

NONPARAMETRIC ESTIMATION OF A PERIODIC FUNCTION

Peter Hall¹, James Reimann², and John Rice^{1,3}

ABSTRACT. Motivated by applications to brightness data on periodic variable stars, we study nonparametric methods for estimating both the period and the amplitude function from noisy observations of a periodic function made at irregularly spaced times. It is shown that nonparametric estimators of period converge at parametric rates and attain a semiparametric lower bound which is the same if the shape of the periodic function is unknown as if it were known. Also, first-order properties of nonparametric estimators of the amplitude function are identical to those that would obtain if the period were known. Numerical simulations and applications to real data show the method to work well in practice.

KEY WORDS AND PHRASES. frequency estimation, nonparametric regression, semiparametric estimation, Nadaraya-Watson estimator, MACHO project, variable star data.

SHORT TITLE. Estimation of a periodic function

¹ Centre for Mathematics and its Applications, Australian National University, Canberra, ACT 0200, Australia

² Biostatistics Department, Genentech, Inc., South San Francisco, CA 94080, USA

³ Department of Statistics, Evans Hall, University of California, Berkeley, CA 94720, USA

1. INTRODUCTION

Periodic variable stars emit light whose intensity, or brightness, changes over time in a smooth and periodic manner. Both the period and the time-varying function that represents brightness are of direct scientific interest, for example as an aid to classifying stars into different categories depending on magnitude, period, and light curve shape, for making inferences about stellar evolution, and for determining astronomical distances (Hoffmeister *et al.*, 1985). It is clear that if observations are made at regularly spaced points in time then identification of the brightness function without a structural model may not be possible. This is one aspect of a range of aliasing problems that arise for observation times that lie on a grid. However, in practice the times of observation are somewhat irregular—a star might be observed on most nights but not at the same time each night.

Our work was motivated by the analysis of light curves of variable stars collected by the MACHO project (Axelrod *et al.*, 1994). The main goal of this project is to search for dark matter in the halo of our galaxy by monitoring the intensity of millions of stars. A valuable byproduct of this search was the identification of more than 40,000 variable stars in the Large Magellanic Cloud, whose periods and light curves were found by methods such as we discuss, based on (Reimann, 1994). Large inventories of variable star light curves are also being produced by other similar surveys (Ferlet *et al.*, 1997). In addition to purely periodic functions, the light curves of variable stars can exhibit non-periodic behaviour, including slowly drifting periods and simultaneous pulsation at two incommensurate periods. Here we focus on the strictly periodic case.

In the present note we analyse a nonparametric approach to estimating both the period and the corresponding periodic function, exploiting the properties of ‘staggered’ observation times. We present a semiparametric lower bound for the accuracy with which the period can be determined, and show that this bound can be asymptotically attained by an estimate based on nonparametric regression. Interestingly, this accuracy is the same as that which could be attained if the shape, but not the period or phase, of the curve were known. We show that the period can be estimated at rates that are more commonly found in completely parametric problems. Indeed, the convergence rate of the estimator of period is $O(n^{-3/2})$, where n denotes the number of observations made. This is the rate obtained by

Quinn and Thompson (1991), for example, who treated a related problem where observation times were regularly spaced and the periodic function was modelled as a finite-parameter trigonometric series.

Since the period can be estimated so accurately, if we construct a nonparametric, smoothed estimator of the brightness function using our estimate of the period, then the resulting nonparametric curve estimator has the same first-order asymptotic properties that it would enjoy if the period were known.

Thus, very little is lost by adopting a nonparametric approach to this problem. And of course there is much to gain, in the absence of physical evidence for an elementary model for the way in which brightness varies with time. Relatively conventional methods, such as local linear smoothing and kernel techniques, may be used to estimate the brightness function itself. The relative susceptibility of some curve estimators to edge effects is not an issue here, since periodicity will be exploited to overcome that difficulty. These properties will be demonstrated both theoretically and numerically in subsequent sections, and applications to star data will be illustrated.

Trigonometric models for regression means have a long history, including for example the contributions of Fisher (1929), Whittle (1959), Walker (1971), Hannan (1973), Brillinger (1986), Rice and Rosenblatt (1988), and Reimann (1999). The case of trigonometric models where all frequencies are harmonics of a single, fundamental frequency is closer to the context of the present paper, and has been treated by Hannan (1974) and Quinn and Thompson (1991) in the setting of regularly spaced observations. Smoothing methods in such problems have been discussed by McDonald (1986), and Reimann (1994). See also Bickel, Klaassen, Ritov and Wellner (1993, p. 107).

2. MODELS AND METHODOLOGY

Let g denote a periodic function, with period θ_0 say, and suppose we make successive observations (X_i, Y_i) , $1 \leq i \leq n$, generated according to the model

$$Y_i = g(X_i) + \epsilon_i, \tag{2.1}$$

where $0 < X_1 \leq \dots \leq X_n$ and, conditional on the X_i 's, the ϵ_i 's are independent and identically distributed random variables with zero mean and finite variance.

From these data we wish to estimate θ_0 and g , making only periodic-smoothness assumptions about the latter.

To this end we construct a nonparametric estimator $\hat{g}(\cdot|\theta)$ of g on $(0, \theta]$ under the assumption that the period of g is θ . We extend \hat{g} to the real line by periodicity, and, defining

$$S(\theta) = \sum_{i=1}^n \{Y_i - \hat{g}(X_i|\theta)\}^2, \quad (2.2)$$

take our estimator $\hat{\theta}$ of θ_0 to be the minimiser of $S(\theta)$. This procedure was proposed by McDonald (1986). Then, for an appropriate estimator $\tilde{g}(\cdot|\theta)$ of g under the assumption of period θ , we take $\hat{g}_0 \equiv \tilde{g}(\cdot|\hat{\theta})$ to be our estimator of g .

Even if the estimators \hat{g} and \tilde{g} are of the same type, for example both local linear estimators, it is usually not a good idea to take them to be identical. In particular, to ensure optimal estimation of θ the estimator $\hat{g}(\cdot|\theta)$ employed to define $S(\theta)$ at (2.2) should be smoothed less than is appropriate for point estimation of g .

In general the function S will have multiple local minima, not least because any integer multiple of θ_0 can be considered to be a period of g . We take θ_0 to be the smallest period of g , assumed to be well-defined. In particular, we do not permit g to be identically constant, and we minimise S in a sufficiently small neighbourhood of θ_0 . These specifications do not eliminate operational difficulties that can be experienced with local minima of S , but they do remove technical problems associated with making $\hat{\theta}$ well defined.

A range of generalisations is possible for the model at (2.1). For example, we may replace the errors ϵ_i by $\sigma_i\epsilon_i$, where the standard deviation σ_i is either known, as is often effectively the case with data on star brightness, or accurately estimable. Then, appropriate weights should be incorporated into the series at (2.2). Furthermore, our main results about convergence rates and appropriate choice of bandwidth continue to hold when the errors form a sufficiently weakly dependent stationary stochastic process. When constructing $\hat{g}(\cdot|\theta)$ for the purpose of computing $\hat{g}(X_i|\theta)$ at (2.2) we may drop the pair (X_i, Y_i) from the data, without influencing first-order asymptotic properties of our method.

Realistic mathematical models for the spacings between successive X_j 's would involve them not converging to zero as $n \rightarrow \infty$. Moreover, they should be approxi-

mately stochastically independent. One such model is

$$X_j = \sum_{i=1}^j V_i, \quad 1 \leq j \leq n, \quad (2.3)$$

where V_1, V_2, \dots are independent identically distributed nonnegative random variables. Clearly there are limitations, however, to the generality of the distribution allowable for V , representing a generic V_i . In particular, if the distribution is defined on an integer lattice, and if θ_0 is a rational number, then the infinite sequence of numbers X_j modulo θ_0 , for $1 \leq j < \infty$, is restricted to a lattice with rational span in the interval $[0, \theta_0)$. This means that, even if the errors ϵ_i are identically zero and we actually observe the infinite sequence $g(X_j)$, $1 \leq j < \infty$, g is not identifiable under just smoothness conditions. These difficulties vanish if we assume that the X_j 's are generated by (2.3) where the distribution of $V > 0$ is absolutely continuous with an integrable characteristic function and that all moments of V are finite. Call this model $(M_{X,1})$. The fact that the characteristic function should be integrable excludes the case where the X_i 's are points of a homogeneous Poisson process, but this context is readily treated separately.

Another class of processes \mathcal{X} is the sequence $X_j = X_j(n) \equiv nY_{nj}$, where $Y_{n1} < \dots < Y_{nn}$ are the order statistics of a random sample Y_1, \dots, Y_n from a Uniform distribution on the interval $[0, y]$, say. Call this model $(M_{X,2})$. Models $(M_{X,1})$ and $(M_{X,2})$ are similar, particularly if V has an exponential distribution. There, if $\mathcal{X}(n+1) = \{X_1, \dots, X_{n+1}\}$ is a sequence of observations generated under $(M_{X,1})$, if $\mathcal{X}'(n) = \{X'_1, \dots, X'_n\}$ is generated under $(M_{X,2})$ with $y = 1$, and if we define $X_{\text{tot}} = \sum_{i \leq n+1} X_i$, then $\{X_1/X_{\text{tot}}, \dots, X_n/X_{\text{tot}}\}$ has the same distribution as $\mathcal{X}'(n)$.

A third class of processes \mathcal{X} is the 'jittered' grid of Akaike (1960), Beutler (1970) and Reimann (1994), where $X_j = j + U_j$, $j \geq 1$, and the variables U_j are independent and Uniformly distributed on $(-\frac{1}{2}, \frac{1}{2})$. Call this model $(M_{X,3})$. All three models have the property that the spacings $X_j - X_{j-1}$ are identically distributed and weakly dependent.

Next we discuss candidates for \hat{g} . Under the assumption that the true period of g is θ , the design points X_i may be interpreted modulo θ as $X_i(\theta) = X_i - \theta[X_i/\theta]$, for $1 \leq i \leq n$, where $[x]$ denotes the largest integer strictly less than x . Then, the design points of the set of data pairs $\mathcal{Y}(\theta) \equiv \{(X_i(\theta), Y_i), 1 \leq i \leq n\}$, the

‘phased’ or ‘folded’ data, all lie in the interval $(0, \theta]$. We suggest repeating *ad infinitum* the scatterplot represented by $\mathcal{Y}(\theta)$, so that the design points lie in each interval $((j-1)\theta, j\theta]$ for $-\infty < j < \infty$; and computing $\hat{g}(\cdot|\theta)$, restricted to $(0, \theta]$, from the data, using a nonparametric smoother. Possible smoothers include kernel estimators, local linear estimators, regression and smoothing splines, and truncated trigonometric series expansions. In practice we would usually need to repeat the design only in each of $(-\theta, 0]$, $(0, \theta]$ and $(\theta, 2\theta]$, since the effective bandwidth would be less than θ . We define $\hat{g}(\cdot|\theta)$ on the real line by $\hat{g}(x) = \hat{g}(x - \theta|x/\theta|\theta)$.

In view of the periodicity of g it is not necessary to use a function estimation method, such as a local linear smoother, which accommodates boundary effects. Indeed, our decision to repeat design points in blocks of width θ means that we do not rely on the boundary-respecting properties of such techniques. To stress this feature of our method, our theoretical analysis in Section 4 will focus on the case where \hat{g} is the Nadaraya-Watson estimator, which suffers notoriously from boundary problems:

$$\hat{g}(x|\theta) = \frac{\sum_i Y_i K_i(x|\theta)}{\sum_i K_i(x|\theta)}, \quad 0 \leq x \leq \theta, \quad (2.4)$$

where $K_i(x|\theta) = K[\{x - X_i(\theta)\}/h]$, K is a kernel, h is a bandwidth, and the two series on the right-hand side of (2.4) are computed using repeated blocks of the data $\mathcal{Y}(\theta)$.

If g has r bounded derivatives; if the estimator \hat{g} is of r 'th order, meaning that its asymptotic bias is of size h^r and its variance of size $(nh)^{-1}$; and if $h = h(n)$ has the property that for some $\eta > 0$, $n^{-(1/2)+\eta} \leq h = o(n^{-1/(2r)})$; then $\hat{\theta} = \operatorname{argmin} S(\theta)$ is consistent for θ_0 and, under regularity conditions,

$$n^{3/2} (\hat{\theta} - \theta_0) \rightarrow N(0, \tau^2) \quad (2.5)$$

in distribution, where

$$\tau^2 = 12 \sigma^2 \theta_0^3 \mu^{-2} \left\{ \int_0^{\theta_0} g'(u)^2 du \right\}^{-1}, \quad (2.6)$$

$\sigma^2 = \operatorname{var}(\epsilon_i)$ and $\mu = \lim_{j \rightarrow \infty} E(X_j - X_{j-1})$, assumed to be finite and nonzero. In section 4 we formally establish this result for the second order Nadaraya-Watson estimate; it can be shown to hold for other kernel methods, such as Gasser-Müller, higher order kernels, and local linear estimators. We believe that it holds generically for reasonable smoothers, such as regression and smoothing splines and orthogonal

series estimators. Formula (2.5) implies that $\hat{\theta}$ converges to θ_0 at a parametric rate. The asymptotic variance is that given by the semiparametric lower bound presented in Section 4. In Quinn and Thompson's (1991) analysis of a periodogram estimate in a closely related problem with equispaced observations, they obtained the same limit theorem for $\hat{\theta}$, albeit with a different value of τ^2 . A similar result for jittered data with the periodic function being a sinusoid was obtained by Reimann (1994).

Formula (2.6) implies that estimators of period have lower variance when the function g is 'less flat', i.e. when g has larger mean-square average derivative. This of course conforms to intuition, since a flat function g does not have well-defined period, and more generally, the flatter g is, the more difficult it is to visually determine its period.

If $h \sim Cn^{-1/(2r)}$ for a constant $C > 0$, and \hat{g} is an r 'th order regression estimator, then $n^{3/2}(\hat{\theta} - \theta_0)$ remains asymptotically Normally distributed but its asymptotic bias is no longer zero. In the r 'th order case, $h = O(n^{-1/(2r)})$ is the largest order of bandwidth that is consistent with the 'parametric' convergence rate, $\hat{\theta} = \theta_0 + O_p(n^{-3/2})$. Since such h is optimal for estimation of g given θ_0 , the phased data should be undersmoothed for estimation of the period relative to the amount of smoothing optimal for estimating the underlying light curve itself.

This high degree of accuracy for estimating θ_0 means that, if $\tilde{g}(\cdot|\theta)$ is a conventional estimator of g under the assumption that the period equals θ , then first-order asymptotic properties of $\hat{g}_0 \equiv \tilde{g}(\cdot|\hat{\theta})$ are identical to those of $\tilde{g}(\cdot|\theta_0)$. That is, from an asymptotic viewpoint the final estimator \hat{g}_0 behaves as though the true period were known. These results follow by Taylor expansion. For example, if $\tilde{g}(\cdot|\theta)$ is the Nadaraya-Watson estimator defined at (2.4), but with a different bandwidth h_0 say, satisfying $h_0 \geq n^{-(1/2)+\xi}$ for some $\xi > 0$, then a Taylor expansion argument such as that given in the appendix shows that for all $\eta > 0$,

$$\begin{aligned} \tilde{g}(\cdot|\hat{\theta}) &= \tilde{g}(\cdot|\theta_0) + O_p \left[n |\hat{\theta} - \theta_0| h_0^{-1} \{h_0 + (nh_0)^{-1/2}\} n^\eta \right] \\ &= \tilde{g}(\cdot|\theta_0) + o_p \{ (nh_0)^{-1/2} \}. \end{aligned} \tag{2.7}$$

The remainder $o_p \{ (nh_0)^{-1/2} \}$ here is of smaller order than the error of $\tilde{g}(\cdot|\theta_0)$ about its mean. And, despite the Nadaraya-Watson estimator $\tilde{g}(\cdot|\theta)$ being susceptible to edge effects, the order of approximation at (2.7) is available uniformly in $x \in (0, \theta_0]$, since the effects of errors at boundaries are of order $n |\hat{\theta} - \theta_0| = o_p \{ (nh_0)^{-1/2} \}$.

3. NUMERICAL EXAMPLES

The MACHO project measures light intensities in two frequency bands, a ‘red’ and a ‘blue’ band. Figure 1 shows the measurements versus time in each of the two bands for an eclipsing binary, i.e. two stars orbiting each other in the plane of the observer. There are 300 measurements in the red band and 305 in the blue band over a period of about 400 days. These plots are rather uninformative, but when the times are transformed modulo a period of 2.4713 days, the periodicity is apparent, as shown in Figure 2. The curve shape is typical of an eclipsing binary system; the brightness is fairly constant when the stars are side by side and troughs occur when one star passes in front of the other. Corresponding to the very fast convergence rate of the estimate of period, when the period is changed very slightly, the dispersion of the phased plot increases dramatically.

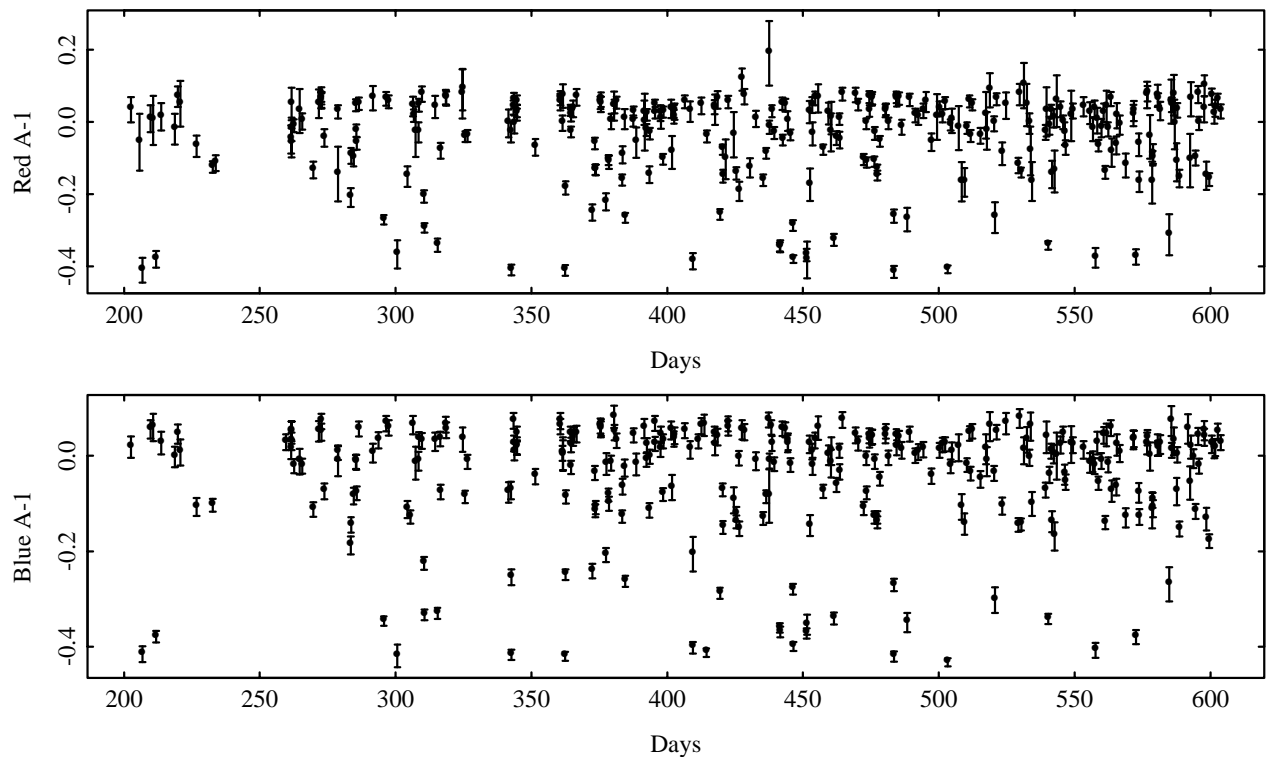


Figure 1: Brightness plotted against time for the red and blue bands of star 77043:4317.

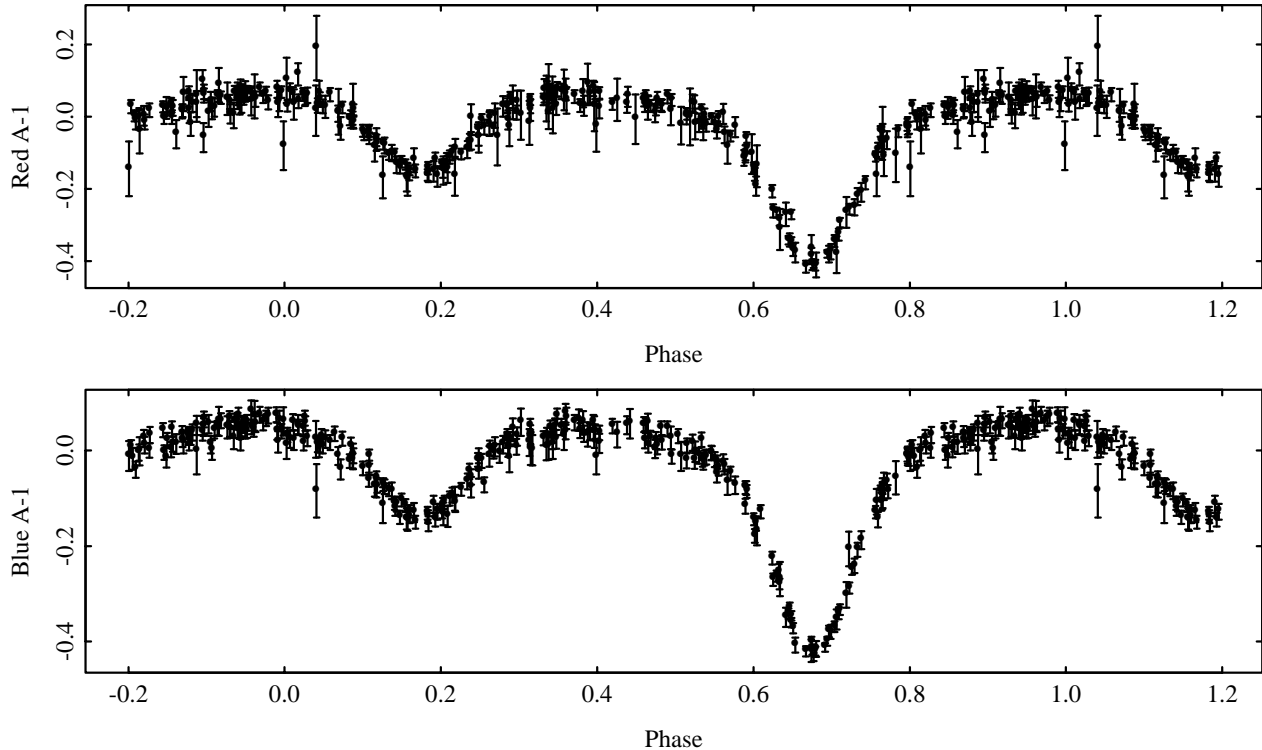


Figure 2: Phase plot of star 77043:4317 at period 2.47133 days.

The residual sum of squares (2.2) used to estimate the period is typically highly oscillatory. Furthermore, for some of the methods, the residual sum of squares is not a continuous function of the period—discontinuities can occur when the ordering of two phase values changes. Thus for the MACHO data, the minimiser was found in two stages: a relatively crude initial grid search selected regions that were then searched on a finer grid. Reimann (1994) provides a more complete discussion.

We now briefly describe the results of one simulation. An asymmetric curve derived from data from a Cepheid was used for the true underlying function, with the maximum and minimum scaled to be 0.5 and -0.5. Actual observation times, 200 in all spanning 243 days, were taken from the MACHO database. Two levels of noise were used: independent and identically distributed Normals with standard deviations .05 and .20. Typical realisations of the phased light curves with period 1.4432 days are shown in Figure 3.

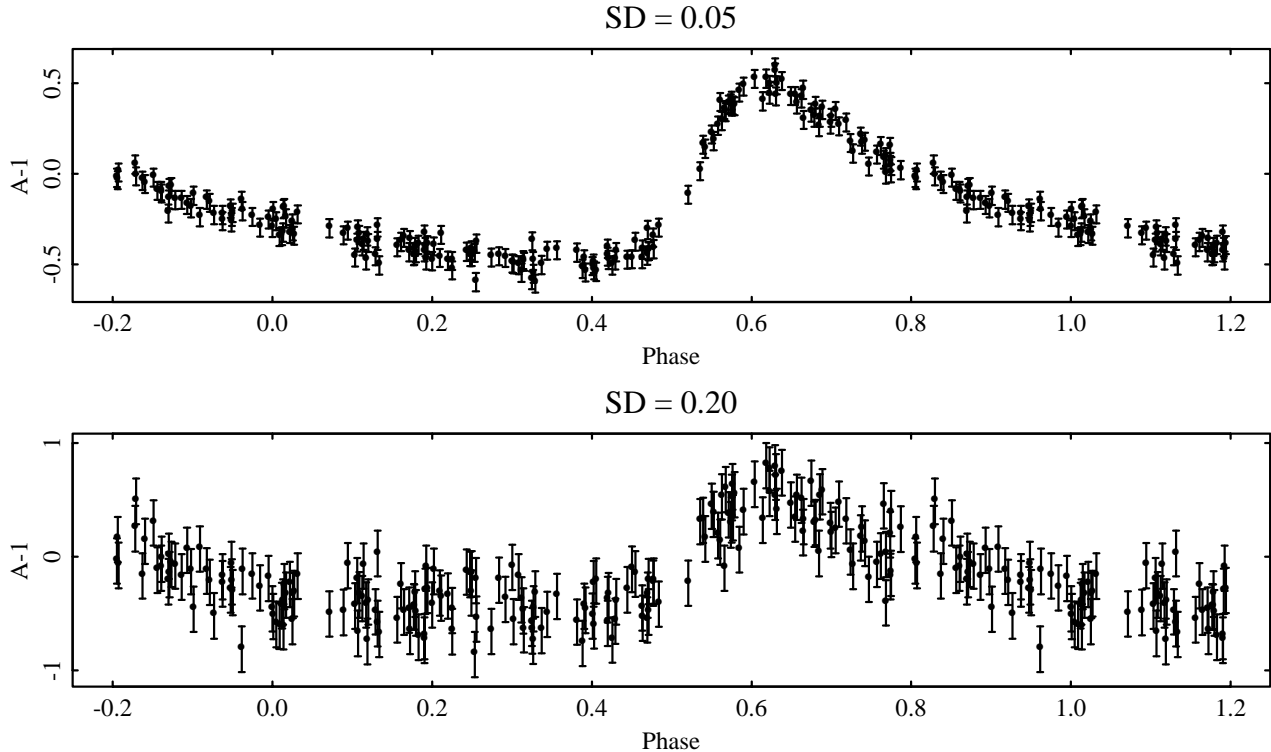


Figure 3: Phase plots of simulated light curves from model A at period 1.4432 days.

Estimates of period were computed for a variety of methods for 100 simulations. The following nonparametric regression estimates considered for use in MACHO were used in (2.2): trigonometric expansions of order one, two, four and six, cubic periodic regression splines with five, nine, and thirteen knots, and SuperSmoother, a cross-validated variable span local linear smoother (Friedman, 1984) for which the sum of absolute residuals rather than sum of squared residuals at (2.2) was minimised. Note that these techniques would be particularly difficult to treat in the manner we have done for kernel techniques. We also considered the maximiser of the periodogram

$$I(\omega) = \frac{1}{n} \left| \sum_{j=1}^n Y_j e^{i\omega X_j} \right|^2.$$

In the astronomy literature, this is referred to as the Deeming periodogram; its properties in estimating frequency are discussed in Deeming (1975). A version of the periodogram which is equivalent to least squares with a single sinusoid was introduced by Lomb (1976) and Scargle (1982), and is referred to as the Lomb-Scargle periodogram. We also examine the behaviour of two other methods proposed in the astronomy literature. The method of Lafler and Kinman (1965), in the case of equal

error variance, chooses the period to minimise the quantity

$$LF(\theta) = \sum_{j=1}^n \{Y_{j+1}^*(\theta) - Y_j^*(\theta)\}^2.$$

where the $Y_j^*(\theta)$ are the values of the $Y_i(\theta)$ ordered as are the $X_i(\theta)$. The related method of Renson (1978) chooses the period to minimise

$$REN(\theta) = \sum_{j=1}^n \frac{\{Y_{j+1}^*(\theta) - Y_j^*(\theta)\}^2}{\{X^{(j+1)}(\theta) - X^{(j)}(\theta)\}^2 + b^2}$$

where the $X^{(j)}(\theta)$ are the ordered values of the $X_i(\theta)$. Here b^2 is a constant chosen so as to prevent the denominator from being too small; Renson recommends taking it to be the ratio of the noise variance to four times the squared amplitude of the signal. This latter method gives greater weight to differences which are close in phase. All the estimates were evaluated on a grid which was much finer than the standard error provided by the semiparametric lower bound.

Table 1: Standardized bias, variance, and MSE of the frequency estimates

Method	Bias/SE	SD = 0.05		SD = 0.20		
		Variance $\times 10^{-10}$	MSE $\times 10^{-10}$	Bias/SE	Variance $\times 10^{-10}$	MSE $\times 10^{-10}$
PG	-16.7	6.3	24.1	-3.9	93	107
C1	-8.7	9.3	16.4	-1.7	128	132
C2	6.9	4.4	6.4	0.8	58	59
C4	-12.1	2.9	7.2	-3.9	50	56
C6	-4.8	3.3	4.0	-2.3	68	72
S5	68.5	1.2	59.3	11.3	33	76
S9	-27.7	4.5	39.0	-8.2	51	85
S13	-4.7	6.3	7.6	-2.5	84	89
LF	-0.7	16.5	16.5	0.0	254	254
RN	-0.7	16.5	16.5	0.0	253	254
SM	-1.9	6.3	6.4	-1.2	140	142

The results of the simulation are summarised in Table 1. The lower bounds for the variances of the frequency estimates are 2.5×10^{-10} and 4.1×10^{-9} for the low and high noise cases. In the low noise case, the estimate with smallest MSE

was the six term Fourier fit, whose mean squared error was 1.6 times the variance bound. The five and nine knot spline estimates were subject to surprisingly large biases, especially in light of the fact that they had the same number of degrees of freedom as the respective Fourier fits. For both the Fourier expansion methods and the spline methods, the mean squared error generally decreased as the order increased. For the higher variance case, the Fourier methods had the smallest mean squared error, followed by the spline methods. The smallest MSE, that of the four term Fourier method, was equal to 1.4 times the lower bound. SuperSmoother was not as effective in the high noise case as in the low noise case.

In general, the mean squared errors were larger than those given by our asymptotics and the biases were large relative to the standard errors. This may be in part due to the estimates being determined by global minimisation, whereas the asymptotic analysis is highly localised. The Lafler and Renson methods, although having little bias, were not competitive due to their large variances. Note that they are crudely analogous to kernel estimates with an extremely small bandwidth. At the other end of the continuum, the periodogram method and the estimate based upon fitting a simple cosinusoid were not competitive with the more flexible smoothing methods, and of these the less smoothed estimates performed better. However, in general, all the methods were impressively accurate considering that only 200 observations were made over 243 days—the worst estimate in Table 1 has root mean square error equal to about 16×10^{-5} cycles per day, which translates into a root mean square error in the estimate of period of about 29 seconds. In this and other examples, the variance bound was a reasonable guide to attainable precision and can thus be recommended for aid in choosing a grid size for global optimisation.

4. ASYMPTOTIC PROPERTIES OF $\hat{\theta}$

We first state the semiparametric lower bound and indicate the proof. We assume that the errors are independent and identically distributed, independent of the observation times, and from a density which satisfies mild regularity conditions, which hold in particular if it is Gaussian. Scaled to have period one, the regression curve $s(x)$ belongs to a set the members of which have uniformly bounded second derivatives and integrated squared first derivatives uniformly bounded below. The parameter space for the period is a bounded open interval. The observation times are generated as in model $(M_{X,2})$. Then if the regression function is known except

for phase, or if it is unknown, the information bound for the estimate of the period is given by (2.6), where σ^2 is replaced by the reciprocal of Fisher information, I , on a location parameter in the error density; the scaling factor is $n^{3/2}$ as at (2.5).

The proof for the case of known shape and unknown phase follows by calculating the limiting Fisher information for this finite parameter model and verifying that the conditions for local asymptotic normality hold, using Theorem II.6.1 of Ibragimov and Has'minskii (1981). For the semiparametric case, with unknown shape, we follow the lines of Bickel *et al.* (1993), who derive the information for this periodic regression model with independent and identically distributed sampling times. Their arguments need to be slightly modified for our sampling model, a triangular array. The observation times in this model are $X_j = nU_j$ where the U_j are (unordered) uniform variates. Let s denote the unknown periodic function, normalised to period one, and h denote the uniform density. Then, following Bickel *et al.*, the information in the first n observations for θ is

$$\Psi^2(n, \theta) = nI E[s'(nU_1\theta^{-1})^2 \{nU_1 - E(nU_1 | \theta)\}^2]$$

where, with both summations over $-\infty < k < \infty$,

$$E(nU_1 | \theta) = \frac{n \sum_k (u_1 + n^{-1}k\theta)h(u_1 + n^{-1}k\theta)}{\sum_k h(u_1 + n^{-1}k\theta)}$$

which equals $nE(U_1) + O(1)$. Thus

$$n^{-3} \Psi^2(n, \theta) = I E[s'(nU_1\theta^{-1})^2 \{U_1 - E(U_1)\}^2] + O(n^{-1}),$$

which converges uniformly to $I \text{var}(U_1) \|s'\|_2^2$. If the error density is Gaussian, the information is $I = \sigma^{-2}$, and we obtain the asymptotic variance at (2.6). Local asymptotic normality follows by Proposition II.1.2 of Bickel *et al.*

We next state a limit theorem for $\hat{\theta}$ under specific regularity conditions, and then we discuss generalisations. We assume that $\hat{g}(\cdot|\theta)$ is defined by (2.4), where (a) K is a symmetric, compactly supported density with three continuous derivatives on $(-\infty, \infty)$; (b) $h = h(n)$ satisfies $h \sim Cn^{-a}$, where $C > 0$ and $\frac{1}{4} < a < \frac{1}{3}$; (c) g is periodic with period θ_0 , has two bounded derivatives, and is not identically constant; (d) $E|\epsilon_i|^{2+\delta} < \infty$ for some $\delta > 0$; and (e) distributions of the observation times $X_1 \leq \dots \leq X_n$ satisfy $(M_{X,1})$, $(M_{X,2})$ or $(M_{X,3})$, given in section 2. Let $0 < b < \frac{1}{3} - a$, and let $\hat{\theta}$ denote that value of θ in the interval $|\theta - \theta_0| \leq n^{b-(3/2)}$

which minimises $S(\theta)$. Then, $\hat{\theta}$ is asymptotically Normal $N(0, \tau^2)$, where τ is defined at (2.6). Under condition (e), $E(X_j - X_{j-1}) = \mu + O(j^{-1})$, where $\mu = E(V)$ when $(M_{X,1})$ holds, $\mu = y^{-1}$ under $(M_{X,2})$, and $\mu = 1$ under $(M_{X,3})$.

If instead of condition (a) we assume that K has k continuous derivatives then, using arguments in this appendix, the allowable range of values of a may be increased to $\frac{1}{4} < a < (k-1)/2k$ in which case the value of b should satisfy $0 < b < (k-1)/(2k) - a$. Longer arguments may be used to substantially relax these smoothness conditions on K . Likewise, the assumption that $|\theta - \theta_0| \leq n^{b-(3/2)}$ can be relaxed to $|\theta - \theta_0| \leq \delta$ for $\delta > 0$ sufficiently small. Of course, if k is large then $(k-1)/2k$ is close to $\frac{1}{2}$, and then $h \sim Cn^{-a}$ and $\frac{1}{4} < a < (k-1)/2k$ requires h to be of smaller order than $n^{-1/4}$ and a little larger than $n^{-1/2}$.

If $h \sim Cn^{-1/4}$ then a term in h^2 that appears in Taylor expansions, such as (4.4) in the proof below, contributes a quantity of size $h^2 \sim C^2 n^{-1/2}$ to the bias of $n(\hat{\theta} - \theta_0)$, in which case the Normal limit at (2.5) would have nonzero mean. More generally, if the estimator \hat{g} is of r 'th order, for example based on fitting polynomials of degree $r-1$, and if g is sufficiently smooth, then the bias term is of size h^r , the condition $\frac{1}{4} < a < (k-1)/2k$ changes to $(2r)^{-1} < a < (k-1)/2k$, and a nonzero mean is required in the Normal limit distribution at (2.5) if $h \sim Cn^{-1/r}$.

Finally we outline a proof of the asymptotic normality with variance given at (2.6) under the specific conditions stated above. Observe that we may write $\hat{g}(x|\theta) = \gamma(x|\theta) + \Delta(x|\theta)$, where $\gamma(x|\theta) = G(x|\theta)/A(x|\theta)$, $\Delta(x|\theta) = D(x|\theta)/A(x|\theta)$, $G(x|\theta) = \sum_i g(X_i)K_i(x|\theta)$, $D(x|\theta) = \sum_i \epsilon_i K_i(x|\theta)$ and $A(x|\theta) = \sum_i K_i(x|\theta)$. We shall Taylor-expand about $\theta = \theta_0$, arguing as follows. Assume that $|\theta - \theta_0| \leq n^{-1-\eta_1}h$, where $\eta_1 > 0$ is fixed, that K vanishes outside $[-1, 1]$ and that $x \in [2h, \theta_0 - 2h]$. Then, X_i/θ_0 is further than h away from all integers, uniformly in indices i such that either $K_i(x|\theta)$ or $K_i(x|\theta_0)$ is nonzero. It follows that $\lfloor X_i/\theta \rfloor = \lfloor X_i/\theta_0 \rfloor$, uniformly in such i , for all sufficiently large n . Therefore,

$$X_i(\theta) - X_i(\theta_0) = \theta_0 \lfloor X_i/\theta_0 \rfloor - \theta \lfloor X_i/\theta \rfloor = (\theta_0 - \theta) \lfloor X_i/\theta_0 \rfloor,$$

the latter identity holding uniformly in indices i such that either $K_i(x|\theta)$ or $K_i(x|\theta_0)$ is nonzero. Hence,

$$K_i(x|\theta) - K_i(x|\theta_0) = \sum_{j=1}^2 j^{-1} (\theta_n X_{ni})^j K^{(j)}(x|\theta_0) + \frac{1}{6} (\theta_n X_{ni})^3 K^{(3)}(x|\theta^{(i)}), \quad (4.1)$$

where $K^{(j)}(x|\theta) = K^{(j)}[\{x - X_i(\theta)\}/h]$, $\theta^{(i)} = \theta^{(i)}(x)$ lies between θ and θ_0 , $\theta_n = n(\theta - \theta_0)/h$ and $X_{ni} = \lfloor X_i/\theta_0 \rfloor/n$.

Substituting the expansion at (4.1) into formulae for $G(x|\theta)$, $D(x|\theta)$ and $A(x|\theta)$ we deduce Taylor expansions, in θ and about $\theta = \theta_0$, of the latter quantities, provided $x \in [2h, \theta_0 - 2h]$. This leads to Taylor expansions of $\gamma(x|\theta)$ and $\Delta(x|\theta)$. In particular, with probability 1 and uniformly in $x \in [2h, \theta_0 - 2h]$, we have for all $\eta_3 > 0$,

$$\begin{aligned} \gamma(x|\theta) - \gamma(x|\theta_0) &= \theta_n \left(\frac{\sum_i X_{ni} g(X_i) K'_i(x|\theta_0)}{A(x|\theta_0)} - \frac{G(x|\theta_0) \sum_i X_{ni} K'_i(x|\theta_0)}{A(x|\theta_0)^2} \right) \\ &\quad + O\left[\theta_n^2 \{h^2 + (nh)^{-1/2} n^{\eta_3}\} + |\theta_n|^3\right], \end{aligned} \quad (4.2)$$

$$\begin{aligned} \Delta(x|\theta) - \Delta(x|\theta_0) &= \theta_n \left(\frac{\sum_i X_{ni} \epsilon_i K'_i(x|\theta_0)}{A(x|\theta_0)} - \frac{D(x|\theta_0) \sum_i X_{ni} K'_i(x|\theta_0)}{A(x|\theta_0)^2} \right) \\ &\quad + O\{\theta_n^2 (nh)^{-1/2} n^{\eta_3} + |\theta_n|^3 n^{\eta_3}\}. \end{aligned} \quad (4.3)$$

In deriving these results we make use of the following consequence of assumptions $(M_{X,1})$, $(M_{X,2})$ or $(M_{X,3})$ about the design sequence. If $h_1 = h_1(n) \rightarrow 0$ in such a manner that for some $\xi > 0$, $n^{-1+\xi} \leq h_1 \leq n^{-\xi}$, then the number of values of $X_i(\theta)$, $1 \leq i \leq n$, that lie in an interval of width h_1 is asymptotic to nh_1/θ , in the sense that for each $0 < \theta_1 < \theta_2 < \infty$ and each $0 < \delta < 1$,

$$\begin{aligned} P \left[(1 - \delta)nh_1/\theta \leq \sum_{i=1}^n I\{X_i(\theta) \in (x, x + h_1)\} \leq (1 + \delta)nh_1/\theta \right. \\ \left. \text{for all } 0 \leq x \leq \theta - h_1 \text{ and all } \theta \in [\theta_1, \theta_2] \right] = 1 - O(n^{-C}) \end{aligned}$$

for all $C > 0$. This property, and results about orders of magnitude that are used throughout our proof, may be derived using moment methods. In the case of model $(M_{X,2})$, where the X_i 's represent ordered values of independent random variables, moment calculations are relatively straight forward. This is also true under model $(M_{X,3})$. To calculate moments for the case $(M_{X,1})$ we first delete indices i for which $1 \leq i \leq n^\eta$, where $\eta > 0$ is fixed but arbitrarily small. It is readily shown that by taking η sufficiently small, the error incurred through this step is of smaller order than the bound that is being derived. Then, writing \sum'_i to denote summation over $n^\eta < i \leq n$, we calculate the k 'th moment of $\sum'_i f\{X_i(\theta)\}$, where $f(u)$ might for example denote $I[u \in (x, x + h_1)]$, by writing the moments as

$$\sum'_{i_1} \dots \sum'_{i_k} E[f\{X_{i_1}(\theta)\} \dots f\{X_{i_k}(\theta)\}].$$

This series may be approximated up to a remainder of order n^{-C} , for any given $C > 0$, by using an Edgeworth expansion of the joint density of $(X_{i_1}, \dots, X_{i_k})$. Of course, the event $X_i(\theta) \in A$, for any Borel set A , is equivalent to $\bigcup_k [\{X_i \in (k\theta, (k+1)\theta]\} \cap \{X_i - k\theta \in A\}]$.

With probability 1 the coefficient of θ_n on the right-hand side of (4.2) equals

$$\begin{aligned} & \frac{1}{2}(\mu/\theta_0) h^{-1} \int_0^{\theta_0} g(u) K' \{(x-u)/h\} du + O\{h^2 + (nh)^{-1/2} n^{\eta_3}\} \\ & = \frac{1}{2}(\mu/\theta_0) h g'(x) + O\{h^2 + (nh)^{-1/2} n^{\eta_3}\}, \end{aligned} \quad (4.4)$$

uniformly in $x \in [2h, \theta_0 - 2h]$, for all $\eta_3 > 0$. In the same sense, the coefficient of θ_n on the right-hand side of (4.3) equals $O\{(nh)^{-1/2} n^{\eta_3}\}$, and $\gamma(x|\theta_0) = g(x) + O\{h^2 + h(nh)^{-1/2} n^{\eta_3}\}$. Therefore,

$$\hat{g}(x|\theta) = g(x) + \Delta(x|\theta_0) + cn(\theta - \theta_0)g'(x) + \alpha_1(x|\theta), \quad (4.5)$$

where (i) $c = \mu/(2\theta_0)$, (ii) the stochastic function α_1 satisfies, for $j = 1$,

$$\sup_{x \in [2h, \theta_0 - 2h]} |\alpha_j(x|\theta)| = O\{|\zeta_n(\theta)|\} \quad (4.6)$$

with probability 1, uniformly in θ for which $|\theta - \theta_0| \leq n^{-1-\eta_1}h$, and (iii) for any $\eta_3 > 0$,

$$\zeta_n(\theta) = |\theta_n| \{h^2 + (nh)^{-1/2} n^{\zeta_3}\} + |\theta_n|^3 n^{\zeta_3} + h^2 + h(nh)^{-1/2} n^{\zeta_3}. \quad (4.7)$$

Taking $x = X_i(\theta)$, noting that $X_i(\theta) - X_i(\theta_0) = (\theta_0 - \theta) \lfloor X_i/\theta_0 \rfloor$ if $X_i(\theta) \in [2h, \theta_0 - 2h]$ and $|\theta - \theta_0| \leq n^{-1-\eta_1}h$, and Taylor-expanding on the right-hand side of (4.5), we deduce that

$$\begin{aligned} \hat{g}\{X_i(\theta)|\theta\} &= g\{X_i(\theta_0)\} + \Delta\{X_i(\theta_0)|\theta_0\} \\ &+ n(\theta - \theta_0)(c - X_{ni})g'\{X_i(\theta_0)\} + \alpha_2\{X_i(\theta)|\theta\}, \end{aligned}$$

where α_2 satisfies (4.6). Therefore, since $g\{X_i(\theta_0)\} = g(X_i)$ and $g'\{X_i(\theta_0)\} = g'(X_i)$, then

$$Y_i - \hat{g}\{X_i(\theta)|\theta\} = \epsilon_i - \Delta\{X_i(\theta_0)|\theta_0\} + n(\theta - \theta_0)(X_{ni} - c)g'(X_i) + \beta_i(\theta), \quad (4.8)$$

where $\beta_i(\theta)$ denotes a random function satisfying

$$\sup_{i \in \mathcal{I}} |\beta_i(\theta)| = O\{|\zeta_n(\theta)|\} \quad (4.9)$$

with probability 1, uniformly in θ such that $|\theta - \theta_0| \leq n^{-1-\eta_1}h$, and $\mathcal{I} = \mathcal{I}(n)$ is the set of integers $1 \leq i \leq n$ such that $X_i(\theta_0) \in [3h, \theta_0 - 3h]$.

If $i \in \tilde{\mathcal{I}} \equiv \{1, \dots, n\} \setminus \mathcal{I}$ then, in view of the periodicity of g , (4.8) and (4.9) continue to hold, provided we interpret $\Delta\{X_i(\theta_0)|\theta_0\}$ in the sense that design points, and the associated response variables, are repeated in $(-\theta_0, 0]$ and in $(\theta_0, 2\theta_0]$, and an extra term of the order of $|g\{X_i(\theta)\} - g(X_i)|$, i.e. the order of $n|\theta - \theta_0|$, is added to the right-hand side of (4.9). That is, (4.7) holds for $i \in \tilde{\mathcal{I}}$, with

$$\sup_{i \in \tilde{\mathcal{I}}} |\beta_i(\theta)| = O\{|\zeta_n(\theta)| + n|\theta - \theta_n|\} \quad (4.10)$$

with probability 1.

Combining (4.8)–(4.10); noting that the number of elements of $\tilde{\mathcal{I}}$ equals $O(nh)$; and observing that under the assumptions made about h in condition (b) stated early in this section, (4.7) implies $\zeta_n(\theta) = O(n^{-(1/2)-\eta})$ for some $\eta > 0$, uniformly in θ for which $|\theta - \theta_0| \leq n^{-1-\eta_1}h$, provided η_3 in the definition of $\zeta_n(\theta)$ is chosen sufficiently small; we conclude that

$$n(\theta_0 - \hat{\theta}) = \frac{\sum_i (X_{ni} - c) g'(X_i) [\epsilon_i - \Delta\{X_i(\theta_0)|\theta_0\}]}{\sum_i \{(X_{ni} - c) g'(X_i)\}^2} + o_p(n^{-1/2}).$$

The ratio on the right-hand side is asymptotically Normally distributed with zero mean and variance

$$\begin{aligned} & \frac{\sum_i (X_{ni} - c)^2 g'(X_i)^2}{\{\sum_i (X_{ni} - c)^2 g'(X_i)^2\}^2} \sigma^2 \\ & \sim \left\{ n(\mu/\theta_0)^2 \int_0^1 (u - \frac{1}{2})^2 du \cdot \theta_0^{-1} \int_0^{\theta_0} g'(u)^2 du \right\}^{-1} \sigma^2 = n^{-1} \tau^2. \end{aligned}$$

In deriving the asymptotic equivalence property claimed just above, observe that $g'(X_i) = g'\{X_i(\theta_0)\}$, that $X_i(\theta_0)$ is asymptotically Uniformly distributed on $(0, \theta_0)$, and that if $i = i(n) \rightarrow \infty$ in such a manner that $i/n \rightarrow u$ where $0 < u < 1$, then $X_{ni} \rightarrow \mu u/\theta_0$ in probability as $n \rightarrow \infty$.

REFERENCES

- AKAIKE, H. (1960). Effect of timing error on the power spectra of sampled-data. *Ann. Inst. Statist. Math.* **11**, 145–165.
- AXELROD, T., ALCOCK, C., ALLSMAN, R., ALVES, D., BECKER, A., BENNETT, D., COOK, K., FREEMAN, K., GRIEST, K., GUERN, J., LEHNER, M., MARSHALL, S., PETERSON, B., PRATT, M., QUINN, P., RODGERS,

- A., STUBBS, C., SUTHERLAND, W., and WELCH, D. (1994). Statistical issues in the MACHO project. In *Statistical Issues in Modern Astronomy*, Feigelson, E. and Babu, G. (eds). Springer, New York.
- BEUTLER, F.J. (1970). Alias-free randomly timed sampling of stochastic processes. *IEEE Trans. Inform. Theor.* **IT-16**, 147–152.
- BICKEL, P.J., KLAASSEN, C.A.J., RITOV, Y. AND WELLNER, J.A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins, Baltimore.
- BRILLINGER, D.R. (1986). Regression for randomly sampled spatial series: the trigonometric case. *Ann. Appl. Prob.* **23**, 275-289.
- DEEMING, T.J. (1975). Fourier analysis with unequally-spaced data. *Astrophysics and Space Science* **36**, 137-158.
- FERLET, R., MAILLARD, J.P. AND RABAN, B. (1997). *Variable Stars and the Astrophysical Returns of Microlensing Surveys*. Editions Frontières, Gif-sur-Yvette, France.
- FRIEDMAN, J. (1984). A variable span smoother. Technical Report No. 5, Laboratory for Computational Statistics, Department of Statistics, Stanford University.
- FISHER, R.A. (1929). Tests of significance in harmonic analysis. *Proc. Roy. Soc. Ser. A* **125**, 54–59.
- HANNAN, E.J. (1973). The estimation of frequency. *J. Appl. Probab.* **10**, 510–519.
- HANNAN, E.J. (1974). Time series analysis. *IEEE Trans. Auto. Control* **AC-19**, 706–715.
- HOFFMEISTER, C., RICHTER, G. AND WENZEL, W. (1985). *Variable Stars*. Springer, New York.
- IBRAGIMOV, I.A. AND HAS'MINSKII, R.Z.(1981). *Statistical Estimation: Asymptotic Theory*. Springer, New York.
- IVANOV, A.V. (1979). A solution of the problem of detecting hidden periodicities. *Theor. Probab. Math. Statist.* **20**, 51–68.
- LAFLER, J. and KINMAN, T.D. (1965), An RR Lyrae survey with the Lick 20-inch astrograph II. The calculation of RR Lyrae periods by electronic computer. *Astrophysical Journal Supplement Series* **11**, 216–222.
- LOMB, N.R. (1976). Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science* **39**, 447-462.
- MCDONALD, J.A. (1986). Periodic smoothing of time series. *SIAM J. Scientific Statist. Comput.* **7**, 665–688.
- QUINN, B.G. AND THOMPSON, P.J. (1991). Estimating the frequency of a periodic function. *Biometrika* **78**, 65–74.
- REIMANN, J.D. (1994). *Frequency Estimation Using Unequally-Spaced Astronom-*

- ical Data*. PhD Thesis, Department of Statistics, University of California, Berkeley.
- REIMANN, J.D. (1999). Frequency estimation by randomly jittered sampling. Manuscript.
- RENSON, R. (1978). Méthode de recherche des périodes des étoiles variables. *Astronomy and Astrophysics* **63**, 125–129.
- RICE, J. AND ROSENBLATT, M. (1988). On frequency estimation. *Biometrika* **75**, 477–484.
- SCARGLE, J.D. (1982). Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. *Astrophysical Journal* **263**, 835–853.
- WALKER, A.M. (1971). On the estimation of a harmonic component of a time series with stationary independent residuals. *Biometrika* **58**, 21–36.
- WHITTLE, P. (1959). Sur la distribution du maximum d'un polynome trigonométrique à coefficients aléatoires. *Colloques Int. Centre Nat. Recherche Sci.* **87**, 173–184.