumptions the matrix g_{kl} , Eq. (32), simplifies because $c = 0$ and $a = b = N_R = N_I = N$, where N is the total complex data values. The sufficient statistic, Eq. (37), reduces to

$$\overline{h^2} = \frac{R(f)^2 + I(f)^2}{N} \quad (40)$$

and is the power spectrum defined in Eq. (10). The functions $R(f)$ and $I(f)$ were defined earlier, Eqs. (8) and (9), and are the real and imaginary parts of the discrete Fourier transform.

The Schuster periodogram, or power spectrum, is the sufficient statistic for frequency estimation in uniformly simultaneously sampled quadrature data given a stationary sinusoidal model, but we already have two generalizations to the discrete Fourier transform that were not contained in the definition, Eq. (1). First, the frequency appearing in Eq. (40) is a continuous parameter; it is not in any way restricted to discrete values. Probability theory indicates that there is information

in the data at frequencies between the f_k in the discrete Fourier transform. Second, the frequency, f , is unbounded. Probability theory does not indicate that the frequency must be less than the Nyquist critical frequency. Consequently, probability theory is telling one that aliases are real indications of the presence of a frequency and absolutely nothing in the data can tell one which is the correct frequency, only prior information can do that.

The Schuster periodogram is an optimal frequency estimator for uniformly simultaneously sampled quadrature data. However, this is not true for real data where the statistic was originally proposed. To see this suppose we have a real data set, so that either $N_I = 0$ or $N_R = 0$. The g_{kl} matrix does simplify—many of the sums appearing in Eqs. (33)-(35) are zero—but the matrix remains nondiagonal and so Eq. (37) is the sufficient statistic for this problem, and is numerically equal to the Lomb-Scargle periodogram. The Schuster periodogram, however, is never a sufficient statistic for either real uniformly or nonuniformly sampled data. It can only be derived as an approximation to the sufficient statistic for this problem. To derive it two approximations must be made in the g_{kl} matrix. First, the off-diagonal element must be much smaller than the diagonal and so can be approximated as zero. The second approximation assumes the diagonal elements may be approximated by $a = b = N/2$. Both approximations ignore terms on the order of \sqrt{N} and are good approximations when one has large amounts of data. In this discussion we have implicitly assumed that the times appearing in the cosine and sine transforms of the data making up the Schuster periodogram were evaluated at the times one actually has data. Trying to interpolate the data onto a uniform grid and then using a power spectrum is not justified under any circumstances. However, its hard to see how any bad results that one might obtain after doing this are to be blamed on the Schuster periodogram.

Suppose now the signal is exponentially decaying with time, but otherwise the data remain uniformly and simultaneously sampled. Examining Eq. (32), we see that when the data are simultaneously acquired, $c = 0$ and $a = b = N_{\text{eff}}$, where N_{eff} is an effective number of complex data values and is given by

$$N_{\text{eff}} = \sum_{i=0}^{N-1} \exp \{-2\alpha t_i\}. \quad (41)$$

The sufficient statistic becomes

$$\overline{h^2} = \frac{R(f, \alpha)^2 + I(f, \alpha)^2}{N_{\text{eff}}}. \quad (42)$$

The effective number of complex data values is equal to the number of complex data values, N , when the decay rate constant, α , is zero, and is approximately, $1/2\alpha$, for densely sampled signals which decay into the noise. The reason for this behavior should be obvious: as the decay rate constant increases, for a fixed dwell time, fewer and fewer data values contribute to the estimation process. When the decay rate constant is large, the effective number of data values goes to zero and the data are uninformative about either the frequency or the decay rate constant.

The functions $R(f, \alpha)$ and $I(f, \alpha)$ are the real and imaginary parts of the discrete Fourier transform of the complex data weighted by an exponential of the decay rate constant α . In a weighted discrete Fourier transform, one multiplies the complex data by a weighting function

$$\text{Complex Weighted Data} = \mathbf{d}(t_i) \exp\{-\alpha t_i\} \quad (43)$$

and then performs the discrete Fourier transform on the weighted data.

In spectroscopic applications many different weighting functions are used. Indeed Varian NMR software comes with exponential, Gaussian, sine-bell and a number of others. In all of these cases, use of the weighting function in conjunction with a discrete Fourier transform amounts to estimating the frequency of a single sinusoid having a decay envelope described by the weighting function. If the weighting function does not mimic the decay envelope of the signal, then these procedures are, from the standpoint of parameter estimation, less than optimal. Of course most of these weighting functions were developed with very different ideas in mind than parameter estimation. For example, the sine-bell is intended to increase resolution of multiple close lines, just as a Gaussian is used to transform the line shape of the resonances from Lorentzian to Gaussian in the hope that the Gaussian will be narrower in the frequency domain. Nonetheless, probability theory indicates that all of these procedures are estimating the frequency of a single sinusoid having a decay envelope described by the weighting function. The better this weighting function describes the true decay of the signal, the better the parameter estimates will be.

For exponential weighting, the spectroscopist must choose a value of α . This is typically done so that the decay envelope of the weighting function matches the decay of the signal. This is equivalent to trying to locate the maximum of the joint posterior probability for the frequency and decay rate constant, Eq.(27). If one makes a contour plot of this joint posterior probability, this plot will have nearly elliptical contours with the major axis of the ellipse nearly parallel to the decay rate constant axis (decay rate constants are less precisely estimated than frequencies), while the minor axis will be nearly parallel to the frequency axis. Thus, estimates of the frequency and decay rate constant are nearly independent of each other. Consequently, if the spectroscopist can guess the decay rate constant, even approximately, the frequency estimate he obtains will be almost as good as that obtained by doing a thorough search for the location of the joint maximum.

It is commonly believed that matched weighting functions (matching the shape of the weighting function to the observed decay of the data) increases the signal-to-noise ratio in the resulting power spectrum at the expense of broadening the peak, thereby decreasing the frequency resolution of the discrete Fourier transform. This is correct if the discrete Fourier transform is used as a spectral estimation procedure and one then tries to estimate multiple frequencies from this spectrum. However, from the standpoint of probability theory, it is only the single largest peak in the weighted discrete Fourier transform power spectrum that is relevant to frequency estimation, and then it is only a very small region around the location of the maximum that is of interest. All of the details in the wings around

that peak are irrelevant to the estimation problem. If there are multiple peaks in the power spectrum, probability theory will systematically ignore them: *the weighted power spectrum is the sufficient statistic for single frequency estimation*; it does not estimate multiple frequencies, although one can show that under many conditions the sufficient statistic for the multiple frequency estimation problem is related to the multiple peaks in a power spectrum [11]. Given a multiple-frequency model, probability theory will lead one to other statistics that take into account the nonorthogonal nature of the sinusoidal model functions. These multiple-frequency models always result in parameter estimates that are either better than or essentially identical to those obtained from a power spectrum. They will be essentially identical to the power spectrum results when multiple, very well separated sinusoids are present, and they will always be better when overlapping resonances are present—see Bretthorst [7–9,11,15–17] for details.

Suppose we have nonuniformly but simultaneously sampled data. What will happen to the resulting Bayesian calculations? When we repeat the Bayesian calculation we find that absolutely nothing has changed. The number of effective data values, N_{eff} , is given by Eq. (41), the matrix g_{kl} is given by Eq. (32), just as the sufficient statistic is given by Eq. (42), and the joint posterior probability for the frequency and decay rate constant is given by Eq. (36). Here the generalization to the discrete Fourier transform accounts for the fact that the times must be evaluated explicitly, the formula, $t_j = j\Delta T$, does not hold and one must substitute the actual time, t_i and t_j , into the equations. Missing observations, make no difference whatever to probability theory. Probability theory analyzes the data you actually obtain, regardless of whether the data are uniformly sampled or not. Indeed, in Bayesian probability theory there is no such thing as a missing data problem.

Having allowed the data to be nonuniformly sampled, we are in a position to see what is happening to the aliasing phenomenon. However, before addressing this, there is one last generalization that we would like to make. This generalization allows the data to be nonsimultaneously sampled. Because the samples are nonsimultaneous, the sums appearing in Eqs. (8) and (9) are no longer correct. The terms appearing in these equations are the sine and cosine transforms of the real and imaginary data sets. When the data were simultaneously sampled, the sine and cosine transforms could be combined into a single sum. For nonsimultaneous time samples, this cannot be done. Each sine and cosine transform must have an independent summation index. If you examine the projections of the model onto the nonuniformly nonsimultaneously sampled data, Eqs. (30) and (31), you will find this is exactly what probability theory has done. The function T_1 corresponds to the real part of the discrete Fourier transform and is the cosine transform of the real data minus the sine transform of the imaginary data; however, now it also accounts for the nonuniform nonsimultaneous times. Similarly, up to a minus sign,⁴ T_2 corresponds to the imaginary part of the discrete Fourier transform. So the discrete Fourier transform has been generalized in the sense that the sine and cosine transforms now have separate summation indices. However, there is more to this generalization than using separate summation indices.

⁴The minus sign comes about because of the use of Varian sign conventions in Eq. (17).

In the previous examples the matrix g_{kl} , Eq. (32), was diagonal with diagonal elements equal to the effective number of data values in each channel. For simultaneously sampled data, these are equal. For nonsimultaneously sampled data, the diagonal elements remain the effective number of data values in each channel, but these are no longer equal. For simultaneous data samples, the zero off-diagonal element means that the integrals over the amplitudes are completely independent of each other. In the model, the function multiplying each amplitude may be thought of as an N dimensional vector. With nonsimultaneous samples these vectors are linear combinations of each other. The magnitude of the off-diagonal element is a measure of how far from orthogonal these vectors are. Consequently, the sufficient statistic, Eq. (37), is now taking into account the fact that these vectors are not orthogonal, the real and imaginary data can have different numbers of data values and that the effective number of data values is a function of both frequency and decay rate constant.

5. Aliasing

Now that we have finished discussing the generalizations of the discrete Fourier transform to nonuniformly nonsimultaneously sampled data, we would like to investigate some of the properties of these calculations to show what has happened to the aliasing phenomenon. Earlier, when we investigated the discrete Fourier transform of uniformly sampled data, we showed that, for frequencies outside the bandwidth, the power spectrum was a periodic function of frequency. What will happen to these aliases when we use nonuniformly nonsimultaneously sampled data? Are the aliases still there? If not, where did they go? One thing that should be obvious is that in nonuniformly nonsimultaneously sampled data the Nyquist critical frequency does not apply, at least not as previously defined, because the times are nonuniformly sampled. Consequently, nonuniformly nonsimultaneously sampled data will not suffer from aliases in the same way that uniformly simultaneously sampled data does.

To demonstrate this, we will again use simulated data. The simulated data will have exactly the same signal as the data shown in Fig. 1(A). The only difference between these two data sets will be the times at which the data were acquired. The nonuniformly nonsimultaneously sampled simulated data are shown in Fig. 3(A). In generating the times at which we have data we had to choose a sampling scheme. On some spectrometers it is possible to sample data exponentially, so we choose exponential sampling. In an exponentially sampled data set, a histogram of the time samples would follow an exponential distribution. Thus, there are more data samples at short times, and exponentially fewer at longer times. However, the discussion here pertains to all nonuniformly nonsimultaneously sampled data, not just to data with times that are exponentially sampled. The base 10 logarithm of the posterior probability for the frequency of a stationary sinusoid is shown in Fig. 3(B). This plot spans a frequency interval that is ten thousand times larger than the corresponding plot shown in Fig. 1(B). In Fig. 2(B), when we extended the region of interest to $\pm 10f_{Nc} = \pm 500$ Hz we had 10 aliases. Yet here we have gone to ± 50 kHz and there are no aliases—where did the aliases go? Why do these

Figure 3. A Nonuniformly Nonsimultaneously Sampled Exponentially Decaying Sinusoid

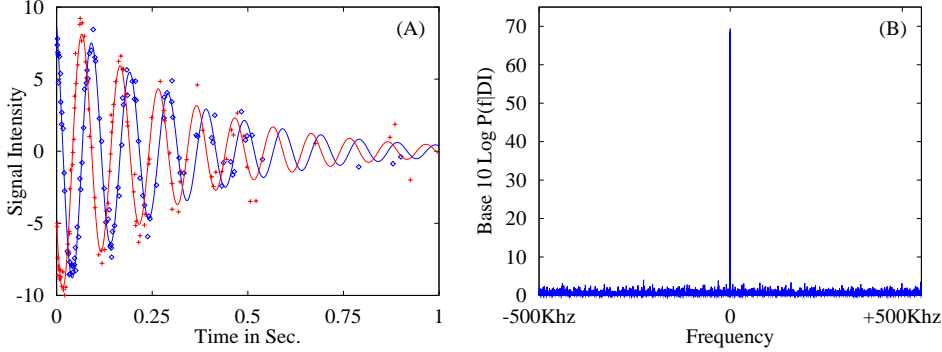


Fig. 3. Panel (A) is computer simulated data. It contains the same exponentially decaying sinusoidal signal as shown in Fig. 1(A) plus noise of the same standard deviation. The lines represent the real and imaginary parts of the sinusoid. The location of the nonuniformly nonsimultaneously sampled data values are denoted by the isolated characters. Panel (B) is the base 10 logarithm of the posterior probability for the frequency given a stationary sinusoid model using these data. Note that this plot spans 10,000 times the Nyquist critical frequency for the uniformly sampled version of this data set shown in Fig. 1(A).

data seem to have a bandwidth that is at least 10,000 times larger than in the first example?

We showed earlier in Eqs. (13)-(16) that aliases come about because of the integer j in $t_j = j\Delta T$, specifying the time at which each uniformly simultaneously sampled data item was acquired. In the present problem, nonuniformly nonsimultaneously sampled data, there is no ΔT such that all of the acquisition times are integer multiples of this time; not if the times are truly sampled randomly. However, all data and times must be recorded to finite accuracy. This is true even of the simulated data shown in Fig. 3(A). Consequently, there must be a largest effective dwell time, $\Delta T'$, such that all of the times (both the real and imaginary) must satisfy

$$t_l = k_l \Delta T' \quad t_l \in \{\text{Real } t_i \text{ or Imaginary } t_j\} \quad (44)$$

where k_l is an integer. The subscript l was added to k to indicate that each of the times t_l requires a different integer k_l to make this relationship true. Of course, this was also true for uniformly sampled data: its just that for uniformly sampled data the integers were consecutive, $k_l = 0, 1, \dots, N - 1$. The effective dwell time is always less than or equal to the smallest time interval between data items, and is the dwell time one would have had to acquire data at in order to obtain a uniformly sampled data set with data items at each of the times t_i and t_j . The effective dwell time, $\Delta T'$, can be used to define a Nyquist critical frequency

$$f_{Nc} = \frac{1}{2\Delta T'}. \quad (45)$$

Aliases *must* appear for frequencies outside the bandwidth defined from $\Delta T'$.

The reason that aliases must appear for frequencies outside this bandwidth can be made apparent in the following way. Suppose we have a hypothetical data set that is sampled at $\Delta T'$. Suppose further, the hypothetical data are zero everywhere except at the times we actually have data, and there the data are equal to the appropriate $d_R(t_i)$ or $d_I(t_j)$. If we now compute the discrete Fourier transform of this hypothetical data set, then by the analysis done in Eqs. (13)-(16) the Nyquist critical frequency of this data set is $1/2\Delta T'$ and frequencies outside the implied bandwidth are aliased. Now look at the definitions of T_1 and T_2 , Eqs. (30) and (31), for the data set we actually have. You will find that these quantities are just the real and imaginary parts of the discrete Fourier transform of our hypothetical data set. The zeros in the hypothetical data set cannot contribute to the sums in the discrete Fourier transform: they act only as place holders, and so the only part of the sums that survive are just where we have data. By construction that is just what Eqs. (30) and (31) are computing. So aliases must appear at frequencies greater than this Nyquist critical frequency.

For the data shown in Fig. 3, the times and the data were recorded to 8 decimal places in an ASCII file. This file was then used by another program to compute the posterior probability for the frequency, Eq. (37). Because the data were recorded to 8 decimal places, a good guess for the effective dwell time would be $\Delta T' = 10^{-8}$ Sec. This would correspond to a Nyquist critical frequency of $f_{Nc} = 5 \times 10^7$ Hz. The first alias of the 10 Hz frequency should appear at 100,000,010 Hz. If we evaluate the base 10 logarithm of the posterior probability at frequencies given $(10 + n \times 10^6)$ Hertz, we should see peaks at $n = 100, 200, 300$ etc. This plot is shown in Fig. 4. Note that the aliases are right at the expected frequencies. An extensive search from zero up to 100 MHz uncovered no aliases prior to the one just above 100 MHz. This suggests that the effective bandwidth of the 100 complex data values shown in Fig. 3(A) is 100 MHz! That is one million times larger than the bandwidth of the data shown in Fig. 1. Indeed, the effective dwell time, $\Delta T'$, was defined as the maximum time for which all of the t_i and t_j are integer multiples of $\Delta T'$. The Nyquist critical frequency computed from $\Delta T'$ is the *smallest* frequency for which the argument of the complex exponential in Eq. (14) is an integer multiple of 2π . Consequently, $1/2\Delta T'$ is the Nyquist critical frequency for these data and there are no aliases within this implied bandwidth.

The fact that the data was recorded to 8 decimal places and that the bandwidth of this simulated data set is 10^8 Hz brings up another curious thing about aliases. Aliasing is primarily a phenomenon that concerns the times in the discretely sampled data, the data itself are almost irrelevant to aliasing. In the example we are discussing the times were recorded to 8 decimal places. If we had recorded the times to only 7 decimal places the aliases would have been in different places. If the last significant digit in the times is truncated, the bandwidth will be at least a factor of 10 lower, and the aliases will be in different places. Of course we must qualify this somewhat, in the case we are discussing, a change in the 8'th decimal place changes the sines and cosines so little that the only noticeable effect is on the location of the aliases. However, if we were to continue truncating decimal places we will eventually reach the point where the times are so far from the correct

Figure 4. The $\text{Log}_{10}P(f|DI)$ At Intervals Of 1 MHz

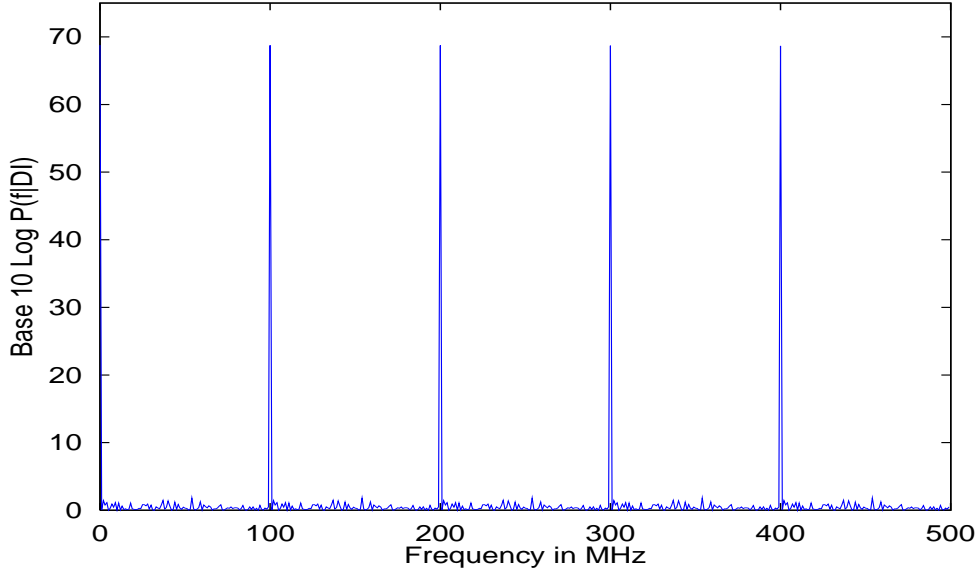


Fig. 4. Given the data shown in Fig. 3(A) aliases should appear every 100 MHz. Here we have evaluated the logarithm of the posterior probability for the frequency of a stationary sinusoid at frequencies given by $10 + n \times 10^6$ Hz. Aliases should appear at $n = 100, 200, 300, \text{etc.}$

values that we are essentially computing nonsense.

So far we have shown that the data have an effective bandwidth of $1/\Delta T'$ but that is not quite the same thing as showing that frequencies anywhere in this 100 MHz interval can be correctly estimated. There is another time, the minimum time between data values, ΔT_M , which might be of some importance. It might be thought that the data can show no evidence for frequencies outside a band of width $1/\Delta T_M$. In the data set shown in Fig. 3(A), $\Delta T_M = 0.00000488$ Sec. This would correspond to a bandwidth of roughly 200 kHz, a tremendous frequency range, but smaller than the effective bandwidth of 100 MHz by a factor of roughly 500. Which is correct?

The arguments given earlier prove that it is the effective dwell time, $\Delta T'$, and not the minimum interval between data values, ΔT_M , that is the important quantity. However, to illustrate that $\Delta T'$ is the critical time, it is a simple matter to generate data that contain a frequency in the range ($[1/\Delta T_M] < f < [1/\Delta T']$) and then compute the base 10 logarithm of posterior probability for the frequency of a stationary sinusoid over the entire range ($0 \leq f \leq 1/\Delta T'$). From a practical standpoint this is nearly impossible for data with a Nyquist critical frequency of 50 MHz; this frequency will have to be lowered. This can be done by truncating the exponentially sampled times to 5 decimal places, and then generating the simulated

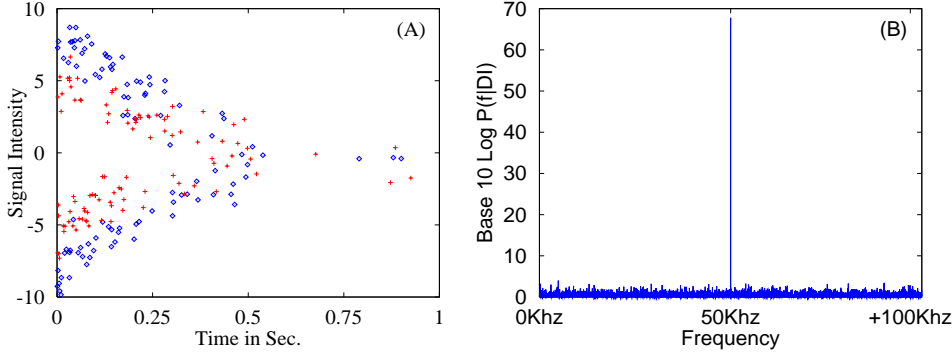
Figure 5. Which Is The Critical Time: $\Delta T'$ Or ΔT_M ?

Fig. 5. Panel (A) is computer simulated data. Except for the frequency, it contains the same signal as that shown in Fig. 1(A). The frequency in these data is 50 kHz. The positions of the nonuniformly nonsimultaneously sampled data values are denoted by the isolated characters. Panel (B) is the base 10 logarithm of the posterior probability for the frequency of a stationary sinusoid given these data.

signal using these truncated times. Truncating the times lowers the Nyquist critical frequency to 50 kHz. Additionally, when these simulated data shown in Fig. 5(A) were generated, no times were generated closer together than 0.0001 Sec. This time corresponds to a critical frequency of 5 kHz; so there is a factor of 10 difference between the Nyquist critical frequency and the frequency calculated from ΔT_M . The simulated acquisition parameters used in generating these data are similar to those used previously, *i.e.*, $N_R = N_I = 100$, $\alpha = 3 \text{ Sec.}^{-1}$, Amplitude = 10, $\sigma = 1$. The only difference is that the simulated frequency is 50 kHz, a full factor of 5 higher than the minimum sampling time would give as the highest resolvable frequency, and equal to the Nyquist critical frequency for these data. However, the region plotted ($0 \leq f \leq 100 \text{ kHz}$), has been shifted upward by 50 kHz, so the 50 kHz frequency is in the middle of the plotted region, and this region should be free of aliases. The base 10 logarithm of the posterior probability for the frequency of a stationary sinusoid is shown in Fig. 5(B). It was evaluated at every 1 Hz from 0 to 100,000 Hz. This frequency resolution is more than enough to ensure that if aliases exist, multiple peaks would be present in this plot. Note that there is a single peak and it is located at 50 kHz, the frequency of the simulated resonance. So the critical time is indeed $\Delta T'$ and the bandwidth of these data is 100 kHz.

Having shown that the critical time is the effective dwell time $\Delta T'$ and that there are no aliases in the full bandwidth implied by $\Delta T'$, one might be tempted to think that that is the end of the story. However, that is not quite correct. While there are no true aliases in the bandwidth implied by $\Delta T'$, it is possible for there to be multiple peaks which are artifacts related to the effective dwell time. These artifacts are not aliases in the sense that they are not exact replicas of the main peak; rather they are evidence for the resonance, and their height is

Figure 6. How Nonuniform Sampling Destroys Aliases

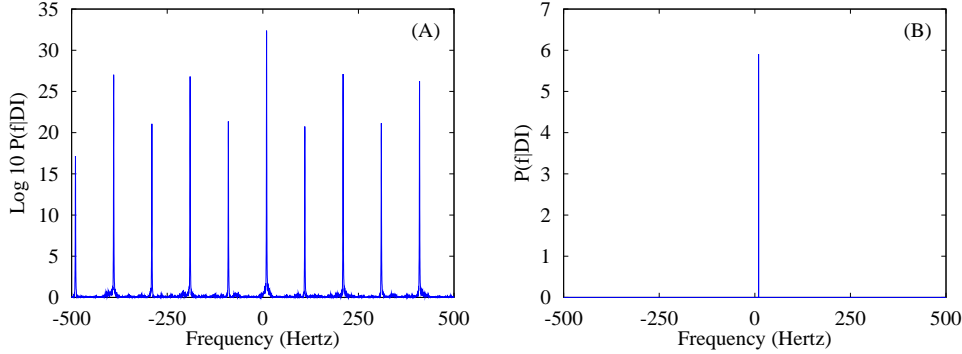


Fig. 6. Panel (A) is the base 10 logarithm of the posterior probability for the frequency of a stationary sinusoid using the data shown in Fig. 1(A) with four additional data values with sample times given by 0.001, 0.015, 0.025, and 0.037 seconds respectively. These four data values were generated using the same signal and signal-to-noise ratio as those shown in Fig. 1(A). The only difference is the nonuniformly sampled times. Note that we now have multiple peaks, but they are not true aliases because they are not of the same height. Panel (B) is the fully normalized posterior probability for the frequency of a stationary sinusoid given these data, note that all of the peaks except the one at the true frequency, 10 Hz, have been exponentially suppressed.

directly related to how strongly the data indicate the presence of a resonance at that frequency. To see how multiple peaks might occur, suppose we have the data shown in Fig. 1(A); with 4 additional nonuniformly but simultaneously sampled complex data values at 0.001, 0.015, 0.025, and 0.037 seconds respectively.⁵ These 4 complex data values were generated by sampling the same signal plus noise as shown in Fig 1(A), but at the four nonuniformly sampled times. Now, according to the analysis done in this paper, the Nyquist critical frequency for the combined data is $1/(2 \times 0.001 \text{ Sec.}) = 500 \text{ Hz}$; this bandwidth is exactly the same as the frequency interval shown in Fig. 2 where we had 10 aliases. In principle, these 4 complex data values should increased the bandwidth by a factor of 10 and should destroy the aliasing phenomenon in the $\pm 500 \text{ Hz}$ frequency band. A plot of the base 10 logarithm of the posterior probability for the frequency of a stationary sinusoid given the combined data is shown in Fig. 6(A). Note that there are 10 peaks in Fig. 6(A), just as there are 10 peaks in Fig. 2, but the peaks are not of the same height, so they are not true aliases. To determine which peak corresponds to the true frequency one must assign a bounded prior probability for the frequency (to eliminate the true aliases at frequencies above 500 Hz and below -500 Hz) and then normalize the posterior probability. The fully normalized posterior proba-

⁵The fact that the 4 complex data values are simultaneously sampled is irrelevant. The results of this analysis would be the same regardless if these 8 total data items were simultaneously sampled or not.

bility for the frequency of a stationary sinusoid is shown in Fig. 6(B). The fully normalized posterior probability density function has a single peak at 10 Hz, the true frequency. In this example, the 4 extra nonuniformly sampled complex data values were enough to raise the probability for the true frequency several orders of magnitude, and when the posterior probability for the frequency was normalized, the vast majority of the weight in the posterior distribution was concentrated around the true frequency, consequently all of the spurious peaks seen in Fig. 6(A) were exponentially suppressed.

In preparing this example the number of nonuniformly sampled data values had to be chosen. Initially 6 complex nonuniformly sampled data values were generated. But this was abandoned because, when the base 10 logarithm of the posterior probability for a stationary frequency was plotted, the spurious peaks were only about 5 percent of the main peak and so did not illustrate the points we wanted to make. However, it does indicate that it does not take many nonuniformly sampled data values to eliminate these spurious peaks. As few as 10 percent should completely eliminate them. Nonetheless, it is possible for the fully normalized posterior probability for the frequency to have multiple peaks in data that contain only a single frequency. If this happens, it indicates that the data simply cannot distinguish which of the possibilities are the true frequency. The only recourse will be to obtain more measurements, preferably at nonuniformly nonsimultaneously sampled times.

The preceding example reiterates what has been said several times: the discrete Fourier transform power spectrum, the Schuster periodogram, a weighted power spectrum, the Lomb-Scargle periodogram, and the generalizations to these presented in this paper are sufficient statistics for frequency estimation given a *single* frequency model. Multiple peaks in the discrete Fourier transform and its generalizations are not necessarily evidence for multiple frequencies. The only way to be certain that multiple frequencies are present is to postulate models containing one, two, etc. frequencies and to then compute the posterior probability for these models. Depending on the outcome of that calculation, one can then estimate the frequencies from the appropriate model.

6. Parameter Estimates

So far the discussions have concentrated on the discrete Fourier transform, how probability theory generalizes it to nonuniformly nonsimultaneously sampled data, and how these generalizations affect aliases. Now we are going to discuss the effect of nonuniform nonsimultaneous sampling on the parameter estimates. In this discussion we are going to estimate the parameters using the data shown in Fig. 1(A) and in Fig. 3(A). These two data sets contain exactly the same signal and each data set contains Gaussian white noise drawn from a Gaussian random number generator of unit standard deviation. However, the noise realizations in each data set are different, and this will result in slightly different parameter estimates for each data set. Nonetheless these two data sets provide an excellent opportunity to demonstrate how nonuniform nonsimultaneous sampling affects the parameter estimates.

Figure 7. Estimating The Parameters

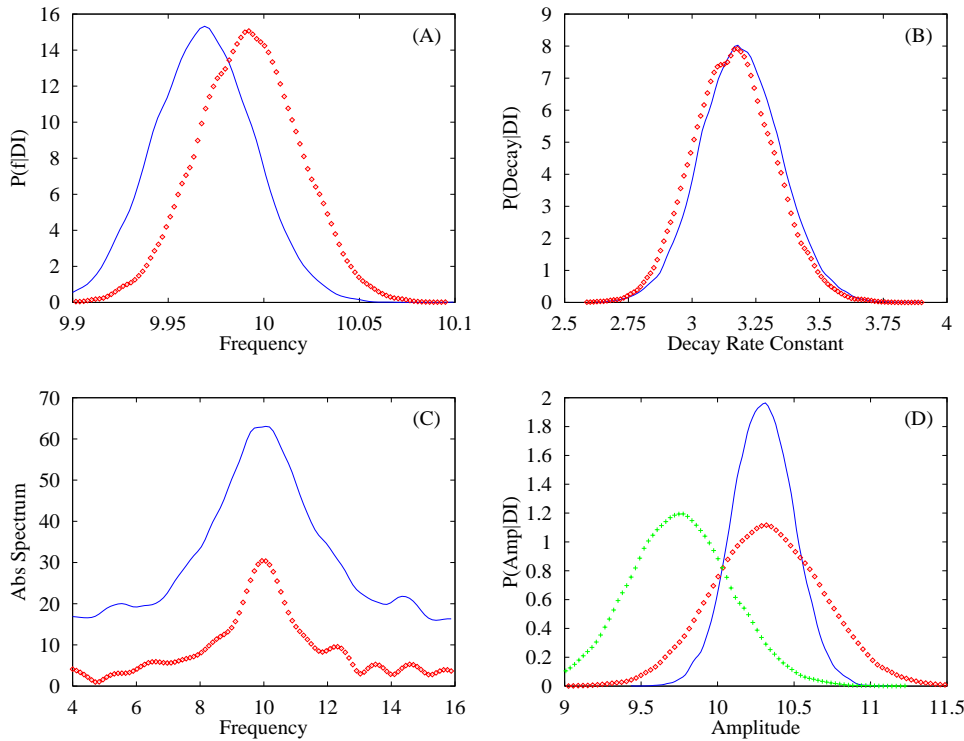


Fig. 6. The posterior probability of the parameter of interest was computed given the nonuniformly nonsimultaneously sampled data, solid lines, and was computed given the uniformly sampled data, open characters. Panel (A) is the posterior probability for the frequency, (B) the posterior probability of the decay rate constant, (D) the posterior probability for the amplitude. Panel (C) is the absolute-value spectrum computed for the uniformly sampled data and for the nonuniformly nonsimultaneously sampled data. The extra curve in panel (D), the plus signs, is the posterior probability for the amplitude computed from a nonuniformly nonsimultaneously sample data set having exactly the same signal with exactly the same signal-to-noise level but with times that were generated from a uniform random number generator.

We will discuss estimation of the frequency, decay rate constant and the amplitude. We will not discuss estimation of the phase and standard deviation of the noise as these are of less importance. The posterior probability for the frequency, decay rate constant, and amplitude are shown in panels (A), (B) and (D) of Fig. 7 respectively. Each of these plots is the fully normalized marginal posterior probability for the parameter of interest independent of all of the other parameters appearing in the model. Panel (C) contains the absolute-value spectra computed from these two data sets and will be used to compare Fourier transform estimation

procedures to the Bayesian calculations. The solid lines in these plots were computed from the nonuniformly nonsimultaneously sampled data shown in Fig. 3(A); while the curves drawn with open characters were computed using the uniformly sampled data shown in Fig. 1(A).

A Markov chain Monte Carlo simulation was used to compute the marginal posterior probability for each parameter. All of the parameters appearing in the model were simulated simultaneously, thus the Markov chain Monte Carlo simulation simulated the joint posterior probability for all the parameters. This was done for computational convenience; *i.e.*, it was easier to do a single Markov chain Monte Carlo simulation than to do five separate calculations, one for each parameter appearing in the model. Because the probability density functions shown in panels (A), (B) and (D) were formed by computing a histogram of the Markov chain Monte Carlo samples there are small irrelevant artifacts in these plots that are related to the number of samples drawn from the simulation. For more on Markov chain Monte Carlo methods and how these can be used to implement Bayesian calculations see Gilks [18] and Radford [19].

The marginal posterior probability for the frequency is shown in panel (A). This is the fully normalized marginal posterior probability for the frequency independent of all of the other parameters, Eq. (38). Note that the true frequency, 10 Hz, is well covered by the posterior probability computed from both the uniformly (open characters) and nonuniformly nonsimultaneously (solid line) sampled data. Also note that these distributions are almost identical in height and width. Consequently, both the uniform and nonuniformly nonsimultaneously sampled data have given the same parameter estimates to within the uncertainty in these estimates. Of course the details for each estimated differ, because the noise realizations in each data set differ. Consequently, the frequency estimate is not strongly dependent on the sampling scheme. Indeed this can be derived from the rules of probability theory with the following proviso: the two sampling schemes must cover the *same* total sampling time and must sample the signal in a reasonably dense fashion so that sums may be approximated by integrals, [11,16]. Having said this, we must reemphasize that this is only true for frequency estimates using data having sampling schemes covering the *same* total sampling time; it is not true if the sampling times differ nor is it necessarily true of the other parameters appearing in the model. Indeed one can show that for a given number of data values, the precision of the frequency estimate for a stationary sinusoid is inversely proportional to the total sampling time, provided one samples the signal in a reasonably dense fashion. Thus, sampling 10 times longer will result in frequency estimates that are 10 times more precise. As noted in Bretthorst [11] this is equivalent to saying that for frequency estimation data values at the front and back of the data are most important in determining the frequency, because it is in these data that small phase differences are most highly magnified by the time variable. This could be very useful in some applications; but of course it will not help in NMR applications where the signal decays away.

In addition to the posterior probability for the frequency, we have also plotted the absolute-value spectra computed from these two data sets, Fig. 7(C). Note that the peaks of these two absolute-value spectra are at essentially the same

frequency as the corresponding peaks in panel (A); although they are plotted on differing scales. If the absolute value spectrum is used to estimate the frequency, one would typically use the peak frequency as the estimate, and then claim roughly the half-width-at-half-height as the uncertainty in this estimate. For these two data sets that is about 10 plus or minus 2 Hz. The two fully normalized posterior probabilities shown in panel (A) span a frequency interval of only 0.2 Hz. This frequency interval is roughly 6 standard deviations. Thus the frequency has been estimated to roughly 10 Hz with an uncertainty of $0.2/6 \approx 0.03$ Hz; a 60 fold reduction in the uncertainty in the frequency estimate.

One last note before we begin the discussion of estimating the decay rate constant, we reiterate that all of the details in the wings of the absolute-value spectrum shown in panel (C) are irrelevant to the frequency estimation process. The posterior probability for the frequency has peaked in a region that is very small compared to the scale of these wings, all of the information about the frequency estimate is contained in a very small region around the largest peak in the absolute-value spectrum.

The marginal posterior probability for the decay rate constant is shown in Fig. 7(B). Here we again find that the parameter estimates from both data sets are essentially identical in all of their relevant details. Both probabilities peak at nearly the same value of the decay rate constant, both have nearly the same width, and therefore the same standard deviation; thus like frequency estimates, the estimates for the decay rate constants do not strongly depend on the sampling scheme. In principle the accuracy of the estimates for the decay rate constants scale with time just like the frequency estimates, of course, with decaying signals this is of little practical importance. Note that the decay rate constant has been estimated to be about $3.2 \pm 0.3 \text{ Sec.}^{-1}$ at one standard deviation. The true value is 3 Sec.^{-1} , so both sampling schemes give reasonable estimates of the decay rate. If one were to try and estimate the decay rate constant from the absolute-values spectrum, the half-width-at-half-height would normally be used, here that is about 2 Sec.^{-1} and no claim about the accuracy of the estimate would be made.

The marginal posterior probability for the amplitude of the sinusoid is shown in Fig. 7(D). The amplitude, A , was defined in Eq. (17). In this paper we did not directly talk about amplitude estimation (see Bretthorst [16] for a discussion of this subject), rather we treated the amplitudes of the sine and cosine model functions as nuisance parameters and removed them from the posterior probability for the other parameters. We did this because we wished to explore the relationships between frequency estimation using Bayesian probability theory and the discrete Fourier transform. However, the Markov chain Monte Carlo simulation used $A \cos(2\pi ft + \theta) \exp\{-\alpha t\}$ as the model for the real data, so it was a trivial matter to compute the posterior probability for the amplitude. If you examine Fig. 7(D) you will note that now we do have a difference between the uniform (open characters) and the nonuniformly nonsimultaneously sampled data (solid lines). The amplitude estimates from the nonuniformly nonsimultaneously sampled data are a good factor of 2 more precise than the estimates from the uniformly sampled data. One might think that this is caused by the nonuniform nonsimultaneous sampling and this would be correct, but not for the obvious rea-

sons. If you examine panel (D) you will note that we have plotted a third curve (plus signs). This curve is the posterior probability for the amplitude computed from data with the exact same signal and signal-to-noise ratio, but having times that are nonuniformly nonsimultaneously sampled where the times were generated from a uniform random number generator. We will call this data set the uniform-randomly sampled data. Note that the height and width of the posterior probabilities computed from both the uniformly and the uniform-randomly sampled data are essentially the same, so by itself the nonuniform nonsimultaneous sampling did not cause the amplitude estimates to improve. The amplitude estimate improved because exponential sampling gathered more data where the signal was large. The accuracy of the amplitude estimate is proportional to the standard deviation of the noise and inversely proportional to square root of the effective number of data values. Because exponential sampling gathered more data where the signal was large, its effective number of data values was larger and so the amplitude estimate improved. In this case, the improvement was about a factor of 2, so the exponential sampling had an effective number of data values that was about a factor of 4 larger than for the uniformly or uniform-randomly sampled data. This fact is also reflected in differing heights of the absolute value spectra plotted in Fig. 7(C). The peak height of an absolute value spectrum is proportional to the square root of the effective number of data values. In panel (C) the spectra computed from the uniformly sampled data set, open characters, is roughly a factor of 2 lower than the height of the spectrum computed from the exponentially sampled data set, solid line.

7. Summary and Conclusions

Probability theory generalizes the discrete Fourier transform roughly as follows: in uniformly simultaneously sampled complex data, the sufficient statistic for estimating the frequency of a single stationary sinusoid is the power spectrum or Schuster periodogram. When the signal is exponentially decaying, the appropriate generalization is an exponentially weighted normalized power spectrum. The normalization constant is an effective number of data values. When the data are nonuniformly simultaneously sampled the sufficient statistic remains unchanged except the times must be explicitly used in the calculation of the periodogram. Finally, when the data are nonuniformly nonsimultaneously sampled, the sine and cosine transforms making up the discrete Fourier transform must properly account for the differing acquisition times in each channel, thus each sine and cosine transform must have its own summation index. The sufficient statistic must also take into account the differing numbers of data in each channel as well as the nonorthogonality of the nonsimultaneously sampled sinusoidal model functions.

In a literal sense, probability theory does no such thing as generalize the discrete Fourier transform. Probability theory simply tells one how to analyze a particular problem optimally. For estimation of a harmonic frequency, the sufficient statistics turn out to be related to the discrete Fourier transform. This was, for us, a happy coincidence because it enabled us to interpret the results of the analysis in a way that sheds light on the discrete Fourier transform and how it should be used. In the

appropriate limits, the discrete Fourier transform power spectrum, the Schuster periodogram, the Lomb-Scargle periodogram and the generalizations presented in this paper are all optimal frequency estimators for the *single* sinusoidal case. However, when the true signal deviate from this model, for example when there are multiple sinusoids or the data contain a constant offset, then these statistics are *never* optimal frequency estimators, and there are always other statistics that will improve the resolution of the multiple frequencies or properly account for a constant offset in the data, see [11,7–9].

Aliasing is a general phenomenon and exists in both uniformly and nonuniformly nonsimultaneously sampled data for exactly the same reason. It is the fact that all of the times may be expressed as an integer multiple of an effective dwell time that is the primary cause of aliasing. Two data sets differing only in how precisely the times are recorded generally have different Nyquist critical frequencies.

The analysis in this paper generalized the concept of bandwidth and showed that uniformly simultaneously sampled data have the smallest possible bandwidth. The addition of any nonuniformly nonsimultaneously sampled data always increases the Nyquist critical frequency and thus increases the bandwidth. The Nyquist critical frequency for nonuniformly nonsimultaneously sampled data may be many orders of magnitude greater than that for uniformly simultaneously sampled data having similar acquisition parameters. Consequently, nonuniformly nonsimultaneously sampled data can have tremendous advantages over uniformly sampled data because the critical time is not how fast one can sample data, but how accurately one can vary the acquisition of each data item. This opens up the possibility of measuring very high frequencies with bandwidths much larger than previously possible.

Acknowledgements

The author would like to thank Joe Ackerman, Jeffrey Neil, Dmitriy Yablonskiy, Jim Quirk, C. Ray Smith, Jeffrey Scargle, Pauline Bretthorst, and Josh Crowell for their comments on preliminary versions of this paper.

A. The Lomb-Scargle Periodogram

The Lomb-Scargle periodogram can be derived as the sufficient statistic for frequency estimation using Bayesian probability theory. As in all such estimation problems, the data must be related to the parameter of interest through a model. The model used by Lomb [1] is

$$d(t_i) = A \cos(2\pi f t_i - \theta) + B \sin(2\pi f t_i - \theta) \quad (46)$$

with θ chosen as a frequency dependent phase shift that makes the sine and cosine model functions orthogonal on the discretely sampled time points; namely

$$\theta = \frac{1}{2} \tan^{-1} \left(\frac{U}{V} \right) \quad (47)$$

with

$$U \equiv \sum_{i=1}^N \sin(4\pi f t_i) \quad \text{and} \quad V \equiv \sum_{i=1}^N \cos(4\pi f t_i). \quad (48)$$

Using the Lomb model, Eq. (46), if we now apply the rules of probability theory to compute the posterior probability for the frequency, f , independent of the amplitudes and variance of the noise, one obtains

$$P(f|DI) \propto \frac{1}{\sqrt{CS}} \left[N\overline{d^2} - \overline{h^2} \right]^{\frac{2-N}{2}} \quad (49)$$

where $\overline{d^2}$ is the mean-square data value. The sufficient statistic, $\overline{h^2}$, is given by

$$\overline{h^2} = \frac{R_{LS}(f)^2}{C} + \frac{I_{LS}(f)^2}{S} \quad (50)$$

and

$$\begin{aligned} R_{LS}(f) &\equiv \sum_{i=1}^N d(t_i) \cos(2\pi f t_i - \theta), \\ I_{LS}(f) &\equiv \sum_{i=1}^N d(t_i) \sin(2\pi f t_i - \theta), \\ C &\equiv \sum_{i=1}^N \cos^2(2\pi f t_i - \theta), \\ S &\equiv \sum_{i=1}^N \sin^2(2\pi f t_i - \theta). \end{aligned} \quad (51)$$

In deriving this result, uniform prior probabilities were used for A and B and a Jeffreys' prior was used for the standard deviation of the noise. Written in this form the sufficient statistic, Eq. (50), is an exact analogue to a normalized power spectrum of the data, and so brings out the relationship to a discrete Fourier transform in a very elegant way.

The Lomb model, Eq. (46), may be transformed into the stationary sinusoidal model used in this paper by the following change of variables

$$\begin{aligned} A' &= A \cos \theta - B \sin \theta \\ B' &= B \cos \theta + A \sin \theta \end{aligned} \quad (52)$$

and the Lomb model becomes

$$d(t_i) = A' \cos(2\pi f t_i) + B' \sin(2\pi f t_i) \quad (53)$$

where the amplitudes, A' and B' , are the amplitudes used in the formulation of the problem given in this paper. This simple change of variables cannot change the resulting sufficient statistics; because the total projection of the model onto the

data remains the same. Consequently, Eq. 37, and Eq. (50) give identical frequency estimates (after account for differing sign conventions). However, use of the Lomb model to estimate the sine and cosine amplitudes of the sinusoid will result in estimates that are shifted from the true values because of the phase θ . If the Lomb model were rewritten to include the amplitude of the sinusoid, $\sqrt{A^2 + B^2}$, and this model used to estimate the amplitude, this transformed model would estimate the amplitude of the sinusoid correctly even though it does not estimate the sine and cosine amplitudes correctly.

Last, the Lomb-Scargle periodogram is usually normalized by an estimate of the standard deviation of the noise. For example, in Numerical Recipes [14] it is divided by essentially 2 times the mean-square data value. However, this cannot possibly be a very sensible normalization if for no other reason than as the noise goes to zero the accuracy of the frequency estimate remains finite; while the posterior probability, Eq. (49), goes into a delta function. As explained in Bretthorst [11], around the maximum of the posterior probability, the posterior probability is well approximated by

$$P(f|DI) \approx \exp \left\{ \frac{\overline{h^2}}{2\langle\sigma^2\rangle} \right\} \quad (54)$$

where $\langle\sigma^2\rangle$, the expected variance of the noise, is given by

$$\langle\sigma^2\rangle = \frac{1}{N-4} \left[N\overline{d^2} - \overline{h^2} \right] \quad (55)$$

and this is essentially the mean-square residual as a function of frequency. As the noise goes to zero, the expected variance of the noise goes to zero, and so, this normalization of the periodogram goes into a delta function, as it should. Consequently, the Lomb-Scargle periodogram, at least as implemented in Numerical Recipes, will misestimate the uncertainty in the parameter estimates and this error could be quite large depending on the signal-to-noise ratio of the data.

References

1. Lomb, N. R. (1976) “Least-Squares Frequency Analysis of Unevenly Spaced Data,” *Astrophysical and Space Science*, **39**, pp. 447-462.
2. Scargle, J. D. (1982) “Studies in Astronomical Time Series Analysis II. Statistical Aspects of Spectral Analysis of Unevenly Sampled Data,” *Astrophysical Journal*, **263**, pp. 835-853.
3. Scargle, J. D. (1989) “Studies in Astronomical Time Series Analysis. III. Fourier Transforms, Autocorrelation and Cross-correlation Functions of Unevenly Spaced Data,” *Astrophysical Journal*, **343**, pp. 874-887.
4. Schuster, A., (1905), “The Periodogram and its Optical Analogy,” *Proceedings of the Royal Society of London*, **77**, p. 136.
5. Nyquist, H., (1928), “Certain Topics in Telegraph Transmission Theory,” *Transactions AIEE*, **3**, p. 617.
6. Nyquist, H., (1924), “Certain Factors Affecting Telegraph Speed,” *Bell System Technical Journal*, **3**, p. 324.
7. Bretthorst, G. Larry (1990), “Bayesian Analysis I, Parameter Estimation” *J. Magn. Reson.*, **88**, pp. 533-551.
8. Bretthorst, G. Larry (1990), “Bayesian Analysis II, Model Selection” *J. Magn. Reson.*, **88**, pp. 552-570.

9. Bretthorst, G. Larry (1990), "Bayesian Analysis III, Examples Relevant to NMR" *J. Magn. Reson.*, **88**, pp. 571-595.
10. Woodward, P. M. (1953), *Probability and Information Theory, with Applications to Radar*, McGraw-Hill, N. Y.
11. Bretthorst, G. Larry (1988), "Bayesian Spectrum Analysis and Parameter Estimation," in *Lecture Notes in Statistics*, **48**, J. Berger, S. Fienberg, J. Gani, K. Krickenberg, and B. Singer (eds), Springer-Verlag, New York, New York.
12. Jaynes, E. T. (1993), "Probability Theory—The Logic of Science," copies of this manuscript are available from "bayes.wustl.edu".
13. Bretthorst, G. Larry (1999), "The Near-Irrelevance of Sampling Frequency Distributions," in *Maximum Entropy and Bayesian Methods*, W. von der Linden *et al.* (eds.), pp. 21-46, Kluwer Academic Publishers, the Netherlands.
14. Press, William H., Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling (1992), *Numerical Recipes: The Art of Scientific Computing, Second Edition*, Cambridge Univ. Press, Cambridge.
15. Bretthorst, G. Larry (1991), "Bayesian Analysis. IV. Noise and Computing Time Considerations," *J. Magn. Reson.*, **93**, pp. 369-394.
16. Bretthorst, G. Larry (1992), "Bayesian Analysis. V. Amplitude Estimation for Multiple Well-Separated Sinusoids," *J. Magn. Reson.*, **98**, pp. 501-523.
17. Bretthorst, G. Larry (1992), "Estimating The Ratio Of Two Amplitudes In Nuclear Magnetic Resonance Data," in *Maximum Entropy and Bayesian Methods*, C. R. Smith *et al.* (eds.), pp. 67-77, Kluwer Academic Publishers, the Netherlands.
18. Gilks, W. R., S. Richardson and D. J. Spiegelhalter (1996), "Markov Chain Monte Carlo in Practice," Chapman & Hall, London.
19. Neal, Radford M. (1993), "Probabilistic Inference Using Markov Chain Monte Carlo Methods," technical report CRG-TR-93-1, Dept. of Computer Science, University of Toronto.