THE ASTRONOMICAL JOURNAL VOLUME 90, NUMBER 11 NOVEMBER 1985

A TEST OF SIGNIFICANCE FOR PERIODS DERIVED USING PHASE-DISPERSION-MINIMIZATION TECHNIQUES

A. F. LINNELL NEMEC
Institute of Ocean Sciences, P.O. Box 6000, 9860 West Saanich Road, Sidney, British Columbia V8L 4B2, Canada

JAMES M. NEMEC
Division of Physics, Mathematics, and Astronomy, 105-24, California Institute of Technology, Pasadena, California 91125
Received 18 June 1985; revised 31 July 1985

ABSTRACT

A test for determining the statistical significance of periods derived using phase-dispersion-minimization techniques (Lafler and Kinman 1965; Stellingwerf 1978) is described. The test is based on Fisher's Method of Randomization, is simple to perform, and can be easily incorporated in standard period-finding programs. It is most valuable in situations where measurement signal-to-noise ratios are low. The secondary periods of five recently discovered double-mode RR Lyrae stars are evaluated using the method.

I. INTRODUCTION

Phase-dispersion-minimization (PDM) techniques (Lafler and Kinman 1965; Stellingwerf 1978) have been used successfully for many years to determine pulsation periods of variable stars from light (and radial velocity) measurements. Despite the wide use of PDM methods, few studies address the question of whether the derived periods are statistically significant, i.e., the corresponding light (or radial velocity) curve is other than what might be expected by chance, in the absence of periodic behavior. In many cases, the significance of a derived period is apparent from a visual inspection of the light curve, and a formal statistical test may only confirm the obvious. In other cases, such as in the identification of secondary pulsations in Cepheid and RR Lyrae variable stars, the light curves may exhibit less obvious structure, making it difficult to determine if the period is real. In such cases, an objective statistical test of significance obviously is desirable.

Stellingwerf (1978) describes an F test for the statistical significance of periods derived using his PDM θ statistic. The test is parametric since it is based on the assumption that the random errors follow a normal distribution. Heck *et al.* (1985) point out that a major drawback of this approach is that it leads to an intractable, if not impossible, calculation of the distribution of a ratio of two quadratic forms. Stellingwerf's assumption that the ratio has an F distribution is not immediately obvious and may be incorrect (Heck *et al.* 1985). Heck *et al.* show that similar distributional difficulties are encountered when parametric methods are applied to the PDM techniques of Lafler and Kinman (1965) and Renson (1978). In the present paper, Fisher's Method of Randomization (see Bradley 1968) is used to define an alternative distribution-free test for the significance of a given period. Randomization is a sound statistical procedure that requires no assumptions regarding the distributional form of the random errors, and eliminates the complicated calculations that were required in previous parametric approaches. The test is simple to apply and can be easily incorporated in existing period-finding programs.

Secondary periods used to determine masses of double-mode RR Lyrae stars in the globular cluster M15, and in the Draco dwarf galaxy, were subjected to the randomization test. For the five stars tested, the secondary period was found to be significant (p value < 0.01) for three of the stars, marginally significant ($0.01 < p$ value < 0.10) for one star, and probably not significant for the last star.

II. NOTATION

Let $y_1, y_2, \ldots, y_n$ be n measurements of a star's magnitude (or radial velocity), corresponding to the n observation times $t_1, t_2, \ldots, t_n$. It will be assumed that the y's are generated according to the model

$$y_i = f(t_i) + e_i \quad (i = 1,2,\ldots n),$$

where $e_1, e_2, \ldots, e_n$ are independent and identically distributed random errors and $f(t_i)$ is an unknown, presumably periodic, function of time. We wish to test the null hypothesis that there is no periodicity, i.e.,

$$H_0\colon f(t) = u = \text{constant for all } t \geqslant 0,$$

against the alternative that f is a periodic function, i.e.,

$$H_1\colon f(t \pm kP) \text{ for all } t \geqslant 0 \text{ and } k = 0,1,2,\ldots,$$

where P is the period and is usually unknown.

For a given trial period P, let

$$\phi(t_i, P) = \text{ fractional part of } (t_i - t_0)/P \quad (i = 1,2,\ldots,n)$$

be the phases of the observation times, where t_0 is arbitrarily chosen to have phase 0. Let $y_1(P), y_2(P), \ldots, y_n(P)$ denote the magnitudes ordered according to phase, for the given P, and let $\theta(P)$ be a suitably normalized version of the sum of squares

$$SS(P) = \sum_{i=1}^{n} \{y_i(P) - \hat{y}_i(P)\}^2,$$

where $\hat{y}_i(P)$ is some function that depends on the magnitudes and phases. The θ statistic measures the dispersion of the observations about a mean light curve corresponding to the trial period P. In general, phase-dispersion-minimization methods attempt to minimize $\theta(P)$, thereby obtaining the "best" possible light curve. Different methods differ only in the definition of $\hat{y}_i(P)$. Lafler and Kinman (1965) use $\hat{y}_i(P) = y_{i-1}(P)$, while Stellingwerf (1978) divides the observations into phase bins and uses $\hat{y}_i(P) = \bar{y}_i(P)$, where $\bar{y}_i(P)$ is the appropriate bin mean. Renson (1978) uses a weighted version of Lafler and Kinman's $\theta(P)$.

Let P_n be the period that minimizes $\theta(P)$. If the observations are periodic, P_n provides an estimate of the period and $\theta(P_n)$ can be used to test the null hypothesis against the alternative. Under H_0, there is no periodic component and $\theta(P)$ reflects random variation regardless of the value of P. When H_1 is true, there is a periodic component and $\theta(P)$ measures random variation only if P is approximately equal to the true period; otherwise, $\theta(P)$ includes both random variation and variation due to phase shifts of the light curve. The minimum value $\theta(P_n)$ should, therefore, be considerably smaller than minima that are expected to occur by chance under H_0. Precise criteria for deciding if $\theta(P_n)$ is sufficiently small to reject H_0 are considered in Sec. III.

III. TESTING THE NULL HYPOTHESIS

a) The Parametric Approach

The usual parametric way of testing the null against the alternative hypothesis is to assume that the random errors have a specific distribution, such as normal. Given the distribution of the errors, it should be possible, at least in theory, to calculate the distribution of the test statistic $\theta(P_n)$ under the null hypothesis. Using this null distribution we can then test whether the observed value of $\theta(P_n)$ is small enough to cast serious doubt on the truth of the null hypothesis. The p value, or probability, under H_0, of obtaining a value of $\theta(P_n)$ at least as extreme as (i.e., less than or equal to) the one observed, is a convenient measure of the strength of evidence against H_0. The smaller the p value the more likely we are to reject H_0 and accept H_1. Typically, a p value less than the nominal value of 0.05 is taken to be sufficient for the rejection of H_0.

To compute the p value, it is necessary to derive the null distribution of $\theta(P_n)$. Consider first the distribution of $\theta(P)$ for fixed value of P. For the Lafler and Kinman (1965), Stellingwerf (1978), and Renson (1978) methods, Heck *et al.* (1985) show that $\theta(P)$ can be written as the ratio of two quadratic forms. It is well known that if two random variables have chi-square distributions and are independent, then their ratio has an F distribution, apart from a constant multiple. However, Heck *et al.* show that if the errors are normally distributed, then, in each of the three cases, the quadratic form that occurs in the numerator of $\theta(P)$ has a distribution other than chi square. Consequently, the usual argument for an F distribution (which Stellingwerf appears to have used) does not apply and the distribution is much more difficult to derive. In some instances, it may be possible to obtain an adequate approximation using numerical devices, such as the method described by Davies (1980). Even if this is possible, the distribution will be quite complicated and will probably involve considerable computation time.

The above discussion is restricted on $\theta(P)$. In practice, one is actually interested in $\theta(P_n)$. Since the argument of θ is no longer fixed, but is the random variable P_n, a further complication is introduced. The properties of P_n are unknown, and even if they were known, results that hold for a fixed P will not necessarily carry over to $\theta(P_n)$. Because there does not appear to be an immediate solution to the problem, it is worthwhile to consider the Method of Randomization as a practical alternative.

b) A Randomization Test

Fisher's Method of Randomization is a powerful tool that can be used to obtain a distribution-free test in any hypothesis-testing situation. It is particularly useful in situations, such as the present one, where computational difficulties, failure to meet model assumptions, or other problems arise in the application of customary tests. A randomization test is based on the idea that the observed arrangement of the data can be considered one of many possible arrangements, all of which are equally likely under the null hypothesis. Given a test statistic that is sensitive to the alternative hypothesis, the statistic is evaluated for each distinguishable arrangement of the data. The p value is the proportion of arrangements that result in a value that is at least as unusual as the one for the data as originally recorded. Unlike the parametric approach, where the p value is defined with respect to all possible samples, the p value in a randomization test is calculated with respect to all possible arrangements of the observed data. The test is distribution free, or nonparametric, since no distributional assumptions are required. A good description of Fisher's Method of Randomization can be found in Bradley (1968).

To apply the Method of Randomization to the period-testing problem of Sec. II, it is necessary to define an appropriate test statistic and to identify all distinguishable (with respect to the test statistic) permutations of the data that are equally likely under the null hypothesis of no periodicity. The most obvious choice for the test statistic is the one used in the parametric approach, namely the minimum (standardized) dispersion, $\theta(P_n)$. Assuming that there is no periodicity in the data, magnitude is independent of observation time. Consequently, the magnitudes $y_1, y_2, ..., y_n$, observed at the times $t_1, t_2, ..., t_n$, are just as likely to have occurred in any other order $y_{r(1)}, y_{r(2)}, ..., y_{r(n)}$, where $r(1), r(2), ..., r(n)$ is a random permutation of the subscripts, $1,2,...,n$. To test the hypothesis that there is no periodic component in the data, $\theta(P_n)$, evaluated for the original ordering of the magnitudes [which is denoted by $\theta(P_n; y_1, y_2, ..., y_n)$] is compared to the randomization distribution, obtained by evaluating $\theta(P_n)$ for each of the $n!$ equally likely permutations of the magnitudes. The p value is the proportion of permutations that give a value of $\theta(P_n)$ less than or equal to $\theta(P_n; y_1, y_2, ... y_n)$. If the null hypothesis is true, randomization is expected to have little effect on $\theta(P_n)$ and the p value should not be particularly small. On the other hand, if the alternative is true, the relationship between magnitude and time accounts for some fraction of the variation in magnitude, and the unexplained or random dispersion, measured by $\theta(P_n)$, is expected to be less than when there is no periodic structure, provided that P_n is close to the true period. Since randomization tends to disturb any such relationship between magnitude and time, this reduction cannot be realized unless the true data ordering is preserved. As a result, randomization is likely to increase $\theta(P_n)$, if there is a real periodic component, and the p value is expected to be small under the alternative.

Typically, $n!$ is so large that it is impractical to calculate $\theta(P_n)$ for every permutation of the data. It is possible, however, to estimate the p value by computing $\theta(P_n)$ for a random sample of m permutations. The proportion of permutations p that yield a value less than or equal to $\theta(P_n; y_1, y_2, ..., y_n)$ provides an estimate of the p value. When m is sufficiently large ($m > 100$), the standard error of the estimated p value can be approximated by the well-known formula $[\mathrm{p}(1-\mathrm{p})/m]^{1/2}$ (see Robbins and Van Ryzin 1975), which has a maximum value of $0.5\,m^{-1/2}$, and an approximate 95% confidence interval for the true p value can be written as $\mathrm{p} \pm 2[\mathrm{p}(1-\mathrm{p})/m]^{1/2}$. Ideally, each of the m random permutations of the data should be subjected to the same kind of period search as the original data. However, to minimize computation costs it may be necessary to restrict the period-search interval and to choose a slightly larger period step size. The performance of the test will, of course, depend on the effectiveness of the period searches, as well as the properties of the measure of dispersion θ and the properties of the sample, such as the spacing of the observation times and the signal-to-noise ratio.

The randomization test can also be used to test the null hypothesis that the period is equal to a specified value, against the alternative that the period is some other value, i.e., H_0: $P = P_0$ versus H_1: $P \neq P_0$, where P_0 is specified. In this case, the test is carried out as before, except that θ is evaluated at P_0 instead of P_n. The value $\theta(P_0; y_1, y_2, \ldots y_n)$, for the original data, is compared to $\theta(P_0)$ for random orderings of the magnitudes. This calculation is considerably easier than the previous one, where P_0 is not given, since no period searches are required.

IV. APPLICATIONS

Variable stars that exhibit complex pulsational behavior generally do so for reasons that are at present poorly understood. Such stars are in fact not rare, but comprise a significant fraction of all variables. In recent years, considerable work has been done deriving periods for, and attempting to explain, the observed radial pulsations of double-mode Cepheids and double-mode RR Lyrae stars (see Cox 1982). Other studies aimed at discovering more such double-mode stars are being made to build up a body of statistics for establishing the frequency of their occurrence (Barrell 1982; Nemec 1985a,b; Nemec, Norris, and Nemec 1985, in preparation; Clement *et al.* 1985, in preparation).

Double-mode variables pulsate simultaneously in the fundamental and the first-overtone modes in such a way that although the primary periods are sometimes very different, the ratios of the periods of two modes remain approximately constant for a given class of variables. For double-mode Cepheids, the primary periods range from 2 to 6 days, while the period ratios $P_1/P_0 = P$ (first overtone)/P(fundamental) tend to lie in a small range around 0.705; for double-mode RR Lyrae stars, the primary periods are in the range $P \sim 0^d35$–0^d42, and the period ratios are $P_1/P_0 \sim 0.746$; for double-mode dwarf Cepheids, the primary periods are $P \sim 0^d04$ to 0^d09, and the period ratios are $P_1/P_0 \sim 0.773$. By comparing the periods derived from observations with the periods predicted by theoretical models (Petersen 1973; Cox, Hodson, and Clancy 1983; Nemec 1985a,b) total masses can be determined. It is of obvious interest to know the extent to which errors in derived periods affect the mass determinations, and even more fundamentally, whether or not the low-amplitude oscillations are in fact real, and not simply chance occurrences. In some cases, there may be more than one secondary pulsation component. For example, the RR Lyrae star AC Andromeda is known to have three pulsation modes (Cox, King, and Hodson 1978). Failure to recognize the third component causes a serious error in the mass determination. Since mass determination using Petersen's (1973) method depends on primary and secondary periods, it is important to know that all the periods are, in fact, real.

In practice, several factors are taken into consideration when determining periods using the PDM methods. These include the depth of the minimum θ value, the appearance of characteristic sidelobes in the θ transform, the influence of outliers, and, in the case of secondary periods, a period ratio that is appropriate for the type of variable. However, the shape of the mean light curve and the dispersion of points about that curve is the single most important indication that the true period has been found. The randomization test is an objective way of assessing the quality of the light curves.

The randomization test, described in Sec. III*b*, was used to test the significance of the period estimates for the recently discovered double-mode RR Lyrae stars V41, 51, 54, and 61 in the globular cluster M15 (Cox, Hodson, and Clancy 1983; Nemec 1985b) and the star V112 in the Draco dwarf galaxy (Nemec 1985b). The stars were selected for study because of the relatively small amplitudes of the secondary pulsations. The photometry for the stars in M15 was from Sandage, Katem, and Sandage (1981), and the photometry of the Draco star was taken from Baade and Swope (1961). Stellingwerf's (1978) measure of dispersion θ_s was used to derive the periods, and the test statistic $\theta_s(P_n)$ was used in the randomization test. All p values were estimated from a sample of $m = 250$ random permutations, which guarantees standard errors no greater than 0.03.

Light curves corresponding to the estimated periods are shown in Fig. 30 of Nemec (1985a) and Fig. 16 of Nemec (1985b). Clearly, the selected stars have the lowest amplitudes and the secondary (fundamental mode) oscillations are good candidates for the randomization test. In all cases, the primary (first-overtone) period is obviously significant since the light curves show a clear dependence on the phase. However, the significance of the secondary periods is less apparent. The stars were deliberately chosen to illustrate the performance of the randomization test in a variety of situations.

For each of the five stars, the primary period gave a p value very near 0.0. Residual magnitudes, after prewhitening with the primary period using the method of Stobie (1970), were used to test the secondary periods. Table I summarizes the test results for the secondary oscillations. For each random permutation of the residuals, a period search of the interval 0^d52–0^d54 in the case of the four M15 stars, and of the interval 0^d56–0^d58 in the case of the Draco star, was conducted. A step size of 0^d0001 was used for all the stars. The step size and period-search intervals were chosen to approximate, as closely as possible, the search strategy used to find the periods for the original data. Since real features in the original θ_s transforms have full widths at half-minimum of typically 0^d0004, a step of 0^d0001 is sufficiently small to detect minima in the randomized data.

Column (3) of Table I gives the Stellingwerf $\theta_s(P_n; y_1, y_2, \ldots, y_n)$ value for the secondary period P_n listed in column (2), column (4) gives the p value for the randomization test, and column (5) gives the p value for Stellingwerf's F test. In the case of the relatively large-amplitude (0^m23) secondary for V61, according to the randomization test, there is an almost zero chance that the light curve could have occurred by chance ($0.00 < p$ value < 0.01, with 95% confidence). Although the Stellingwerf θ_s transform for the star V51 (Fig. 7 of Nemec 1985b) shows very little evidence of a secondary component, the randomization test p value suggests that a minimum θ_s value as small as 0.77 is unlikely to have occurred by chance. Two possible outliers are present in the light curve for V54 (Fig. 16 of Nemec 1985b). Randomization test p values were calculated with and without the two points, and in both cases the secondary period is only marginally significant. The θ_s transform for the V41 residuals (Fig. 6 of Nemec 1985b) provides weak evidence for the existence of a secondary component, but the rather large p value suggests that it may not be significant. Finally, V112 in

TABLE I. Results of the randomization tests.

Star	Period (days)	θ_s statistic	Randomization p value	F test p value	No. obs.
(1)	(2)	(3)	(4)	(5)	(6)
M15-V41	0ᵈ523964	0.85	0.188(0.025)	0.280	61
M15-V51	0ᵈ533703	0.77	0.012(0.010)	0.162	62
M15-V54	0ᵈ537645	0.83	0.072(0.016)	0.249	59
		0.82	0.104(0.019)	0.249	57
M15-V61	0ᵈ536100	0.56	0.000	0.019	60
Draco-V112	0ᵈ574420	0.74	0.000	0.126	64
White noise	0ᵈ538254	0.88	0.368(0.031)	0.318	63

Notes to TABLE I

(1) All randomization test p values are based on a sample of 250 random permutations of the data. The approximate standard errors (given in parentheses) were obtained using the formula $[p(1-p)/m]^{1/2}$. For the two cases where the p value is 0.000, the 95% confidence interval can be obtained using the binomial distribution, and is given by 0.0 to 0.01 (see Snedecor and Cochran 1973). (2) The p values for the Stellingwerf F test and the randomization test are one-sided. (3) The two rows for V54 are with (top) and without (bottom) the two outliers seen in Fig. 16 of Nemec (1985b). Since the two points are systematically overbright in the raw, pure fundamental and pure first-overtone light curves, it is probable that the photometric magnitudes are incorrect.

Draco is one of the least likely double-mode RR Lyrae stars among the ten discovered by Nemec (1985a), although the θ_s transform (Fig. 10 of Nemec 1985a) and the light curves strongly suggest that there is a secondary component.* The presence of a secondary period is supported by the randomization test, which gives a p value of 0.000 (95% confidence interval 0.0 to 0.01).

The last row of Table I was obtained by applying the randomization test to an artificial M15 data set containing no periodic component. This was done by generating a random normally distributed magnitude for each observation time. The artificial (white noise) magnitudes were then period searched in the same way as the real data, resulting in the period and θ_s statistic given in columns (2) and (3). As expected, the p value of 0.368 is too large to reject the null hypothesis that the data have no period in the range searched.

In general, the F test appears to be much more conservative than the randomization test, failing to reject the null hypothesis for all but one star (M51-V61). The discrepancies between the F test and the randomization test p values are not unexpected, however, since the F test has no obvious theoretical basis. In contrast, the randomization test is based on a sound statistical procedure and, judging by its performance for the small number of stars in Table I, is a useful and practical alternative to the F test.

*The secondary period for V112 in Table IV of Nemec (1985a) should be changed from 0ᵈ57422 to 0ᵈ57442.

REFERENCES

Baade, W., and Swope, H. H. (1961). Astron. J. **66**, 300.

Barrell, S. W. (1982). Mon. Not. R. Astron. Soc. **200**, 139.

Bradley, J. V. (1968). *Distribution-Free Statistical Tests* (Prentice-Hall, New York), p. 68.

Cox, A. N. (1982). In *Pulsations in Classical and Cataclysmic Variable Stars*, edited by J. P. Cox and C. J. Hansen (J.I.L.A., University of Colorado and National Bureau of Standards, Boulder, CO), p. 157.

Cox, A. N., King, D. S., and Hodson, S. W. (1978). Astrophys. J. **224**, 607.

Cox, A. N., Hodson, S. W., and Clancy, S. H. (1983). Astrophys. J. **266**, 94.

Davies, R. V. (1980). Algorithm AS 155, Appl. Stat. **29**, 323.

Heck, A., Manfroid, J., and Mersch, G. (1985). Astron. Astrophys. Suppl. **59**, 63.

Horman, M. (1980). Master's thesis, University of Liege.

Lafler, J., and Kinman, T. D. (1965). Astrophys. J. Suppl. **11**, 216.

Nemec, J. M. (1985a). Astron. J. **90**, 204.

Nemec, J. M. (1985b). Astron. J. **90**, 240.

Petersen, J. O. (1973). Astron. Astrophys. **27**, 89.

Renson, P. (1978). Astron. Astrophys. **63**, 125.

Robbins, H., and Van Ryzin, J. (1975). *Introduction to Statistics* (Science Research Associates, Chicago), p. 167.

Sandage, A., Katem, B., and Sandage, M. (1981). Astrophys. J. Suppl. **46**, 41.

Snedecor, G. W., and Cochran, W. G. (1973). *Statistical Methods*, 6th ed. (Iowa State University, Ames, IA).

Stellingwerf, R. F. (1978). Astrophys. J. **224**, 953.

Stobie, R. S. (1970). Observatory, **90**, 20.