Aplikace matematiky

Tomáš Cipra

Improvement of Fisher's test of periodicity

Aplikace matematiky, Vol. 28 (1983), No. 3, 186–193

Persistent URL: http://dml.cz/dmlcz/104025

IMPROVEMENT OF FISHER'S TEST OF PERIODICITY

TOMÁŠ CIPRA

(Received June 21, 1982)

Fisher's test of periodicity in time series and Siegel's version of this test for compound periodicities are investigated in the paper. An improvement increasing the power of the test is suggested and demonstrated by means of numerical simulations.

1. INTRODUCTION

Fisher's test (see [2]) is a suitable instrument for testing periodicity in time series. It can be applied in the following situation: Let a series $\{w_t\}$ arise from the model

$$w_t = \zeta_t + a_t\,, \quad t = 1, \ldots, n\,, \tag{1.1}$$

where $\{\zeta t\}$ represents the unobservable true values of the investigated series and $\{a_t\}$ is the white noise (i.e. a_t's are iid random variables with distribution $N(0, \sigma^2)$, $\sigma^2 > 0$). To make some statistical inference about the periodic behaviour of $\{\zeta_t\}$ let us consider the null hypothesis that there is no periodic activity

$$H_0 : \zeta_1 = \ldots = \zeta_n = 0\,. \tag{1.2}$$

In Fisher's test this hypothesis is tested against the alternative that $\{\zeta_t\}$ contains a significant periodicity. To simplify the calculations n is assumed to be odd, i.e.

$$n = 2s + 1 \tag{1.3}$$

(the case of n even is described in [3]). Let $I(\lambda)$ be the periodogram of $\{w_t\}$ defined as

$$I(f) = a^2(f) + b^2(f)\,, \quad -\tfrac{1}{2} \leqq f \leqq \tfrac{1}{2}\,, \tag{1.4}$$

where

$$a(f) = \sqrt{\frac{2}{n}} \sum_{t=1}^{n} w_t \cos 2\pi f t\,, \quad b(f) = \sqrt{\frac{2}{n}} \sum_{t=1}^{n} w_t \sin 2\pi f t\,. \tag{1.5}$$

Fisher's test uses the values of this periodogram for the frequencies

$$f_j = j/n\,, \quad j = 1, \ldots, s\,. \tag{1.6}$$

However, these values must be normalized to the form

$$Y_j = I(f_j) / \sum_{i=1}^{s} I(f_i) \tag{1.7}$$

to eliminate the effect of σ^2. The alternative hypothesis of Fisher's test corresponds to the situation that the maximum value among $I(f_1), \ldots, I(f_s)$ is much greater than the other ones. Therefore this test is based on the statistic

$$W = \max_{1 \leqq j \leqq s} Y_j \tag{1.8}$$

and rejects H_0 when W exceeds the appropriate critical value. The critical values of Fisher's test are tabulated in [2], [4] or in [5]. They are calculated by means of the following distributional formula for W:

$$P_{H_0}(W > x) = s(1 - x)^{s-1} - \binom{s}{2}(1 - 2x)^{s-1} + \binom{s}{3}(1 - 3x)^{s-1} - \ldots, \tag{1.9}$$

where only such members $\binom{s}{k}(1 - kx)^{s-1}$ are considered on the right-hand side of (1.9) for which $(1 - kx)$ is positive.

When there is an activity at several frequencies in the alternative hypothesis the Siegel's extension of Fisher's test can be used (see [5]). It generally has a higher power for such compound periodicity. Its test statistic has the form

$$T_\lambda = \sum_{j=1}^{s} (Y_j - \lambda g_F)_+ , \tag{1.10}$$

where $(t)_+$ denotes $\max(t, 0)$, g_F is the corresponding critical value of Fisher's test and λ is a parameter chosen between 0 and 1. Fisher's test is the special case when $\lambda = 1$ and the critical area is taken as $T_1 > 0$. The distributional formula for T_λ can be derived and the critical values calculated again (see [5]). The most recommended value of λ is 0·6.

It can be easily shown (see e.g. [5]) that when the alternative hypothesis of Fisher's test has the form

$$\zeta_t = \cos(2\pi f_0 t) , \tag{1.11}$$

where $0 < f_0 < \frac{1}{2}$ is a constant, then the value $I(f_0)$ of the periodogram (1.4) is distinctly high (of order n) while the other values $I(f)$ are relatively small. Since the test statistic W uses only the values $I(f_1), \ldots, I(f_s)$ it frequently occurs that the maximum among $I(f_1), \ldots, I(f_s)$ is significantly smaller than the maximum value $I(f_0)$ of the whole periodogram (especially for smaller values of n because then the gaps between particular frequencies f_j are great). Simulation studies show that this phenomenon decreases the power of Fisher's test unpleasantly. It is proved in Section 2 of the paper that the distributional formula (1.9) holds for the case when such values of the periodogram are considered which are calculated for the frequencies in the middles of the intervals with boundary points f_j defined in (1.6). In the paper we therefore suggest to apply the both tests (for the previous frequencies f_j and for the new ones) at the same significance levels simultaneously. The significance level of this compound test is then equal at most to the double of the level of significance of the both tests which form it. Numerical simulations in Section 3 demonstrate that such procedure can increase the power of Fisher's test substantially.

2. MODIFICATION OF FISHER'S TEST

Let us delete the first observation w_1 in the series $\{w_t\}$ and replace the formula (1.4) by

$$I'(f) = c^2(f) + d^2(f), \quad -\tfrac{1}{2} \leqq f \leqq \tfrac{1}{2}, \tag{2.1}$$

where

$$c(f) = \sqrt{\frac{2}{n-1}} \sum_{t=2}^{n} w_t \cos 2\pi f(t-1), \tag{2.2}$$

$$d(f) = \sqrt{\frac{2}{n-1}} \sum_{t=2}^{n} w_t \sin 2\pi f(t-1).$$

Let us consider the values of (2.1) for the frequencies

$$f'_j = \frac{2j-1}{2(n-1)} = \frac{2j-1}{4s}, \quad j = 1, \ldots, s. \tag{2.3}$$

These frequencies represent the middles of the intervals with the boundary points $f_j = j/n$ if we accept the approximation $1/n \sim 1/(n-1)$. Now let W' be the statistic calculated in the same way as W but with $I'(f)$ and f'_j replacing $I(f)$ and f_j. We shall show that the distributional formula (1.9) holds also for the statistic W'.

It will be sufficient to show that

$$c(f'_1), \ldots, c(f'_s), \quad d(f'_1), \ldots, d(f'_s) \tag{2.4}$$

are iid variables with distribution $N(0, \sigma^2)$ under the null hypothesis H_0 in (1.2) because then Fisher's method of the proof of (1.9) can be further applied without any changes (see also [1]).

First we shall show that

$$\sum_{t=2}^{n} \cos 2\pi(f'_i \pm f'_j)(t-1) = \sum_{t=2}^{n} \sin 2\pi(f'_i \pm f'_j)(t-1) = 0 \tag{2.5}$$

with one exception only, namely

$$\sum_{t=2}^{n} \cos 2\pi(f_i' - f_i')(t-1) = n - 1 . \tag{2.6}$$

To this end we shall use the following general formulas:

$$\cos u + \cos 2u + \ldots + \cos (n-1) u = \frac{\sin (n - 1/2) u}{2 \sin u/2} - 1/2 , \tag{2.7}$$

$$\sin u + \sin 2u + \ldots + \sin (n-1) u = \frac{\cos u/2 - \cos (n - 1/2)u}{2 \sin u/2} \tag{2.8}$$

that hold for $u \neq 2k\pi$. We can write e.g.

$$\sum_{t=2}^{n} \cos 2\pi(f_i' \pm f_j')(t-1) = \frac{\sin 2\pi[(n-1) + 1/2](f_i' \pm f_j')}{2 \sin \pi(f_i' \pm f_j')} - 1/2 = 0$$

since

$$2\pi(n-1)(f_i' \pm f_j') = \pi[(2i-1) \pm (2j-1)]$$

so that

$$\sin 2\pi[(n-1) + \tfrac{1}{2}](f_i' \pm f_j') = \sin \pi(f_i' \pm f_j') .$$

Because

$$\sin xt \sin yt = [\cos (x-y) t - \cos (x+y) t]/2 ,$$

$$\cos xt \cos yt = [\cos (x-y) t + \cos (x+y) t]/2 ,$$

$$\sin xt \cos yt = [\sin (x-y) t + \sin (x+y) t]/2 ,$$

we obtain due to (2.5) and (2.6) that for arbitrary $i, j = 1, \ldots, m$

$$\sum_{t=2}^{n} \cos^2 2\pi f_j'(t-1) = \sum_{t=2}^{n} \sin^2 2\pi f_j'(t-1) = (n-1)/2 , \tag{2.9}$$

$$\sum_{t=2}^{n} \sin 2\pi f_i'(t-1) \cos 2\pi f_j'(t-1) = 0 , \tag{2.10}$$

and for $i \neq j$

$$\sum_{t=2}^{n} \cos 2\pi f_i'(t-1) \cos 2\pi f_j'(t-1) = \tag{2.11}$$

$$= \sum_{t=2}^{n} \sin 2\pi f_i'(t-1) \cos 2\pi f_j'(t-1) = 0 .$$

Therefore

$$\operatorname{var} c(f_j') = \mathrm{E} \frac{2}{n-1} \sum_{t=2}^{n} a_t \cos 2\pi f_j'(t-1) \sum_{s=2}^{n} a_s \cos 2\pi f_j'(s-1) =$$

$$= \frac{2\sigma^2}{n-1} \sum_{t=2}^{n} \cos^2 2\pi f_j'(t-1) = \sigma^2 .$$

In the same way it is possible to prove that var $d(f_j') = \sigma^2$ and that any two variables in (2.4) are uncorrelated (their normality and zero mean value are obvious).

Let us mention that if the original length n of the series $\{w_t\}$ is an even number then we need not delete any observation of the series in the modified test.

As to the practical application we shall use the classical Fisher's test described in Section 1 and the modified test described in this Section simultaneously. In other words, we shall reject the null hypothesis H_0 if at least one of these two tests rejects it. Moreover, if the significance levels of the both „simple“ tests are $\alpha/2$ then the significance level of the compound test will be at most α.

3. RESULTS OF SIMULATION STUDIES

Series of the type

$$w_t = A \cos 2\pi f t + a_t \tag{3.1}$$

with various lengths were generated for various amplitudes A and frequencies f. One hundred replications were performed for each choice of these parameters at the computer Prime 750 at the Department of Statistics of the University of Uppsala. The numbers of rejections of the classical Fisher's test and of its modification described in Section 2 were recorded during those 100 replications. Table 1 shows the results of the simulations for the significance levels $\alpha = 0{\cdot}01$ and $\alpha = 0{\cdot}05$. As to the notation, e.g. the symbol $w'(0{\cdot}05)$ denotes the number of rejections of H_0 in 100 replications for the modified Fisher's test at the significance level $\alpha = 0{\cdot}05$. These simulation results demonstrate that the classical Fisher's test can really have a small power if the true frequency f in (3.1) is close to the middles of the intervals with boundary points f_j defined in (1.6) (see e.g. $n = 21$, $A = 1{\cdot}5$, $f = 5{\cdot}6/21$ or $n = 51$, $A = 1$, $f = 10{\cdot}5/51$). For f close to some f_j the power of the classical Fisher's test is, of course, larger than the power of its modified version (see e.g. $n = 21$, $A = 1{\cdot}5$, $f = 5{\cdot}1/21$ or $n = 31$, $A = 1{\cdot}5$, $f = 7{\cdot}1/31$). The results of the simulations for the white noise $w_t = a_t$ (see $n = 21$, $A = 0$, $f = 0$) justifies to the correctness of the simulation procedure.

Table 2 compares the power of the compound test at the significance level at most 0·05 with the power of the classical Fisher's test at the same significance level. The fourth and fifth columns of this table present the numbers of rejections of H_0 for the classical and modified Fisher's tests at the significance level $\alpha = 0{\cdot}025$ (i.e. $w(0{\cdot}025)$ and $w'(0{\cdot}025)$). These tests together form the compound test whose results are reported in the sixth column (see $w_c(0{\cdot}05)$). If we compare these results with those of the classical Fisher's test at the significance level $\alpha = 0{\cdot}05$ in the last column of Table 2 we can conclude that the compound test is in average more powerful than the classical one. The differences in the cases when it is not so (e.g. $n = 31$, $A = 1$, $f = 4{\cdot}1$) are not too large. On the other hand, the power of the compound test is frequently more than twice larger than that of the classical test at the same significance level (see e.g. $n = 31$, $A = 1\cdot5$, $f = 3\cdot4$ or $n = 51$, $A = 1$, $f = 10\cdot5$).

Table 3 is analogous to Table 1 but the results for Siegel's test are reported in it. Series of the form

$$w_t = A_1 \cos 2\pi f_1 t + A_2 \cos 2\pi f_2 t + a_t \tag{3.2}$$

were generated for this purpose. E.g. the symbol $t'_{0.6}(0\cdot01)$ denotes the number of rejections of H_0 in 100 replications for the modified Siegel's test with $\lambda = 0\cdot6$ at the significance level $\alpha = 0\cdot01$. The modification of Siegel's test consists in the fact that the values $I'(f'_j)$, $j = 1, \ldots, s$ according to (2.1)–(2.3) are used for the construction of T_λ in (1.10) instead of the values $I(f_j)$ according to (1.4)–(1.6). The conclusions which can be drawn from Table 3 are similar to the previous ones.

The results for the compound Siegel's test which is constructed similarly as the compound Fisher's test are not reported in this paper.

Table 1. The number of rejections of H_0 in 100 replications for the classical and modified Fisher's tests ($w(0\cdot01)$ relates to the classical Fisher's test at the significance level $\alpha = 0\cdot01$, etc.)

n	A	f	$w(0\cdot01)$	$w'(0\cdot01)$	$w(0\cdot05)$	$w'(0\cdot05)$
11	1·5	2·5/11	0	9	2	28
11	1·5	2/11	17	6	38	18
21	0	0	1	0	4	4
21	1·5	2·7/21	24	60	48	83
21	1·5	5·6/21	9	50	18	73
21	1·5	5·1/21	63	16	85	39
21	1	2·5/21	0	19	7	36
21	1	5·7/21	9	20	21	35
31	1·5	7·6/31	26	91	52	98
31	1·5	7·4/31	18	47	44	70
31	1·5	7·1/31	91	42	98	67
31	1·5	8·2/31	73	22	90	47
51	1·5	5·3/51	87	91	98	98
51	1	5·3/51	41	40	65	60
51	1	10·5/51	15	66	31	88
51	0·75	10·7/51	16	38	35	54

Table 2. Comparison of powers of the compound and classical Fisher's tests

n	A	f	$w(0{\cdot}025)$	$w'(0{\cdot}025)$	$w_c(0{\cdot}05)$	$w(0{\cdot}05)$
21	2	3/21	100	47	100	100
21	1·5	3/21	76	22	77	85
21	2	3·5/21	12	91	93	25
31	1·5	3·4/31	23	77	33	82
31	1·5	3·9/31	93	58	95	95
31	1	4·1/31	45	11	49	59
31	1·5	7·6/31	38	89	90	48
31	1	7·6/31	15	39	40	21
31	1·5	8·2/31	85	28	86	88
31	1·5	10·5/31	20	51	61	27
51	1	5/51	85	24	86	89
51	1·5	7·6/51	83	100	100	89
51	1	10·1/51	85	46	85	90
51	1	10·5/51	23	77	82	35
51	1·5	15·5/51	71	97	100	79
51	1	15·5/51	25	54	61	31
51	1	18·4/51	36	34	54	47
51	1	18·6/51	38	63	67	52

Table 3. The number of rejections of H_0 in 100 replications for Siegel's test and its modification ($t'_{0,6}(0{\cdot}01)$ relates to the modified Siegel's test with $\lambda = 0{\cdot}6$ at the significance level $\alpha = 0{\cdot}01$, etc.)

n	A_1	f_1	A_2	f_2	$t_{0,6}(0{\cdot}01)$	$t'_{0,6}(0{\cdot}01)$	$t_{0,6}(0{\cdot}05)$	$t'_{0,6}(0{\cdot}05)$
21	1·5	2·5/21	1·5	6·5/21	0	16	2	52
21	1·5	2/21	1·5	6/21	23	2	77	23
21	1·5	2·7/21	1·5	6·4/21	0	6	5	28
31	1·5	5·6/31	1·5	19·7/31	3	26	24	65
31	1·5	5·1/31	1·5	20·2/31	20	3	53	30
31	1·5	5/31	1·5	19·6/31	27	6	74	19
51	1	5·5/51	1	15·5/51	9	54	27	87
51	1	5·6/51	1	18·7/51	20	83	56	95
51	1	5·6/51	1	19·2/51	41	70	73	88
101	0·75	10·5/101	0·75	22·6/101	44	98	70	100
101	0·75	10·1/101	0·75	22·2/101	91	35	98	55
101	0·5	10·1/101	0·75	22·6/101	40	73	65	88

References

[1] *J. Anděl:* Statistical analysis of time series. SNTL, Prague, 1976 (in Czech).

[2] *R. A. Fisher:* Tests of significance in harmonic analysis. Proc. Roy. Soc. A 125 (1929), 54—59.

[3] *R. A. Fisher:* The sampling distribution of some statistics obtained from non-linear equations. Annals of Eugenics 9 (1939), 238—249.

[4] *R. A. Fisher:* On the similarity of the distribution found for the test of significance in harmonic analysis and in Steven's problem in geometrical probability. Annals of Eugenics 10 (1940), 14—17.

[5] *A. F. Siegel:* Testing for periodicity in a time series. Journal of the American Statistical Association 75 (1980), 345—348.

Souhrn

ZLEPŠENÍ FISHEROVA TESTU PERIODICITY

TOMÁŠ CIPRA

V článku je vyšetřován Fisherův test periodicity v časových řadách a Siegelova verze tohoto testu pro složené periodicity. Je navrženo zlepšení tohoto testu, které zvětšuje jeho sílu, a je demonstrováno pomocí numerických simulací.

Author's address: RNDr. *Tomáš Cipra*, CSc., Matematicko-fyzikální fakulta UK, katedra pravděpodobnosti a matematické statistiky, Sokolovská 83, 186 00 Praha 8.