

# Frequency Estimation And Generalized Lomb-Scargle Periodograms

G. Larry Bretthorst

**ABSTRACT** Using Bayesian probability theory we demonstrate that the Lomb-Scargle periodogram may be generalized in a straightforward manner to nonuniformly nonsimultaneously sampled quadrature data when the sinusoid has arbitrary amplitude modulation. This generalized Lomb-Scargle periodogram is the sufficient statistic for single frequency estimation in a wide class of problems ranging from stationary frequency estimation in real uniformly sampled data, to frequency estimation for a single sinusoid having exponential, Gaussian, or arbitrary amplitude modulation. In addition we define the bandwidth of a nonuniformly sampled data set and show that the phenomenon of aliases exists in both uniformly and nonuniformly sampled data and that the phenomenon has the same cause in both types of data. Finally, we show that nonuniform sampling does not affect the accuracy of the frequency estimates; although it may affect the accuracy of the amplitude estimates.

## 1 Introduction

The problem of estimating the frequency or period of a sinusoid arises in an a host of contexts in the sciences. For example, in NMR the signals are sinusoidal with exponential decay. In Meteorology temperature data obviously fluctuate sinusoidal on a daily and yearly basis. In astrophysics the period of variables stars may be on the order of days to years with nonstationary nonsinusoidal oscillations about a trend. The data gathered by the different sciences is almost as varied as phenomena being observed. In NMR the quadrature data are almost always uniformly sampled (a quadrature measurement is one in which a measurement of both the real and imaginary components of the sinusoids has been made). In astrophysics the data are often magnitudes sampled at unevenly spaced intervals; while other astrophysics data may consists of measured velocities, etc.

The standard way to deal with such data is to compute a discrete Fourier transform of the data and then view the transform as an absorption spectrum, a power spectrum, a Schuster periodogram (Schuster 1905), or a Lomb-Scargle periodogram (Lomb 1976, and Scargle 1982 and 1989), see Priestley (1981) and Marple 1987 for a review of classical spectral estima-

tion techniques. The problem with all such techniques is that they have not been derived from any single set of unifying principles that tell one what is the optimal way to estimate the period. In this paper we change that by using Bayesian probability theory to derive the discrete Fourier transform, the power spectrum, the weighted power spectrum, the Schuster periodogram and the Lomb-Scargle periodogram as special cases of a generalized Lomb-Scargle periodogram, and show that the generalized Lomb-Scargle periodogram is a sufficient statistic for single frequency estimation, (a sufficient statistic summarizes all of the information in the data relevant to the question being asked).

In Bayesian probability theory are two basic rules for manipulating probabilities, the product rule and the sum rule; all other rules may be derived from these. If  $A$ ,  $B$ , and  $C$  stand for three arbitrary hypotheses, then the product rule states

$$P(AB|C) = P(A|C)P(B|AC), \tag{1.1}$$

where  $P(AB|C)$  is the joint probability that “ $A$  and  $B$  are true given that  $C$  is true,”  $P(A|C)$  is the probability that “ $A$  is true given  $C$  is true,” and  $P(B|AC)$  is the probability that “ $B$  is true given that both  $A$  and  $C$  are true.”

In Aristotelian logic, the hypothesis “ $A$  and  $B$ ” is the same as “ $B$  and  $A$ ,” so the numerical value assigned to the probabilities for these hypotheses must be the same. The order may be rearranged in the product rule, Eq. (1.1), to obtain:

$$P(BA|C) = P(B|C)P(A|BC), \tag{1.2}$$

which may be combined with Eq. (1.1) to obtain a seemingly trivial result

$$P(A|BC) = \frac{P(A|C)P(B|AC)}{P(B|C)}. \tag{1.3}$$

This is Bayes’ theorem. It is named after Rev. Thomas Bayes, an 18th century mathematician who derived a special case of this theorem. Bayes’ calculations were published in 1763, two years after his death. This theorem, as generalized by Laplace, is the basic starting point for inference problems using probability theory as logic.

The second rule of probability theory, the sum rule, relates to the probability for “ $A$  or  $B$ .” The operation “or” is indicated by a  $+$  inside a probability symbol. The sum rule states that given three hypotheses  $A$ ,  $B$ , and  $C$ , the probability for “ $A$  or  $B$  given  $C$ ” is

$$P(A + B|C) = P(A|C) + P(B|C) - P(AB|C). \tag{1.4}$$

If the hypotheses  $A$  and  $B$  are mutually exclusive, that is the probability  $P(AB|C)$  is zero, the sum rule becomes:

$$P(A + B|C) = P(A|C) + P(B|C). \tag{1.5}$$

The sum rule is especially useful because it allows one to investigate an interesting hypothesis while removing an uninteresting or nuisance hypothesis from consideration.

To illustrate how to use the sum rule to eliminate nuisance hypotheses, suppose  $D$  stands for the data,  $f$  the hypothesis “the frequency of a sinusoidal oscillation was  $f$ ,” and  $B$  the hypothesis “the amplitude of the sinusoid was  $B$ .” Now suppose one wishes to compute the probability for the frequency given the data,  $P(f|D)$ , but the amplitude  $B$  is present and must be dealt with. The way to proceed is to compute the joint probability for the frequency and the amplitude given the data, and then use the sum rule to eliminate the amplitude from consideration. Suppose, for argument’s sake, the amplitude  $B$  could take on only one of two mutually exclusive values  $B \in \{B_1, B_2\}$ . If one computes the probability for the frequency and ( $B_1$  or  $B_2$ ) given the data one has

$$P(f|D) \equiv P(f[B_1 + B_2]|D) = P(fB_1|D) + P(fB_2|D). \quad (1.6)$$

This probability distribution summarizes all of the information in the data relevant to the estimation of the frequency  $f$ . The probability  $P(f|D)$  is called the marginal probability for the frequency  $f$  given the data  $D$ .

The marginal probability  $P(f|D)$  does not depend on the amplitudes at all. To see this, the product rule is applied to the right-hand side of Eq. (1.6) to obtain

$$P(f|D) = P(B_1|D)P(f|B_1D) + P(B_2|D)P(f|B_2D) \quad (1.7)$$

but

$$P(B_1|D) + P(B_2|D) = 1 \quad (1.8)$$

because the hypotheses are exhaustive. So the probability for the frequency  $f$  is a weighted average of the probability for the frequency given that one knows the various amplitudes. The weights are just the probability that each of the amplitudes is the correct one. Of course, the amplitude could take on more than two values; for example if  $B \in \{B_1, \dots, B_m\}$ , then the marginal probability distribution becomes

$$P(f|D) = \sum_{j=1}^m P(fB_j|D), \quad (1.9)$$

provided the amplitudes are mutually exclusive and exhaustive. In many problems, the hypotheses  $B$  could take on a continuum of values, but *as long as only one value of  $B$  is realized when the data were taken* the sum rule becomes

$$P(f|D) = \int dB P(fB|D). \quad (1.10)$$

Note that the  $B$  inside the probability symbols refers to the hypothesis; while the  $B$  appearing outside of the probability symbols is a number or

index. A notation could be developed to stress this distinction, but in most cases the meaning is apparent from the context.

## 2 Frequency Estimation—The Generalized Lomb-Scargle Periodogram

The problem addressed is the estimation of the frequency,  $f$ , of a sinusoid having known arbitrary amplitude modulation given nonuniformly non-simultaneously sampled quadrature data. To apply Bayesian probability theory to any problem one must relate the parameter of interest, here the frequency, to the measured data. This is done through a model. For a sinusoid having arbitrary amplitude modulation, the frequency may be related to the real data by

$$d_R(t_i) = A \cos(2\pi f t_i - \theta) Z(t_i) + B \sin(2\pi f t_i - \theta) Z(t_i) + n_R(t_i) \quad (1.11)$$

where  $d_R(t_i)$  denotes the real data measured at time  $t_i$ ,  $A$  and  $B$  are the cosine and sine amplitudes of the sinusoid, and  $n_R(t_i)$  denotes the noise at time  $t_i$ . Following Lomb's example,  $\theta$  will be defined in such a way as to make the cosine and sine functions orthogonal on the discretely sampled times. The function  $Z(t_i)$  specifies the amplitude modulation of the sinusoid;  $Z(t)$  could be an exponential, a Gaussian, or any other function appropriate to the signal being modeled. If  $Z(t)$  is a function of any parameters, these parameters are presumed known; for example, if  $Z(t)$  is a decaying exponential, then we assume the decay rate constant is known. Of course, in any Bayesian analysis we could turn our attention to the parameters in  $Z(f)$  and estimate them, but for this problem we will consider them as known and suppress them from the notation.

In a quadrature data set one also has a measurement of the imaginary or quadrature part of the signal. The imaginary data are  $90^\circ$  out of phase with the real data. Here this means that model for the imaginary data is  $90^\circ$  out of phase with the model for the real data:

$$d_I(t'_j) = -A \sin(2\pi f t'_j - \theta) Z(t'_j) + B \cos(2\pi f t'_j - \theta) Z(t'_j) + n_I(t'_j). \quad (1.12)$$

We have labeled the times at which the imaginary data were acquired with a prime superscript to distinguish them from the times at which the real data were acquired and we have added a subscript,  $I$ , to several quantities to indicate that these quantities refer to the imaginary part of the signal. The total number of data values will be designated as  $N = N_R + N_I$ , where  $N_R$  and  $N_I$  are the number of data values in the real and imaginary channels respectively;  $N_R$  and  $N_I$  need not be equal and can be zero.

In Bayesian probability theory all of the information in the data relevant to the problem being solved is summarized in a probability density function.

For the problem of estimating the frequency, this probability is denoted as  $P(f|D_R D_I I)$ , where this should be read as the posterior probability for the frequency  $f$  given the real data  $D_R$ , the imaginary data  $D_I$  and the prior information  $I$ . In this probability all of the arguments are hypotheses. For example  $f$  stands for a hypotheses of the form “at the time the data were take the value of the frequency was  $f$ .” Thus probability theory is ranking a whole series of models, one for each value of  $f$ , and the width of the posterior probability is a natural measure of how uncertain one is of the frequency. The hypotheses  $I$  appearing in  $P(f|D_R D_I I)$  refers to all of our prior information—including the model equations—and does not refer to the imaginary data; rather it refers to the general background information that goes into making this a well posed problem.

Using the sum rule of probability theory, Eq. (1.5), the posterior probability for the frequency is computed from the joint posterior probability for all of the parameters:

$$P(f|DI) = \int dA dB d\sigma P(fAB\sigma|D_R D_I I) \quad (1.13)$$

where  $\sigma$  is the standard deviation of the Gaussian noise prior probability. The right-hand side of this equation may be factored using Bayes’ theorem, Eq. (1.3), and the product rule, Eq. (1.1); one obtains

$$\begin{aligned} P(f|D_R D_I I) &\propto \int dA dB d\sigma P(f|I)P(A|I)P(B|I)P(\sigma|I) \\ &\times P(D_R|fAB\sigma I)P(D_I|fAB\sigma I) \end{aligned} \quad (1.14)$$

where we have assumed logical independence of the parameters, and that the standard deviation of the noise prior probability is the same for both the real and imaginary data; i.e., our prior information indicate that real and imaginary data have the same noise levels.

If we assign uniform prior probabilities to  $P(f|I)$ ,  $P(A|I)$ ,  $P(B|I)$ , a Jeffreys’ prior  $(1/\sigma)$  to  $P(\sigma|I)$ , and assign the two likelihoods using Gaussian noise prior probabilities, one obtains:

$$P(f|DI) \propto \int_{-\infty}^{\infty} dA \int_{-\infty}^{\infty} dB \int_0^{\infty} d\sigma \sigma^{-(N+1)} \exp \left\{ -\frac{Q}{2\sigma^2} \right\} \quad (1.15)$$

where

$$Q \equiv N\overline{d^2} - 2AR(f) - 2BI(f) + A^2C(f) + B^2S(f). \quad (1.16)$$

The mean-square data value,  $\overline{d^2}$ , is defined as

$$\overline{d^2} = \frac{1}{N} \left[ \sum_{i=1}^{N_R} d_R(t_i)^2 + \sum_{j=1}^{N_I} d_I(t'_j)^2 \right]. \quad (1.17)$$

The function  $R(f)$  is defined as

$$\begin{aligned}
 R(f) &\equiv \sum_{i=1}^{N_R} d_R(t_i) \cos(2\pi f t_i - \theta) Z(t_i) \\
 &\quad - \sum_{j=1}^{N_I} d_I(t'_j) \sin(2\pi f t'_j - \theta) Z(t'_j)
 \end{aligned} \tag{1.18}$$

which for uniformly sampled data reduces to the real part of a weighted discrete Fourier transform of the complex data. The function  $Z(t)$  plays the role of the weight or apodizing function. Similarly, the function  $I(f)$  is defined as

$$\begin{aligned}
 I(f) &\equiv \sum_{i=1}^{N_R} d_R(t_i) \sin(2\pi f t_i - \theta) Z(t_i) \\
 &\quad + \sum_{j=1}^{N_I} d_I(t'_j) \cos(2\pi f t'_j - \theta) Z(t'_j)
 \end{aligned} \tag{1.19}$$

which for uniformly sampled data reduces to the imaginary part of a weighted discrete Fourier transform of the complex data. The function  $C(f)$  is defined as

$$C(f) \equiv \sum_{i=1}^{N_R} \cos^2(2\pi f t_i - \theta) Z(t_i)^2 + \sum_{j=1}^{N_I} \sin^2(2\pi f t'_j - \theta) Z(t'_j)^2 \tag{1.20}$$

and is an effective number of data items in the real data, see Bretthorst 2000 for more on this. Similarly the function  $S(f)$  is defined as

$$S(f) \equiv \sum_{i=1}^{N_R} \sin^2(2\pi f t_i - \theta) Z(t_i)^2 + \sum_{j=1}^{N_I} \cos^2(2\pi f t'_j - \theta) Z(t'_j)^2 \tag{1.21}$$

and is the effective number of data items in the imaginary data. Finally, the condition that the cross terms cancel, i.e., that the model functions are orthogonal, is used to determine the value of  $\theta$ . This condition is given by:

$$\begin{aligned}
 0 &= \sum_{i=1}^{N_R} \cos(2\pi f t_i - \theta) \sin(2\pi f t_i - \theta) Z(t_i)^2 \\
 &\quad - \sum_{j=1}^{N_I} \sin(2\pi f t'_j - \theta) \cos(2\pi f t'_j - \theta) Z(t'_j)^2.
 \end{aligned} \tag{1.22}$$

Note that if the data are simultaneously sampled,  $t_i = t'_i$ , Eq. (1.22) is automatically satisfied, so  $\theta$  may be defined to be zero. Otherwise,  $\theta$  is

given by

$$\theta = \frac{1}{2} \tan^{-1} \left[ \frac{\sum_{i=1}^{N_R} \sin(4\pi f t_i) Z(t_i)^2 - \sum_{j=1}^{N_I} \sin(4\pi f t'_j) Z(t'_j)^2}{\sum_{i=1}^{N_R} \cos(4\pi f t_i) Z(t_i)^2 - \sum_{j=1}^{N_I} \cos(4\pi f t'_j) Z(t'_j)^2} \right]. \quad (1.23)$$

The triple integral in Eq. (1.15) may be evaluated as follows: First, the integrals over the two amplitudes are uncoupled Gaussian quadrature integrals and are easily done. One needs only complete the square in the exponent, and a simple change of variables to evaluate them. The remaining integral over the standard deviation of the noise prior probability may be transformed into a Gamma integral and is also easily evaluated. We do not give the details of these evaluations; rather we simply give the results:

$$P(f|DI) \propto \frac{1}{\sqrt{C(f)S(f)}} \left[ Nd^2 - \bar{h}^2 \right]^{\frac{2-N}{2}} \quad (1.24)$$

where the sufficient statistic  $\bar{h}^2$  is given by

$$\bar{h}^2 = \frac{R(f)^2}{C(f)} + \frac{I(f)^2}{S(f)} \quad (1.25)$$

and is a generalization of the Lomb-Scargle periodogram.

The generalized Lomb-Scargle periodogram, Eq. (1.25), has a number of very interesting features. First, when the data are real and the sinusoid is stationary, the sufficient statistic for single frequency estimation is the Lomb-Scargle periodogram; not the Schuster periodogram, i.e., not the power spectrum. Second, when the data are real, but  $Z(t)$  is not constant, then Eq. (1.25) generalizes the Lomb-Scargle periodogram in a very straightforward manner to account for the amplitude modulation of the signal. Third, for uniformly sampled quadrature data when the sinusoid is stationary, Eq. (1.25) reduces to a Schuster periodogram or the power spectrum of the data. So while the Schuster periodogram is not a sufficient statistic for frequency estimation in real nonquadrature data, it is a sufficient statistic for quadrature data. Fourth, for uniformly sampled quadrature data when the sinusoid is not stationary, Eq. (1.25) reduced to a weighted power spectrum of the data. Thus the weighted power spectrum is the sufficient statistic for single frequency estimation when the data are quadrature. Fifth, when the quadrature data are nonuniformly but simultaneously sampled, Eq. (1.25) generalizes the weighted power spectrum to account for the nonuniform samples, but otherwise is the exact analogue of a weighted power spectrum. Finally, when the data are nonuniformly and nonsimultaneously sampled, Eq. (1.25) generalizes to a functional form that is formally identical to a Lomb-Scargle periodogram but adapted to an amplitude modulated quadrature detected sinusoid.

### 3 Aliasing

Now that we have finished deriving the generalized Lomb-Scargle periodogram, we would like to investigate some of its properties, in particular the aliasing phenomenon. First, it is easy to show that the parameter  $\theta$  appearing in the generalized Lomb-Scargle model does not change the location of the peak in the Lomb-Scargle periodogram; fixing  $\theta$  only changes the estimated phase of the sinusoid. Consequently, fixing  $\theta$  simplifies the functional form of the sufficient statistic to formally resemble a power spectrum and indeed the generalized Lomb-Scargle periodogram reduces to a power spectrum for simultaneously sampled quadrature data. Now a power spectrum, i.e., the discrete Fourier transform, is a periodic function of frequency. The period is called the bandwidth, and the bandwidth is the largest frequency interval free of repeats or aliases. Because the generalized Lomb-Scargle periodogram reduces to a power spectrum under appropriate conditions, the bandwidth of the generalized Lomb-Scargle periodogram is exactly the same as the bandwidth of the discrete Fourier transform. The question we would like to investigate in this section, is what happens to these repeats or aliases when the data are nonuniformly nonsimultaneously sampled? Are the aliases still there? If not, where did they go?

First, the discrete Fourier transform may be defined as

$$\mathcal{F}(f_k) \equiv \sum_{j=0}^{N-1} \mathbf{d}(t_j) \exp \{2\pi i f_k t_j\} \quad (1.26)$$

where the complex data  $\mathbf{d}$  is given by  $\mathbf{d} \equiv d_R(t_i) + i d_I(t_i)$ . For uniformly sample data the times are given by

$$t_j = j\Delta T \quad (1.27)$$

and if the fast discrete Fourier transform is used to perform this calculation, the frequencies  $f_k$  are given by

$$f_k = \frac{k}{N\Delta T} \quad k \in \left\{ -\frac{N}{2}, -\frac{N}{2} + 1, \dots, \frac{N}{2} \right\}. \quad (1.28)$$

The time  $\Delta T$  is the time interval between data samples and can be used to define the Nyquist critical frequency,

$$f_{N_c} = \pm \frac{1}{2\Delta T}. \quad (1.29)$$

The Nyquist critical frequency may be used to define the bandwidth:

$$\text{bandwidth} \equiv (-f_{N_c} \leq f \leq f_{N_c}). \quad (1.30)$$

It is the largest frequency interval over which the discrete Fourier transform is not a periodic function of frequency.



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 outside the bandwidth:

$$f_k = \frac{k}{N\Delta T}, \quad k = mN + k', \quad k' \in \left\{ -\frac{N}{2}, -\frac{N}{2} + 1, \dots, \frac{N}{2} \right\}. \quad (1.31)$$

The index  $k'$  specifies the nonaliased frequency interval of a discrete Fourier transform. 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. (1.31, and 1.27) into Eq. (1.26), the reason the discrete Fourier transform is periodic becomes readily apparent

$$\mathcal{F}(f_{k'}) \equiv \sum_{j=0}^{N-1} \mathbf{d}(t_j) \exp \left\{ \frac{2\pi \mathbf{i}(mN + k')j}{N} \right\}, \quad (1.32)$$

$$= \sum_{j=0}^{N-1} \mathbf{d}(t_j) \exp \{2\pi \mathbf{i}mj\} \exp \left\{ \frac{2\pi \mathbf{i}k'j}{N} \right\}, \quad (1.33)$$

$$= \sum_{j=0}^{N-1} \mathbf{d}(t_j) \exp \left\{ \frac{2\pi \mathbf{i}k'j}{N} \right\}, \quad (1.34)$$

$$= \sum_{j=0}^{N-1} \mathbf{d}(t_j) \exp \{2\pi \mathbf{i}f_{k'}t_j\}. \quad (1.35)$$

In going from Eq. (1.33) to (1.34) a factor,  $\exp\{\mathbf{i}(2\pi mj)\}$ , was dropped because both  $m$  and  $j$  are integers, so  $(2\pi mj)$  is an integer multiple of  $2\pi$ , 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.

In the present problem, nonuniformly nonsimultaneously sampled data, there is no  $\Delta T$  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. Consequently, there must be a largest effective dwell time,  $\Delta T'$ , such that all of the times (both the real and imaginary) must satisfy

$$t_l = k_l \Delta T' \quad t_l \in \{\text{Real } t_i \text{ or Imaginary } t'_j\} \quad (1.36)$$

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 least common denominator for all of the times. Additionally, the effective dwell time is the dwell time at which one would have had to acquire data 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

$$f_{Nc} = \frac{1}{2\Delta T'}. \tag{1.37}$$

Aliases *must* appear for frequencies outside this bandwidth.

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. (1.32)-(1.35) the Nyquist critical frequency of this data set is  $1/2\Delta T'$  and frequencies outside the bandwidth are aliased. Now look at the definitions of  $R(f)$  and  $I(f)$ , Eqs. (1.18) and (1.19). 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 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. (1.18) and (1.19) are computing. So aliases must appear at frequencies greater than this Nyquist critical frequency. For much more on aliases see Bretthorst 2000.

## 4 Parameter Estimates

The generalize Lomb-Scargle periodogram is a sufficient statistic for the estimation of a frequency in nonuniformly nonsimultaneously sample data. However, the frequency is not the only parameter appearing in the model; the model also implicitly contains an amplitude, phase and possible one or more parameters associated with amplitude modulation of the signal. In this section we would like to investigate what happens to the parameter when the data are nonuniformly nonsimultaneously sampled. In particular we would like to know if the parameter estimates change when the data are nonuniformly nonsimultaneously sampled.

In this discussion we are going to estimate the parameters using the data shown in Fig. 1(A) and (B). These two data sets contain exactly the same

FIGURE 1. Uniformly and Nonuniformly Sampling

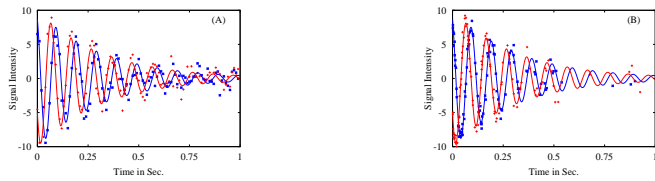


Fig. 1. Panel (A) and (B) are simulated data, each data set has exactly the same signal having exactly the same signal-to-noise. The data sets differ only because panel (A) has been uniformly sampled, while (B) has been nonuniformly sampled. Note the nonuniform samples were taken exponentially, thus there are more samples at the beginning of the data and exponentially fewer at the end of the data.

signal and have exactly the same signal-to-noise, they differ from each other only in that panel (A) has been uniformly sampled while panel (B) has been randomly sampled. These random samples are distributed exponentially. We mention this only because it will become important later when we consider amplitude estimation. The noise realizations in each data set are different, and this will result in slightly different parameter estimates for each data set.

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 prior probability as these are of less importance. The model we will use is given by

$$d_R(t_i) = A \cos(2\pi f t_i + \phi) \exp\{-\alpha t_i\} \quad (1.38)$$

for the real channel. This model is of the general form of the Lomb-Scargle model, but now we have suppressed the extra phase parameter, as its redundant, we have added an exponential decay rate constant to describe the amplitude modulation, and we have written the model in terms of an amplitude and phase rather than sine and cosine amplitudes.

Markov chain Monte Carlo was used to compute the marginal posterior probability for each parameter. All of the parameters appearing in the model were simulated simultaneously, thus the target distribution of Markov chain Monte Carlo simulation was the joint posterior probability for all the parameters. We targeted the joint posterior probability for all of the parameters 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 Fig. 2(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 Neal 1993 and Gilks, et. al. 1996.

The posterior probability for the frequency, decay rate constant, and amplitude are shown in Fig. 2(A), (B) and (D) 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 curves drawn with open characters were computed using the uniformly sampled data shown in Fig. 1(A); while the solid lines in these plots were computed from the nonuniformly nonsimultaneously sampled data shown in Fig. 1(B).

The marginal posterior probability for the frequency is shown in Fig. 2(A). This is the fully normalized marginal posterior probability for the frequency independent of all of the other parameters, Eq. (1.24). 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. 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. Thus, sampling 10 times longer will result in frequency estimates that are 10 times more precise. As noted in Bretthorst 1988 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.

We have also plotted the absolute-value spectra computed from these two data sets, Fig. 2(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 note 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 single largest peak in the spectrum*. In the discrete Fourier transform, the presence of multiple peaks may or may not be an indication of multi-

ple resonances. Indeed it is easy to show that the generalized Lomb-Scargle periodogram may have peaks that are related to the sampling scheme. The only way to be certain that multiple resonances are present, is to postulate a model containing multiple resonances and then compute the posterior probability for the number of resonances.

The marginal posterior probability for the decay rate constant is shown in Fig. 2(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  $3 \text{ Sec.}^{-1}$ , 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  $2 \text{ Sec.}^{-1}$  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. 2(D). In this paper we did not directly talk about amplitude estimation (see Bretthorst 1992 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 Eq. (1.38) 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. 2(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 reasons. 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 uniformly randomly sampled data. Note that the height and width of the posterior probabilities computed from both the uniformly and the uniformly-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. 2(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.

## 5 Summary and Conclusions

Probability theory generalizes the Lomb-Scargle periodogram roughly as follows: in uniformly or nonuniformly sampled real data, the sufficient statistic for estimating the frequency of a single stationary sinusoid is the Lomb-Scargle periodogram. When the function  $Z(t_i)$  is not a constant, probability theory generalized the Lomb-Scargle periodogram to include this modulation. For a stationary sinusoid, when the data are quadrature simultaneously sampled, probability theory simplifies the Lomb-Scargle periodogram to a Schuster periodogram. When the sinusoid is not stationary, the sufficient statistic becomes a weighted power spectrum where the weighting function is given by  $Z(t)$ . Finally, when the data are nonuniformly nonsimultaneously sampled, the sufficient statistic is the generalized Lomb-Scargle periodogram.

In a literal sense, probability theory does no such thing as generalize the discrete Fourier transform or the Lomb-Scargle periodogram. Probability theory simply tells one how to analyze a particular problem optimally. For estimation of a sinusoidal 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 trend, 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 trend in the data, see Bretthorst 1988, and 2000.

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 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.

## 6 References

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.

Bretthorst, G. Larry, (2000) "Nonuniform Sampling: Aliasing and Bandwidth," *Maximum Entropy and Bayesian Methods*, G. Erickson *ed.*, Kluwer academic press, the Netherlands.

Gilks, W. R., S. Richardson and D. J. Spiegelhalter (1996), "Markov Chain Monte Carlo in Practice," Chapman & Hall, London.

Lomb, N. R. (1976) "Least-Squares Frequency Analysis of Unevenly Spaced Data," *Astrophysical and Space Science*, **39**, pp. 447-462.

Marple, S. L. (1987) *Digital spectral Analysis with applications*, Prentice-



Hall, Inc., Englewood Clifts. New Jersey.

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.

Priestley, M. B. (1981), *Spectral Analysis and Time Series*, 2 Vols., Academic Press, Inc., Orlando FL, Combined paperback edition with corrections (1983).

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.

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.

Schuster, A., (1905), "The Periodogram and its Optical Analogy," *Proceedings of the Royal Society of London*, **77**, p. 136.

FIGURE 2. Estimating The Parameters

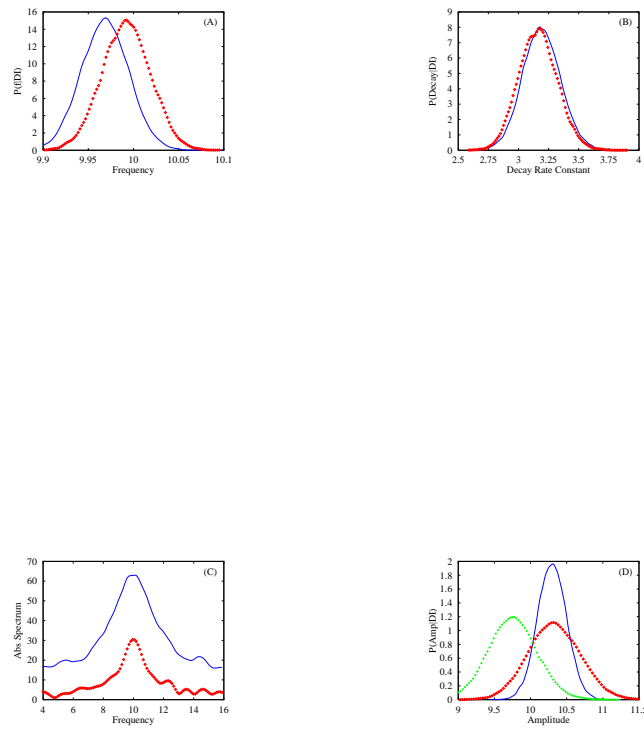


Fig. 2. The posterior probability of the parameters was computed for the uniformly and nonuniformly nonsimultaneously sampled data, open characters and solid lines respectively. Panel (A) is the posterior probability for the frequency, (B) the decay rate constant, (D) the amplitude. Panel (C) is the absolute-value spectrum computed for the two data sets. The extra curve in panel (D), the plus signs, was computed from a nonuniformly nonsimultaneously sample data set having uniformly sampled times, see text for details.