THE ASTROPHYSICAL JOURNAL, 516:315–323, 1999 May 1

OPTIMUM PERIOD SEARCH: QUANTITATIVE ANALYSIS

A. SCHWARZENBERG-CZERNY
Astronomical Observatory of Adam Mickiewicz University, ul. Słoneczna 36, 60-286 Poznań, Poland, and Copernicus Astronomical Center, ul. Bartycka 18, 00-716 Warsaw, Poland; alex@camk.edu.pl
Received 1998 June 2; accepted 1998 December 8

ABSTRACT

Statistical properties of two broad classes of methods used in period search, namely, phase binning and model function methods, are compared. We employ hypothesis-testing theory to study these methods and present closed analytical formulae for evaluation of the sensitivity of period search, for different kinds of signals. Based on this theory, we draw two conclusions: (1) the methods using smooth model functions are generally more sensitive than those using phase binning and (2) the resolution of the model functions should match structures in the detected signal. Both excess and insufficient resolution result in decreased detection sensitivity. Finally, we demonstrate that within the broad class of the methods discussed, methods utilizing the same models but different statistics generally are equally sensitive. Our considerations apply to most existing period-search methods, which enable formulation of statistical detection criteria.

Subject headings: binaries: eclipsing — methods: numerical — methods: statistical — pulsars: general — stars: oscillations — X-rays: stars

1. INTRODUCTION

Astronomical time series often differ by their uneven sampling from the evenly sampled series, encountered in economics and communications and discussed in the vast statistical literature. Thus astronomers are often forced to help themselves by inventing methods to analyze their time series. In this paper we consider methods for period searching, which admit formulation of objective statistical detection criteria, at least in principle. Examples of such methods used in astronomy are given by Lafler & Kinman (1965), Jurkevich (1971), Lomb (1976), Stellingwerf (1978), Ferraz-Mello (1981), Scargle (1982), Schwarzenberg-Czerny (1989, hereafter Paper I), Davies (1990), Foster (1994), Schwarzenberg-Czerny (1996, 1997a, hereafter Papers II and III, respectively). The detection criteria are based on the classical *hypothesis-testing theory* of statistics (e.g., Eadie et al. 1971; Lupton 1993). The null hypothesis H_0, to be tested in the process of period search, states that a time series consists of a pure noise. In other words, H_0 states that the amplitude of the (deterministic) periodic component of an input signal vanishes, $A = 0$. Hence, detection criteria *do not* depend on the properties of the periodic component for $A > 0$.

The wealth of methods employed in period searching raised the question of a choice of the optimum one. In statistics optimum sensitivity is obtained by fitting of data with their (hypothetical) realistic model (e.g., § 2.3). In that sense no single optimum method valid for all types of data and/or signals exists. In other words, (1) for $A > 0$ sensitivity of a method *does* depend on the type of input signal, i.e., on the alternative hypothesis H_1. In practice, the shape of the deterministic component in the input signal is rarely known a priori. Hence, (2) general methods have to rely on a simple mathematical model, as a first guess. It is a fact of profound consequences that the guessed model often differs considerably from the observed process. Both facts, (1) and (2), cause the evaluation of the detection performance of a given method to be a rather complex process. Occasionally, extensive Monte Carlo simulations were conducted to study performance of various methods and to compare them (Heck, Manfroid, & Mersch 1985; De Jager, Swanepoel, & Raubenheimer 1989). In the present paper we take a more systematic approach and develop an analytical theory of the sensitivity of the period search. Our discussion applies to the broad class of methods defined more precisely in § 2.1. Our measure of sensitivity for signal detection is the test power, discussed in § 2.2. Since the sensitivity depends on the properties of a mixture of a noise *and* of the deterministic component, no known analytical probability distributions from general statistics apply. To proceed further, we assume that the deterministic component is so small that the perturbation methods may be employed for construction of the relevant probability distribution (§ 2.3).

Up to this point our considerations are quite general. In order to illustrate our discussion we consider in detail the von Mises family of signal shapes. The example has broad enough scope to be useful in many interesting applications (§ 3). In analysis of our example we expand the input signal in terms of one of two families of the model functions, the orthogonal trigonometric (Fourier) polynomials and the step functions (phase bins). Given the theory and a signal shape, a number of interesting questions concerning the choice of the best method can be answered. These involve the determination of the optimum kind of model (§ 4.1) and its sophistication, i.e., the desired number of its parameters (§ 4.2). Finally, we consider dependence of sensitivity on the signal-to-noise ratio and on the number of the observations (§ 5).

For an excellent introduction to hypothesis testing, the reader may refer to the books by Eadie et al. (1971) and Lupton (1993), or to the summary by Press et al. (1992). Few statistical texts discuss unevenly sampled time series; examples of some that do are those by Mardia (1972), Deeming (1975), and Bloomfield (1976). Statistical functions are contained in many commercial software libraries. We also find useful the source code of the statistical functions published by Press et al. (1992).

2. STATISTICAL HYPOTHESES AND TEST POWER

2.1. *Vector Norms as Detection Statistics*

Our considerations are restricted to a well-defined class of period-search methods, which employ a merit function Θ that is calculated in such a way as to be independent of the kind of data. These methods admit the detection criterion which involves a critical value, Θ_c, independent of the data (§ 2.2). To simplify our derivations, we assume that Θ is a quadratic norm: $\Theta \in L^2$. Most of the methods admitting the data-independent detection criterion known from the literature either belong to the L^2 class or may be approximated by a class member. This class also contains methods in which multiple frequencies ω contribute to a single value of Θ, though in an a priori prescribed way, e.g., cepstrum, multiharmonic Fourier series, etc. However, methods like CLEAN and CLEANEST, in which the number of the frequencies considered depends on the data, are not part of this class. These methods are more complex to evaluate, since they require the sequential tests theory (Eadie et al. 1971; Fisz 1963). An analogy between linear and nonlinear equation solvers exists: the former always work but are occasionally inefficient, while the latter may be very efficient but will often yield wrong solutions.

Let an observed time series consist of n observations x_i obtained at times t_i, $i = 1, \ldots, n$. Unless stated otherwise, we shall assume throughout the paper that x_i are sampled from the Gaussian distribution, for any fixed t_i. It is convenient to consider x_i as components of an n-dimensional vector x. For these vectors we define a scalar product $\langle \cdot, \cdot \rangle$ and a norm $\| \cdot \|$, such that

$$\langle \xi, \eta \rangle = \int_0^{2\pi} \overline{\xi(\phi)} \eta(\phi) d\mu(\phi)$$

$$= \sum_{i=1}^{n} \mu_i \bar{\xi}_i \eta_i \, , \tag{1}$$

$$\|\xi\|^2 = \langle \xi, \xi \rangle \, , \tag{2}$$

where ξ and η are arbitrary vectors and $d\mu$ corresponds to the Stieltjes integral. For the step weight function μ, the integral in equation (1) reduces to a weighted sum. The natural weights for a set of discrete observations are inverse variances:

$$\mu_i = V^{-1}\{x_i\} \, . \tag{3}$$

Let us adopt a set of r orthogonal vectors $\phi^{(l)}$, $l = 1, \ldots, r$, such that $r < n$, $\langle \phi^{(k)}, \phi^{(l)} \rangle = \delta_{kl}$. We shall use linear combinations of these vectors as models of observations. In vector language, $\phi^{(k)}$ form an orthogonal basis of an r-dimensional subspace X_r of the n-dimensional vector space of observations X_n. We call this subspace X_r a subspace of models. An arbitrary vector of observations $x \in X_n$ may be decomposed uniquely into combination of a vector from the model space $x_\parallel \in X_r$ and a vector of *residuals*, $x_\perp$, such that their norms correspond to the following sums of squares:

$$\|x_\parallel\|^2 \equiv \sum_{l=1}^{r} |\langle x, \phi^{(l)} \rangle|^2 \, , \tag{4}$$

$$\|x_\perp\|^2 \equiv \|x\|^2 - \|x_\parallel\|^2 \, , \tag{5}$$

If components of x are independent random variables with the standard normal distributions $N(0, 1)$, the norms $\|x\|^2$, $\|x_\parallel\|^2$, and $\|x_\perp\|^2$ follow the $\chi^2(d)$, $\chi^2(d_\parallel)$, and $\chi^2(d_\perp)$ distributions. Here d, $d_\parallel$, and $d_\perp = d - d_\parallel$ denote numbers of degrees of freedom corresponding to the whole data, model, and residuals, respectively. Depending on whether an average value has been subtracted from all data x or not, $d = n - 1$ and $d_\parallel = r - 1$ or $d = n$ and $d_\parallel = r$. The norms $\|x_\parallel\|^2$ and $\|x_\perp\|^2$ measure the quality of the fit of the model series

$$x_\parallel = \sum_{l=1}^{r} \langle x, \phi^{(l)} \rangle \phi^{(l)} \tag{6}$$

to the observed series x. Large $\|x_\parallel\|^2$ or small $\|x_\perp\|^2$ indicates a good fit, i.e., a possible detection of the model signal. In practical applications, these norms suffer from the scale-factor indeterminacy of μ_i in equation (1). One has to use their ratio in order to get rid of the indeterminacy. By virtue of the Fisher lemma, $\|x_\parallel\|^2$ and $\|x_\perp\|^2$ are statistically independent as long as the null hypothesis H_0 holds (§ 2.2). Hence their ratio is sampled from the Fisher-Snedecor F-distribution. The partial norms $\|x_\parallel\|^2$ and $\|x_\perp\|^2$ divided by the total norm obey the particular kind of beta distribution, which is related to the F-distribution; the statistic on the left is related to the distribution shown at the right (e.g., Abramovitz & Stegun 1971; Bickel & Doksum 1977; Paper III):

$$\begin{array}{ll} \text{Statistic} & \text{Distribution} \\ \Theta_F \equiv \dfrac{d_\perp \|x_\parallel\|^2}{d_\parallel \|x_\perp\|^2} & F(d_\parallel, d_\perp; \Theta_F) \\ \Theta_\parallel \equiv \dfrac{\|x_\parallel\|^2}{\|x\|^2} & I_{\Theta_\parallel}(d_\parallel/2, d_\perp/2) \\ \Theta_\perp \equiv \dfrac{\|x_\perp\|^2}{\|x\|^2} & I_{\Theta_\perp}(d_\perp/2, d_\parallel/2) \end{array} \tag{7}$$

where

$$\Theta_\parallel = 1 - \Theta_\perp = d_\parallel \Theta_F / (d_\perp + \Theta_F) \tag{8}$$

and $I_\Theta(a, b)$ is an incomplete beta function (Abramovitz & Stegun 1971). As discussed elsewhere (e.g., Papers I and III), most of the statistics used in period searching correspond to one of the Θ statistics listed above. In particular, in Paper I we demonstrate how the string-length statistics (Lafler & Kinman 1965; Dvoretsky 1983) may be approximated by a Θ statistic.

2.2. *Sensitivity*

Let us adopt a general model of a signal x, consisting of a Gaussian white noise n and of a deterministic function s scaled by the amplitude factor $Ad^{1/2}$: $x = n + Ad^{1/2}s$. We assume here that the normalization of the deterministic function s is such that $\|s\| = 1$ and $\langle s, 1 \rangle = 0$ (eq. [B3]). The factor $Ad^{1/2}$ is consistent with A^2 being the signal-to-noise ratio (S/N):

$$\text{S/N} \equiv \|Ad^{1/2}s\|^2 / E\{\|n\|^2\} = A^2 d/d \, ,$$

where we exploited the properties of the χ^2 random variable $\|n\|^2$ (eqs. [3] and [A5]).

Detection criteria and sensitivity in time series analysis correspond respectively to hypothesis testing and test power in statistics (e.g., Eadie et al. 1971). In order to detect the deterministic signal s, we try to choose among two hypotheses: a null hypothesis H_0 stating that the observed signal consists of a pure noise, i.e., $A = 0$, and an alternative hypothesis H_1, that the observed signal consists of the noise *and* the deterministic signal of amplitude $A > 0$. Of particular relevance are two probability distributions of a given statistic Θ, one which assumes H_0 is true and another which assumes H_1 is true: $P(0; \Theta)$ and $P(A; \Theta)$, respectively. The null hypothesis H_0 is rejected and the deterministic signal is detected if according to H_0, observation of a value of Θ exceeding its actually recorded value is very improbable: $P(0; \Theta) > 1 - \alpha$, where $\alpha \ll 1$ and $1 - \alpha$ is called the confidence level. The choice of α and $1 - \alpha$ fixes expected fractions of false detections and of correct rejections, respectively. A given fixed value of α corresponds to the critical value of Θ_c such that

$$1 - \alpha = P(0; \Theta_c) . \tag{9}$$

Hence, the detection criterion may be formulated as $\Theta > \Theta_c$. This criterion depends solely on the distribution for H_0, and not on that for H_1.

Sensitivity of the test does depend on the distribution for H_1, however. The more the H_0 and H_1 distributions differ, the more sensitive is the statistic Θ. Quantitatively, sensitivity is measured by the power $1 - \beta$ of the statistic, defined as

$$1 - \beta = 1 - P(A; \Theta_c) , \tag{10}$$

where Θ_c is fixed and corresponds to a given value of α. Thus β and $1 - \beta$ yield expected fractions of missed and of correct detections, respectively. The higher the fraction of correct detections, the more sensitive a given method is. In this paper we adopt the power (eq. [10]) as a quantitative measure of sensitivity of methods for period search. Our comparison of methods and selection of the optimum one rely on test power.

2.3. *Asymptotic Distribution for Small Amplitude*

For the moment we restrict our attention to just one of three Θ statistics, namely the F-statistic $\Theta_F = d_\perp \|x_\parallel\|^2 / d_\parallel \|x_\perp\|^2$. We are unable to derive analytically its exact probability distribution $P(A; \Theta_F)$ for H_1, i.e., for

$$\Theta_F = d_\perp \|(n + A d^{1/2} s)_\parallel\|^2 / d_\parallel \|(n + A d^{1/2} s)_\perp\|^2$$

and for $A > 0$ (cf. § 2.2). To proceed further, we assume that A is small and adopt a perturbation approach. We assume that the perturbed distribution for H_1, $P(A; \Theta_F)$ in equation (10), is close to the unperturbed distribution for H_0, $P(0; \Theta_F)$ in equation (9). Further, we assume that a perturbation affects the mean and variance of the distribution but has little effect on its shape $R(\cdot)$:

$$P(A; \Theta_F) \approx R(\Gamma) , \tag{11}$$

where

$$\Gamma = \frac{\Theta_F - E\{A; \Theta_F\}}{\sqrt{V\{A; \Theta_F\}}} , \tag{12}$$

$$P(0; \Theta_F) \equiv R\left(\frac{\Theta_F - E\{0; \Theta_F\}}{\sqrt{V\{0; \Theta_F\}}}\right) . \tag{13}$$

For large $d_\parallel$ and $d_\perp$, such that $1 \ll d_\parallel \ll d_\perp$, $R(\Gamma)$ approaches the normal distribution $N(0,1)$, by virtue of the central limit theorem. The only remaining task is calculation of moments of the distribution $P(A; \Theta_F)$. We calculate the relevant moments, $E\{A; \Theta_F\}$ and $V\{A; \Theta_F\}$, by expanding powers of Θ_F^1 and Θ_F^2 in a series according to powers of A, our small parameter. The final results are obtained by truncating of terms higher than A^2 and taking the expected values. Details of the calculation are contained in Appendix A. Retaining only the leading term in equation (11), we obtain

$$\Gamma \approx \frac{\Theta_c - E\{0; \Theta_F\}}{\sqrt{V\{0; \Theta_F\}}} - A^2 d \frac{\|s_\parallel\|^2}{\sqrt{2 d_\parallel}} + O\left(A^4, \left(\frac{d_\parallel}{d_\perp}\right)^1\right) , \tag{14}$$

where $E\{0; \Theta_F\}$ and $V\{0; \Theta_F\}$ correspond to the mean and variance of the probability distribution for H_0. In the latter case the Fisher-Snedecor $F(d_\parallel, d_\perp)$ distribution holds, so that $E\{0; \Theta_F\} \sim 1$ and $V\{0; \Theta_F\} \sim 2/d_\parallel$ (eqs. [A8] and [A9]). The A^2 term corresponds to the change in the expected values $E\{0; \Theta_F\} - E\{A; \Theta_F\}$. Indeed, by taking $V\{A; \Theta_F\} \equiv V\{0; \Theta_F\}$ and allowing the mean value $E\{A; \Theta_F\}$ to vary with A and repeating the calculations of Appendix A, one obtains the A^2 term in the same form. Note that the A^2 term is proportional to the S function introduced in Paper I:

$$S \equiv (\partial E\{A, \Theta_F\} / \partial A^2) / \sqrt{2 V\{A, \Theta_F\}} = d \|s_\parallel\|^2 / 2\sqrt{d_\parallel} .$$

In fact, the present derivation constitutes generalization of the results from Paper I.

Equation (14) may be derived in another way, up to terms of order A^2. For this purpose we observe that for $A > 0$ the norm $\|n + A d^{1/2} s\|^2$ has the noncentral $\chi^2(d, A^2 d \|s\|^2)$ distribution. The noncentral χ^2 distribution may be approximated by the scaled central χ^2 distribution (Eadie et al. 1971). Hence, the distribution $P(A; \Theta_F)$ for H_1: $A > 0$ may be approximated by the scaled Fisher-Snedecor distribution. In this way we obtain the A^2 term in the same form as in equation (14) above. Hence, we are confident that for the alternative hypothesis H_1: $A > 0$, equation (14) for the probability distribution of the Θ_F statistic holds under quite general assumptions.

Results obtained so far in this section hold for Θ_F only. Now we extend our results onto the remaining Θ statistics. The Fisher-Snedecor Θ_F statistic is completely equivalent to the two kinds of beta statistics $\Theta_\parallel$ and $\Theta_\perp$ (§ 2.1; see, e.g., Abramovitz & Stegun 1971; Bickel & Doksum 1977; Paper III). Conversion between these statistics and their corresponding distributions is effected by a simple change of variable (eq. [8]). Hence, within our approximation H_1: $A > 0$, equivalence of Θ_F, $\Theta_\parallel$, and $\Theta_\perp$ holds for their distributions: Fisher-Snedecor $P(A; \Theta_F)$, beta $P(A; \Theta_\parallel)$, and another beta $P(A; \Theta_\perp)$. Thus, all conclusions obtained for Θ_F, including equation (14), hold for $\Theta_\parallel$ and $\Theta_\perp$ as long as $A \ll 1$ and a suitable change of variable is applied. For the special case of the basis of step functions this has already been demonstrated in Paper I. In the present derivation nothing was assumed about the sampling pattern and shapes of the signal and model functions. Thus our results apply quite generally, both for fitting of a smooth model function and for phase-folding and binning methods (i.e., for a step function). The only restrictive assumption here is that the amplitude A is small enough to neglect terms of order higher than A^2. Combining equations (9)–(14), one obtains

$$1 - \beta = 1 - R\left(R^{-1}(1 - \alpha) - A^2 d \frac{\|s_\parallel\|^2}{\sqrt{2 d_\parallel}} + O\left(A^4, \frac{d_\parallel}{d_\perp}\right)\right) , \tag{15}$$

where in the asymptotic limit of $1 \ll d_\parallel \ll d$ corresponding to the central limit theorem, the distribution R approaches the standard normal distribution:

$$R(\Gamma) = \frac{1 + \mathrm{erf}(\Gamma/\sqrt{2})}{2} . \tag{16}$$

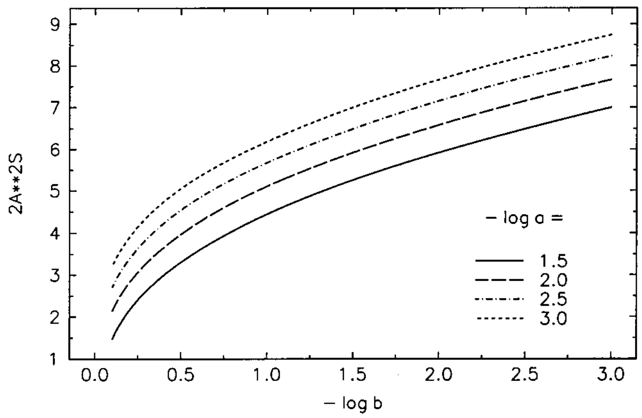

FIG. 1.—Critical values of the data-quality product $2A^2S \equiv A^2d\|s_\parallel\|^2/d_\parallel^{1/2}$ are plotted against required sensitivity level $-\log_{10}\beta$, where $1-\beta$, A^2, d, $d_\parallel$, and $\|s_\parallel\|$ denote the expected fraction of correct detections, signal-to-noise ratio, number of observations and number of parameters of the model, and fraction of power fitted by the model, respectively. Lines are plotted for significance levels $1-\alpha$, where $-\log_{10}\alpha = 1.5, 2.0, 2.5, 3.0$ from bottom to top. Present calculations are valid for $A^2 \ll 1$ and for the standard normal distribution consistent with the limit theorem (eq. [16]). Note that for the optimum model, $\|s_\parallel\| \approx 1$.

The number of observations $d = n$ and signal-to-noise ratio A^2 enter equation (15) only in the combination A^2d. Solving equation (15), one obtains

$$A^2d\,\frac{\|s_\parallel\|^2}{\sqrt{d_\parallel}} = \sqrt{2}[R^{-1}(1-\alpha) - R^{-1}(\beta)]\ . \tag{17}$$

Hence, a given sensitivity level, corresponding to $A^2d = \text{constant}$, may be obtained either for a large number of observations or for a large signal-to-noise ratio. This scaling of sensitivity with A and d is essential for the design of experiments. Equation (17) may be used to calculate the minimum number of observations required to obtain a given signal-to-noise ratio. Note, that for the optimum model $\|s_\parallel\| \approx 1$ (§ 4.2). In Figure 1 we plot the critical value of $A^2d\|s_\parallel\|^2/d_\parallel^{1/2}$, required to achieve sensitivity $1-\beta$, for the standard normal distribution, i.e., for $R^{-1}(P) = \sqrt{2}\,\mathrm{erf}^{-1}(2P-1)$ (eq. [16]).

3. AN EXAMPLE: VON MISES INPUT SIGNAL

Up to this point our considerations have been quite general. Now, to illustrate applications of equation (14), we are going to consider a specific example of the input signal. Without much difficulty similar results may be obtained for different shapes of input signals. This is facilitated by the fact that the test power depends on the shape of the input signal solely via the norm

$$\|s_\parallel\|^2 \equiv 1 - \|s_\perp\|^2 = \sum_{l=1}^{r} |\langle s, \phi^{(l)}\rangle|^2\ . \tag{18}$$

Prior to the calculation of the scalar products $\langle s, \phi^{(l)}\rangle$ one has to select a specific Stieltjes weight function, orthogonal basis vectors $\phi^{(l)}$, and, of course, the shape of the input signal. The weight function μ is related to the sampling of the time series. The continuous uniform weight function adopted here, $d\mu = d\phi/2\pi$, is a suitable approximation for discrete observations either sampled evenly or sampled randomly with the uniform distribution. Thus, for an estimate of the expected value of the norm $\|s_\parallel\|^2$, the Stieltjes integral reduces to the ordinary Riemann integral. This is not a particularly restrictive assumption, since some freedom remains in adopting the orthogonal basis-most suitable for a particular sampling pattern, e.g., the Dirac δ distributions for the discrete case. For the present purposes we consider two different sets of orthonormal basis vectors: (1) the step functions ψ_S, related to phase bins, and (2) the Fourier harmonics ψ_F:

$$\psi_S^{(l)}(\phi) = \begin{cases} \sqrt{r}\,, & l - \dfrac{1}{2} < \dfrac{N\phi}{2\pi} < l - \dfrac{1}{2}\,, \\ 0\,, & \text{otherwise,} \end{cases} \tag{19}$$

$$\psi_{FC}^{(l)}(\phi) = \sqrt{2}\cos l\phi\ , \tag{20}$$

$$\psi_{FS}^{(l)}(\phi) = \sqrt{2}\sin l\phi\ . \tag{21}$$

These two sets cover most models used in practical methods. Because of the symmetry of the von Mises function, its sine components $\langle s, \psi_{FS}\rangle$ all vanish and need no further considerations.

For the deterministic component of the input signal, s, we adopt the von Mises function $s(\phi) \sim e^{\kappa\cos\phi}$, where κ is the shape parameter and ϕ is phase. The von Mises function is a periodic analog of the Gaussian bell function (e.g., Mardia 1972). Its advantage is its ability to mimic a wide range of signals encountered in astronomy, from sinusoidal oscillations ($\kappa = 0$) to the narrow pulses and eclipses ($\kappa \to \infty$) (Fig. 2). Because of the normalization adopted in § 2.2, namely, $\langle s, 1\rangle = 0$ and $\|s\| = 1$, we have to scale the von Mises function accordingly. The detailed formulae are listed in the Appendix B. We do not pretend that our assumptions on the weight function, the model functions, and the input signal are optimal. However, they should emulate reasonably well most cases encountered in practice. In a similar way one can perform the calculations for a different set of the assumptions, if desired.

Expanding the von Mises signal into Fourier cosine harmonics, we obtain $\langle s, \psi_{FC}^{(0)}\rangle = 0$ and

$$\langle s, \psi_{FC}^{(l)}\rangle = \frac{\sqrt{2}}{\pi}\,a(\kappa) \times \left[\int_0^\pi d\phi\cos l\phi\; e^{\kappa\cos\phi} - I_0(\kappa)\int_0^\pi d\phi\cos l\phi\right].$$

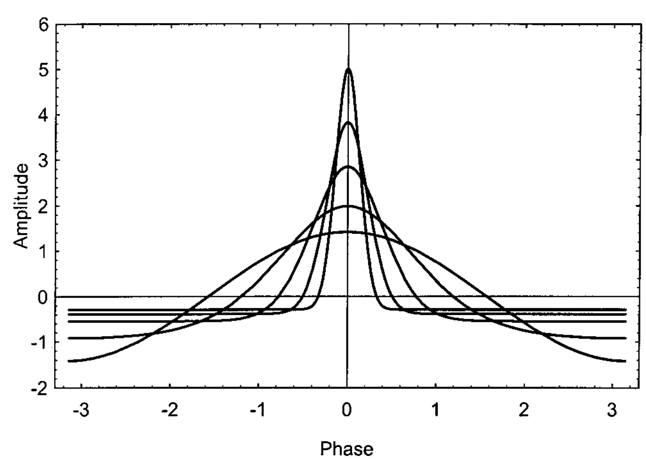

FIG. 2.—Family of von Mises signals, for values of $\kappa = e^i - 1$, $i = 0, 1, \ldots, 4$.

The second integral vanishes and the first one reduces to a Bessel function I_l (Abramovitz & Stegun 1971, eq. [9.6.19]), thus, finally,

$$\langle s, \psi_{\mathrm{FC}}^{(l)} \rangle = \sqrt{2} a(\kappa) I_l(\kappa) \tag{22}$$

for $l > 0$, where $a(\kappa)$ is defined in equation (B2).

For the step function the calculations are a little more involved. By expansion of s in the series of Bessel functions (eq. [9.6.34] of Abramovitz & Stegun 1971) one obtains

$$\langle s, \psi_S^{(l)} \rangle = \frac{\sqrt{r}}{\pi} a(\kappa) \int_{2\pi(l-1/2)/r}^{2\pi(l+1/2)/r} d\phi \sum_{k=1}^{\infty} I_k(\kappa) \cos l\phi \ .$$

By integration of the sum term by term and by subsequent application of the trigonometric identity $\sin(\alpha + \beta) - \sin(\alpha - \beta) = 2 \cos \alpha \sin \beta$, one obtains

$$\langle s, \psi_S^{(l)} \rangle = 2 \frac{\sqrt{r}}{\pi} a(\kappa) \sum_{k=1}^{\infty} \frac{I_k(\kappa)}{k} \sin \frac{k\pi}{r} \cos \frac{2\pi k l}{r} \tag{23}$$

for $l = 1, \ldots, r$. Note that it suffices to evaluate the Bessel functions I_k only once for all l. An efficient algorithm to do so is described by Abramovitz & Stegun (1971, § 9.12 and eqs. [9.6.26] and [9.6.37]).

4. STATISTICAL PROPERTIES

It follows from equation (14) that for a given amplitude A and number of the parameters of the model $d_\parallel$, the effect of the shape of the model functions affects $\|s_\parallel\|^2$ alone. The larger $\|s_\parallel\|^2$ is, the more sensitive the corresponding period statistics for the particular input signal. Different models are compared in § 4.1. For the optimum choice of the number of model parameters, we consider in § 4.2 the ratio $\|s_\parallel\|^2/d_\parallel^{1/2}$. Finally, in § 5 we discuss the trade-offs involved in choosing the optimum number of observations and sensitivity of the period search.

4.1. *Choice of Model*

In Figure 3 we plot the norm of the modeled signal $\|s_\parallel\|^2$ against the pulse-shape parameter κ. The considered shapes range from the pure sinusoidal oscillations ($\kappa = 0$) to the narrow Gaussian pulses ($\kappa = 5$). Two families of curves are plotted for the Fourier series model and the step function model. For the latter family computations were performed for three different offsets of the bin centers and pulse maxima: 0°, 120°, and 240° in phase. For each family two curves were plotted, corresponding to the number of model parameters $r = 3$ and $r = 11$. These models correspond to the Fourier and step function models using $(r - 1)/2$ harmonics or r phase bins, respectively. For a given number of the parameters r the Fourier series yields large $\|s_\parallel\|^2$ compared to that for phase folding and binning. This is true except for the chance coincidence of the signal peak and bin centers, when phase folding may be marginally better for a range of the signal shapes. Inspection of Figure 3 reveals that typical values of $\|s_\parallel\|$ for the step model are a factor of ϵ smaller than the corresponding values for Fourier series, where $0.75 > \epsilon > 0.50$. This result is perhaps obvious, as the smooth von Mises function is fitted better by smooth Fourier harmonics than by a step function. The worse the fit, the worse the sensitivity of the period search. The loss of the sensitivity for phase binning is particularly severe for smooth signals, i.e., for small κ, and for simple models, i.e., for small r. For very narrow pulses, i.e., for large κ, both types of methods perform equally poorly. Whether decrease of $\|s_\parallel\|^2$ causes significant decrease in sensitivity of step methods depends on different factors present in equation (14). This equation constitutes a tool to quantify the loss of sensitivity. To preserve sensitivity, loss in $\|s_\parallel\|^2$ in equation (14) by a factor of $0.75 > \epsilon > 0.50$ should be matched by the increase of an amplitude by the factor $1.15 < 1/\sqrt{\epsilon} < 1.41$ or by the increase of a number of observations by the factor $1.33 < 1/\epsilon < 2.0$.

So far our considerations were valid only for the von Mises signal. Their extension for general signals encountered in astronomy requires caution. Our discussion remains valid for all discrete observations of light curves and radial velocity curves of pulsating and eclipsing stars and radio and X-ray pulsars, which are sufficiently finely sampled in phase to resolve their basic shape. From our point of view such observations differ insignificantly from continuous functions used in our example. For these data, methods involving a smooth model (e.g., power spectrum, Lomb-Scargle spectrum, the multiharmonic Fourier periodogram, and those χ^2 methods involving a smooth model) are more sensitive than those which rely on phase folding and binning, for the same number of model parameters r. Skewness of the shape, e.g., in light curves of Cepheids, and scatter due to random noise do not much affect our conclusions. Our conclusions do not hold for coarse discrete sampling, unable to resolve important features of a signal. For such sampling, the performance of smooth and step models is the same.

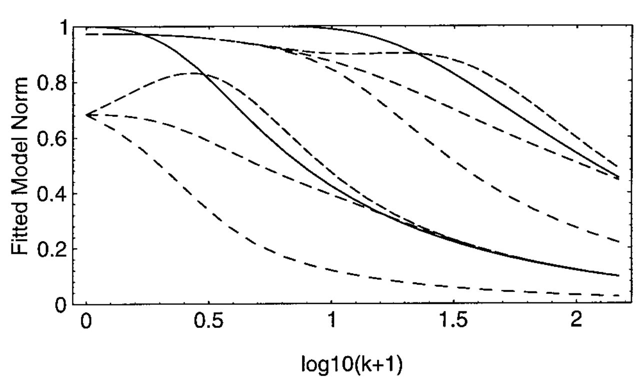

FIG. 3.—Dependence of sensitivity on the shape of the input signal. Specifically we plot the norm $\|s_\parallel\|^2$ of the fitted model against the shape parameter κ of the von Mises signal. The extreme values of $\kappa = 0$ and $\kappa \to \infty$ correspond to sinusoidal and delta function signals, respectively (cf. Fig. 2). Fits were performed for the Fourier series (*continuous line*) and phase-folding-and-binning models (*dashed lines*), for $r = 3$ and 11 parameters, corresponding to 1 and 5 Fourier harmonics or 3 and 11 phase bins, respectively. For phase folding we plot families of curves for the signal peak and bin centers offset by 0, 0.25 and 0.5 bin widths.

4.2. *The Optimum Resolution*

According to equation (14) the sensitivity depends on the size of the model $d_\parallel = r - 1$ solely via the detection power function S (§ 2.3). In Figure 4 we plot S against the number of model parameters r, for a given shape of the input signal. Specifically for this example we used the Fourier series model and the von Mises input signal for $\kappa = 50$. According to equation (14) the maximum of $\|s_\parallel\|^2/d_\parallel^{1/2}$ corresponds to the maximum of sensitivity. Hence, *the model corresponding to the maximum of S is optimal for detection of the given signal.* The optimal model in the present example has $r = 15$ parameters, i.e., 7 harmonics. It is widely appreciated that a good match of model and observations increases sensitivity.

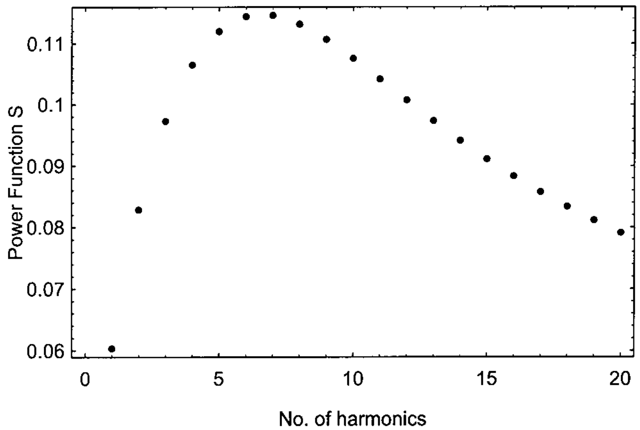

FIG. 4.—Dependence of sensitivity on the complexity of a model. Specifically, we plot the value of the test power function S against the number of harmonics of the model Fourier series. The input signal is the von Mises function for $\kappa = 50$.

However, it is less widely realized that use of excessively fine models, involving a large number of parameters, decreases sensitivity. In the present example all models using more then 7 harmonics are less sensitive then the optimum one.

In order to study dependence of the optimal model on the shape of an input signal, we considered a number of examples similar to the one illustrated in Figure 4, for $0 < \kappa < 400$. In Figure 5 we plot the resolution of the optimum models against the width of their corresponding input signals. Specifically, we plot the FWHM of the highest harmonics of the optimal model, $\pi/(2r - 1)$, against the FWHM of the von Mises signal (eq. [B5]). Inspection of Figure 5 reveals that *resolution of the optimum model matches the characteristic scale of features of the detected signal.*

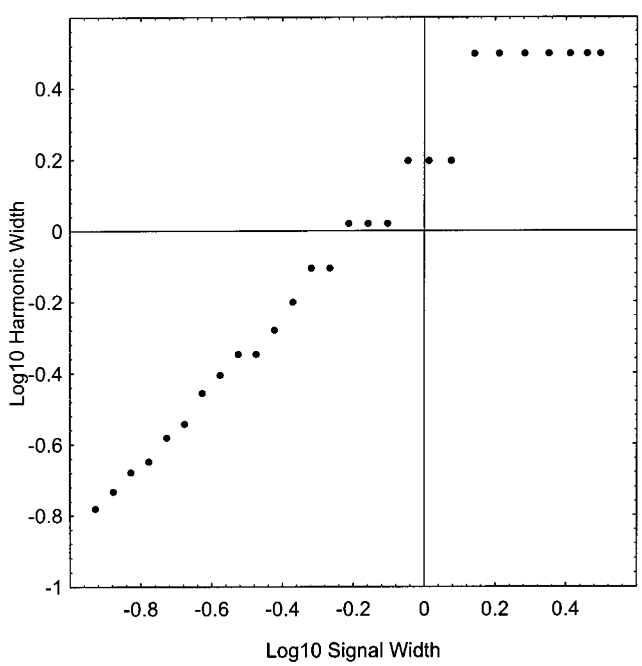

FIG. 5.—Optimum model resolution as function of signal width. Resolution of the Fourier model most sensitive for detection of the von Mises signal is plotted against the FWHM of the signal, for $0 \leq \kappa < 400$, on log scales. As resolution we plot the FWHM of the highest harmonic of a Fourier series.

Our conclusion is that the use of models, which are *either coarser or finer* than the optimal model, *decreases sensitivity* of the period search. Gregory & Loredo (1992) reach the same conclusion using a Bayesian approach. In fact it may be demonstrated that at least in an asymptotic limit Bayesian and classical period-search methods differ only by their effective significance level (Schwarzenberg-Czerny 1998). This inefficiency plagues methods using narrow phase bins, e.g., containing just two observations each, and the related string-length method (Lafler & Kinman 1965; Dvoretsky 1983). In that respect it is important to realize that these methods are not "assumption-free," "model-free," or "nonparametric." On the contrary, they involve implicit models with a large number of parameters, such as averages of all bins, coordinates of string nodes, and so on. A loss of sensitivity corresponds to broader distribution of $dP/d\Theta$ and more overlap of $dP(0; \Theta)/d\Theta$ and $dP(A; \Theta)/d\Theta$ in equations (9) and (10). It may be exhibited either directly or indirectly. The direct manifestation occurs when Θ_c is increased to keep α fixed. The indirect manifestation occurs when, by oversight of the loss, Θ_c is kept fixed, resulting in increase of the effective α. Then the true detections are not lost, but the number of false detections increases. This case applies to methods relying on a large number of models tried in sequence, while by mistake a single trial value of Θ_c is used (e.g., the H test of De Jager et al. 1989). A related phenomenon is called the bandwidth or multiple-trial effect (§ 5).

5. SENSITIVITY VERSUS DATA SIZE

Let us consider two fairly realistic examples of photoelectric observations of an intermediate polar (IP) star and search of RR Lyrae (RR) stars in a CCD survey. We adopt the following general procedure: first we set in equation (15) the values of all parameters except for the one for which the value is unknown; then we solve for that unknown value. The difference is that in the IP case A^2 is known and n unknown, while in the RR Lyrae case n is fixed and A^2 is unknown. In Table 1 we list values of all input parameters, for IP and RR Lyrae cases.

The unknown parameters are indicated in Table 1 by question marks. For both cases we assume a confidence level $1 - \alpha = 0.99$. In equation (15) we use the normal distribution approximation (eq. [16]; cf. Fig. 1).

Rigid statistical criteria are mistrusted by many observers because of the examples of seemingly absurd results. Generally, these examples are caused by errors in application of statistical procedures. The most frequent errors are neglect of (1) the multiple-trial effect (cf. Horne & Baliunas 1986), (2) the correlation-of-residuals effect (cf. Schwarzenberg-Czerny 1991), and (3) use of an invalid probability distribution (cf. Paper III; Schwarzenberg-Czerny 1998). We account for the correlation of residuals in the IP example, applying an equation $d = n/n_{\mathrm{corr}}$ for conversion of the effec-

TABLE 1

PARAMETERS OF SAMPLE DATA SETS

Name	$1-\beta$	$2A^2S$	$d_{\parallel}$	n_{corr}	n	A^2
IP.......	0.5	3.4	3	10	?	0.1
RR......	0.99	6.7	11	1	200	?

tive number of observations d to the true number of observations n. Since usually many frequencies are searched, the multiple-trial correction should apply to the probability distributions discussed here. However, such detailed calculations remain outside the scope of the present simple examples.

5.1. *Monitoring of an Intermediate Polar*

Assume that for a given IP star, the signal-to-noise ratio is fixed by internal stellar properties, pulsation versus flickering, to $A^2 = 0.1$, and that the number of observations n required to detect sinusoidal periodic pulses at least 50% of the time $(1 - \beta = 0.5)$ remains unknown. Further, assume that Fourier models with $d_{\parallel} = 3$, corresponding to pure sinusoid is optimum, i.e., that $\|s_{\parallel}\| = 1$. Next we solve equation (15) for d. Here d denotes the effective number of independent observations. The actual observations are not independent. The consecutive random deviations from the average light curve are correlated because the timescale of flickering is long compared to the sampling interval. For this reason d depends on the number of consecutive correlated observations $n_{\rm corr}$, so that $d = n/n_{\rm corr}$ (§ 5). Finally, for the IP we calculate the required number of observations to be $n = 600$, corresponding to over 3 hours for an integration time of 20 s.

5.2. *Search of RR Lyrae Stars*

In the RR Lyrae example the number of CCD frames $n = 200$ obtained in an observing run is fixed, and the limiting signal-to-noise ratio A^2 corresponding to 99% completeness of the detections remains unknown. In this case we assume that the Fourier model using 5 harmonics $(d_{\parallel} = 11)$ is optimum, hence $\|s_{\parallel}\| = 1$. By solving equation (15) with respect to A^2, we obtain the limit signal-to-noise ratio $A^2 = 0.1$ for 99% completeness. In order to calculate from this ratio the maximum error level of the CCD photometry, one has to assume that a typical amplitude of RR Lyrae stars in question is, say, $D = 0.1$ mag. The prescribed completeness may be reached for those RR Lyrae stars for which CCD photometry for each frame yields an error of $E \leq D/(A^2)^{1/2} = 0.3$ mag.

6. CONCLUSIONS

Our evaluation of methods for period search rests on application of the theory of hypothesis testing for a systematic analytical study of the sensitivity of these methods. Hence our conclusions have wider application than the results of Monte Carlo simulations. Some of our results are quite obvious intuitively:

1. The period-search methods using smooth functions to model the signal are more sensitive than the methods using the step functions, i.e., phase binning (§ 4.1).
2. Sensitivity increases for models that more closely fit features in the signal.
3. However, it is not always appreciated that sensitivity of methods involving an excessive number of parameters is small (however, cf. Gregory & Loredo 1992). In that sense, the so-called nonparametric methods, e.g., involving multiple bins or string length, are actually excessively multiparametric methods.
4. We demonstrated that for optimum sensitivity, the resolution of the model should match the resolution of the signal features (§ 4.2).
5. Our nonobvious result is that the sensitivity of the detection depends on the product of the signal-to-noise ratio A^2 and the number of observations n. The form of this dependence enables prediction of the number of observations required to reach a given level of sensitivity, for a given signal-to-noise ratio.
6. Our important conclusion is that many methods are statistically equivalent, provided that they use a quadratic norm corresponding to any of the Θ statistics with its correct distribution (§ 2.3) and that the model involves the same number and kind of orthogonal functions. Hence, a number of methods relying on phase binning are equivalent for the same number of bins. A number of the methods relying on the Fourier harmonics are statistically equivalent, provided that the same number of harmonics is used. Sensitivity does depend on the model and its number of the parameters and on the kind of detected signal, and not so much on the kind of periodogram statistic. However, one must be aware that many distributions and corresponding detection criteria claimed in the literature are incorrect. One has to employ the correct criteria for validity of the equivalence.
7. Methods using nonorthogonal models are prone to unpredictable variations in their sensitivity, depending on sampling and signal phase. The ordinary power spectrum for uneven sampling suffers from this problem.
8. Observers are reminded to use the correct statistical procedures. These should account for the effective number of searched frequencies and for the correlation of the residuals from the model fit (§ 5). Note that for a large number of methods, incorrect probability distributions are claimed in the literature (see Paper III and Schwarzenberg-Czerny 1997b, 1998, for examples). Our results permit evaluation of the performance of a vast class of period-search methods used in practice. A detailed account of the review of the methods remains outside the scope of the present paper and will be published elsewhere. Below we quote some results without their justification.
9. The sensitivity of our orthogonal multiharmonic AOV method (Paper II) for r chosen to match the signal resolution is optimal. Optimality in the soft sense means that in certain circumstances other methods perform equally well, but none outperforms our method. Specifically, for $r = 2$ or 3 the multiharmonic AOV method is equivalent statistically and computationally to Lomb (1976), Ferraz-Mello (1981), and Scargle (1982) spectra.
10. For signals with sharp pulses or eclipses, requiring $r > 10$, the Akerlof et al. (1994) method is nearly as sensitive as our method, but computationally more efficient. It is more complex to implement, though.
11. Our methods passed extensive tests on over 15,000 light curves obtained by the OGLE team (Udalski et al. 1994).

I would like to thank Don Kurtz for providing an ignition spark to start this work. The organizers of the NATO workshop in Ghent, Belgium, and Chris Sterken in particular, are thanked for their generous hospitality. Thanks are due to the anonymous referee for his helpful general comments and for his meticulous marking of the whole text. Financial support by Polish KBN grants 2 P03C 001 12 and 2 P03D 005 13 is acknowledged.

APPENDIX A

MOMENTS OF F-STATISTICS FOR COMPOSITE SIGNALS

We consider moments of an F-statistic

$$\Theta_F = d_\perp \|(n + A\sqrt{d}s)_\parallel\|^2 / d_\parallel \|(n + A\sqrt{d}s)_\perp\|^2$$

for small $A > 0$ using the perturbation approach. For this purpose we expand the statistic in the power series of the amplitude A, compute its powers Θ_F^1 and Θ_F^2 and take the expected values. First, by using the identity $\|n\|^2 = \langle n, n\rangle$ and the skew linearity of the scalar product we obtain

$$\|n + A\sqrt{d}s\|^2 = \|n\|^2 + A^2 d\|s\|^2 + 2A\sqrt{d}\,\mathrm{Re}\,\langle n, s\rangle\ , \tag{A1}$$

$$\|n + A\sqrt{d}s\|^4 = \|n\|^4 + 2A^2 d\|s\|^2\|n\|^2 + 4A^2 d\,\mathrm{Re}\,\langle n, s\rangle^2 + A^4\|d^2 s\|^4\ , \tag{A2}$$

Using these expansions for the norms we obtain the following expressions for the moments of Θ_F:

$$\begin{aligned}\frac{d_\parallel}{d_\perp} E\{\Theta_F\} &= E\{\|n_\parallel\|^2\}E\left\{\frac{1}{\|n_\perp\|^2}\right\} + A^2 d\left[\|s_\parallel\|^2 E\left\{\frac{1}{\|n_\perp\|^2}\right\}\right.\\ &\quad \left. - \|s_\perp\|^2 E\{\|n_\parallel\|^2\}E\left\{\frac{1}{\|n_\perp\|^4}\right\}\right] + O(A^4)\ ,\end{aligned} \tag{A3}$$

$$\begin{aligned}\frac{d_\parallel^2}{d_\perp^2} E\{\Theta_F^2\} &= E\{\|n_\parallel\|^4\}E\left\{\frac{1}{\|n_\perp\|^4}\right\} + 2A^2 d\left[\|s_\parallel\|^2 E\{\|n_\parallel\|^2\}E\left\{\frac{1}{\|n_\perp\|^4}\right\}\right.\\ &\quad - \|s_\perp\|^2 E\{\|n_\parallel\|^4\}E\left\{\frac{1}{\|n_\perp\|^6}\right\} + 2E\{\mathrm{Re}\,\langle n_\parallel, s_\parallel\rangle^2\}E\left\{\frac{1}{\|n_\perp\|^4}\right\}\\ &\quad \left. - 2E\left\{\frac{\mathrm{Re}\,\langle n_\perp, s_\perp\rangle^2}{\|n_\perp\|^8}\right\}E\{\|n_\parallel\|^4\}\right] + O(A^4)\ .\end{aligned} \tag{A4}$$

Our calculations were simplified by independence of $\|n_\parallel\|^2$ and $\|n_\perp\|^2$ by virtue of Fisher's lemma. The norm $\|n\|^2$ for the noise is sampled from the $\chi^2(d)$ distribution, so

$$E\{\|n\|^{2m}\} = 2^m\Gamma(m + d/2)/\Gamma(d/2) \text{ for } m > -d/2\ . \tag{A5}$$

We have exploited the relation

$$\int x^m \chi^2(d; x)dx \sim \int \chi^2(d + 2m; x)dx = 1\ .$$

Since all n are independent,

$$E\{\mathrm{Re}\,\langle n, s\rangle^2\} \approx \|s\|^2 E\{\|n\|^2\}/d\ .$$

The expressions resulting from equations (A3) and (A4) are lengthy, but they contain many terms proportional to high powers of the small number $d_\parallel/d_\perp \ll 1$. Neglecting all these small terms, one obtains

$$E\{A; \Theta_F\} = E\{0; \Theta_F\} + A^2 d\left(\frac{\|s_\parallel\|^2}{d_\parallel} - \frac{\|s_\perp\|^2}{d_\perp} + \frac{2\|s_\parallel\|^2}{d_\perp d_\parallel}\right) + O\left(A^4, \frac{d_\parallel^2}{d_\perp^2}\right)\ , \tag{A6}$$

$$V\{A; \Theta_F\} = V\{0; \Theta_F\} + 4A^2 d\,\frac{2\|s_\parallel\|^2 - \|s_\perp\|^2}{d_\perp d_\parallel} + O\left(A^4, \frac{d_\parallel^2}{d_\perp^2}\right)\ , \tag{A7}$$

where for the Fisher-Snedecor distribution $F(d_\parallel, d_\perp)$ (e.g., Eadie et al. 1971)

$$E\{0; \Theta_F\} = \frac{d_\perp}{d_\perp - 2}\ , \tag{A8}$$

$$V\{0; \Theta_F\} = \frac{2d_\perp^2(d_\parallel + d_\perp - 2)}{d_\parallel(d_\perp - 2)^2(d_\perp - 4)}\ . \tag{A9}$$

Finally, substituting these moments we obtain

$$\begin{aligned}\frac{\Theta_F - E\{A; \Theta_F\}}{\sqrt{V\{A; \Theta_F\}}} &= \frac{\Theta_F - E\{0; \Theta_F\}}{\sqrt{V\{0; \Theta_F\}}} + \frac{A^2 d}{\sqrt{2}}\left[-\frac{\|s_\parallel\|^2}{\sqrt{d_\parallel}}\right.\\ &\quad \left. + \Theta_F\,\frac{\|s_\perp\|^2 - 2\|s_\parallel\|^2}{d_\perp\sqrt{d_\parallel}}\right] + O\left(A^4, \left(\frac{d_\parallel}{d_\perp}\right)^1\right)\ .\end{aligned} \tag{A10}$$

The series expansions were performed using *Mathematica* (Wolfram 1991). In the calculations the terms higher than those listed were accounted for, to protect against cancellation of terms.

APPENDIX B

VON MISES SIGNAL

Von Mises introduced a periodic analog of the Gaussian bell function $s(\phi) \sim e^{\kappa \cos \phi}$, where ϕ is phase, κ is the shape parameter (e.g., Mardia 1972). In the present paper we adopt the following normalization conditions for the input signal: $\langle s, 1 \rangle = 0$ and $\|s\| = 1$. Evaluating the integrals in the normalization condition and in the mean square derivative $\langle s'^2 \rangle$, using Abramovitz & Stegun (1971) formula [9.6.19], one obtains

$$s(\phi) = a(\kappa)[e^{\kappa \cos \phi} - I(\kappa)] , \tag{B1}$$

where

$$a(\kappa) = \frac{1}{\sqrt{I_0(2\kappa) - I_0(\kappa)^2}} , \tag{B2}$$

$$\langle s, 1 \rangle = 0 , \quad \|s\| = 1 , \tag{B3}$$

$$\|s'^2\|^2 = \frac{\kappa^2 a(\kappa)}{2} [I_0(2\kappa) - I_2(2\kappa)^2] , \tag{B4}$$

$$D_{\rm FWHM} = \frac{1}{\pi} \arccos \left(\frac{\log \cosh \kappa}{\kappa} \right) . \tag{B5}$$

$I_n(\kappa)$ is the modified Bessel function, and $D_{\rm FWHM}$ is the full width at half-maximum. Since the denominators in equations (B1)–(B5) vanish for $\kappa \to 0$, the following asymptotic expansions obtained using *Mathematica* (Wolfram 1991) are useful:

$$s(\phi) = \sqrt{2}\left[\cos \phi + \frac{\kappa}{2} \left(\cos^2 \phi - \frac{1}{2} \right) \right] + O(\kappa^2) , \tag{B6}$$

$$\|s'\|^2 = 1 + \frac{3\kappa^2}{16} + O(\kappa^4) , \tag{B7}$$

$$D_{\rm FWHM} = \frac{\pi - \kappa}{2\pi} + O(\kappa^2) . \tag{B8}$$

In the limit of $\kappa \to \infty$ the von Mises function reduces to the Gaussian bell:

$$s(\phi) \to \sqrt{2(\pi\kappa)^{1/2}}(e^{-\kappa\phi^2/2} - 1/\sqrt{2\pi\kappa}) ; \tag{B9}$$

$$\|s'\|^2 \to \kappa/2 , \tag{B10}$$

$$D_{\rm FWHM} \to (1/\pi)\sqrt{2 \ln 2/\kappa} . \tag{B11}$$

REFERENCES

Abramovitz, M., & Stegun, I. 1971, Handbook of Mathematical Functions (New York: Dover)
Akerlof, C., et al. 1994, ApJ, 436, 787
Bickel, P. J., & Doksum, K. A. 1977, Mathematical Statistics (San Francisco: Holden-Day)
Bloomfield, P. 1976, Fourier Analysis of Time Series: An Introduction (New York: Wiley)
Davies, S. R. 1990, MNRAS, 244, 93
Deeming, T. J. 1975, A&AS, 36, 137
De Jager, O. C., Swanepoel, J. W. H., & Raubenheimer, B. C. 1989, A&A, 221, 180
Dvoretsky, M. M. 1983, MNRAS, 203, 917
Eadie, W. T., Drijard, D., James, F. E., Roos, M., & Sadoulet, B. 1971, Statistical Methods in Experimental Physics (Amsterdam: North-Holland)
Ferraz-Mello, S. 1981, AJ, 86, 619
Fisz, M. 1963, Probability Theory and Mathematical Statistics (New York: Wiley), 536
Foster, G. 1994, AJ, 109, 1889
Gregory, P. C., & Loredo, T. J. 1992, ApJ, 398, 146
Heck, A., Manfroid, J., & Mersch, G. 1985, A&AS, 59, 63
Horne, J. H., & Baliunas, S. 1986, ApJ, 302, 757
Jurkevich, I. 1971, Ap&SS, 13, 154
Lafler, J., & Kinman, T. D. 1965, ApJS, 11, 216
Lomb, N. R. 1976, Ap&SS, 39, 447
Lupton, R. 1993, Statistics in Theory and Practice (Princeton: Princeton Univ. Press)
Mardia, K. V. 1972, Statistics of Directional Data (London: Academic)
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical Recipes (Cambridge: Cambridge Univ. Press)
Scargle, J. H. 1982, ApJ, 263, 835
Schwarzenberg-Czerny, A. 1989, MNRAS, 241, 153 (Paper I)
———. 1991, MNRAS, 253, 198
———. 1996, ApJ, 460, L107 (Paper II)
———. 1997a, ApJ, 489, 941, (Paper III)
———. 1997b, in Proc. Conf. on Time Series in Astronomy (Tel Aviv), ed. D. Maoz & E. Leibovitz (Amsterdam: Kluwer)
———. 1998, MNRAS, 301, 838
Stellingwerf, R. F. 1978, ApJ, 224, 953 (S78)
Udalski, A., Kubiak, M., Szymanski, M., Kaluzny, J., Mateo, M., & Krzeminski, W. 1994, Acta Astron., 44, 317
Wolfram, S. 1991, Mathematica: A System for Doing Mathematics by Computer (Redwood City: Addison-Wesley)