

文章编号: 1000-8055(2003)06-0713-04

非等间距相关系数 $AR(p)$ 序列预测方法

马小兵, 傅惠民

(北京航空航天大学 固体力学研究所, 北京 100083)

摘要: 针对非等间距时间序列预测中存在误差较大的问题, 本文建立了非等间距相关系数 $AR(p)$ 序列预测方法, 详细讨论了非等间距相关系数 $AR(1)$ 序列和 $AR(2)$ 序列的预测公式和误差估计。大量计算表明, 本文方法与通过插值将非等间距序列变换为等间距序列进行预测的传统方法相比, 具有更高的预测精度。

关键词: 航空、航天推进系统; 时间序列; 非等间距序列; 相关系数平稳序列; 预测; 非平稳随机过程

中图分类号: TB 301; O 211 **文献标识码:** A

Prediction Method for Correlation Coefficient $AR(p)$ Series with Unequally Spaced Data

MA Xiao-bing, FU Huimin

(Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract A new prediction method for correlation coefficient stationary series with unequally spaced data is established. The prediction formulas and error estimations for the correlation coefficient $AR(1)$ and $AR(2)$ series are discussed in detail. The traditional interpolation method, which transfers the unequally spaced data to equally spaced data, may cause a great error in prediction. The results of calculations show that the precision of present method is higher than the traditional interpolation method.

Key words: aerospace propulsion system; time series; unequally spaced series; correlation coefficient stationary series; prediction; non-stationary random processes

时间序列预测方法广泛应用于自动控制、性能测试和故障诊断等工程领域, 但是非等间距序列预测问题至今没有得到很好地解决, 传统方法通过插值将非等间距序列转换为等间距序列进行预测, 往往导致较大的误差。文献[1]建立了非等间距相关系数平稳序列分析方法, 相关系数平稳

序列是从非平稳序列中分离出来的一类工程上常见且便于研究的时间序列, 它的均值和方差都可随时间变化, 能够更好地描述实际中的时间序列^[2~5], 传统的平稳序列仅是它的一个特例。本文进一步提出非等间距相关系数 $AR(p)$ 序列预测方法, 与传统方法相比, 大大提高了预测精度。

收稿日期: 2003-09-22; 修订日期: 2003-10-25

基金项目: 国防科技预研项目(413200204)

作者简介: 马小兵(1978—), 男, 陕西铜川人, 北京航空航天大学博士生, 从事信号处理及小样本可靠性技术研究

1 非等间距相关系数 $AR(1)$ 序列 预测公式

设 $X_i = X(t_i)$, $i = 1, 2, \dots, n$ 是来自非等间距相关系数平稳过程 $X(t)$ 的样本函数的一组观测值。若 $X_1, X_2, \dots, X_n$ 为相关系数 $AR(p)$ 、 $MA(q)$ 或 $ARMA(p, q)$ 序列, 则 $Y_1, Y_2, \dots, Y_n$ 为相关函数 $AR(p)$ 、 $MA(q)$ 或 $ARMA(p, q)$ 序列^[2], 其中 $Y(t) = [X(t) - \mu(a, t)]/I(b, t)$, $E[X(t)] = \mu(a, t)$, $\text{Var}[X(t)] = \sigma^2 I^2(b, t)$ 。当 $t_1, t_2, \dots, t_n$ 为非等间距时, 则称 $X = (X_1, X_2, \dots, X_n)$ 为非等间距相关系数平稳序列。

对于非等间距相关系数 $AR(1)$ 序列, 可根据 k 时刻的观测值 X_k 对 $k+1$ 时刻观测值 X_{k+1} 进行最小方差预测(使预测值的均方误差最小):

$$X_{k+1}|k = E(X_{k+1} | X_k) \quad (1)$$

即:

$$X_{k+1}|k = \mu(a, t_{k+1}) + e^{\alpha \tau_{k+1}} I(b, t_{k+1}) y_k \quad (2)$$

式中 $\tau_{k+1} = t_{k+1} - t_k$, $y_k = [X(t_k) - \mu(a, t_k)]/I(b, t_k)$, a, b, α, σ 由文献[1]求得。预测的均方误差为:

$$MSE(X_{k+1}|k) = \sigma^2 I^2(b, t_{k+1}) (1 - e^{2\alpha \tau_{k+1}}) \quad (3)$$

2 非等间距相关系数 $AR(2)$ 序列 预测公式

对于非等间距相关系数 $AR(2)$ 序列, 则可根据 k 时刻及 $k-1$ 时刻的观测值 X_k, X_{k-1} 对 $k+1$ 时刻观测值 X_{k+1} 进行最小方差预测:

$$X_{k+1}|k = E(X_{k+1} | X_k, X_{k-1}) \quad (4)$$

即:

$$X_{k+1}|k = \mu(a, t_{k+1}) + I(b, t_{k+1}) [\eta_1(\tau_{k+1}) y_k + \eta_2(\tau_{k+1}) y_{k-1}] \quad (5)$$

其中 $\tau_{k+1} = (\tau_{(k+1)1}, \tau_{(k+1)2})'$, $\tau_{(k+1)1} = t_{k+1} - t_k$, $\tau_{(k+1)2} = t_{k+1} - t_{k-1}$ 。且:

$$\eta_1(\tau_{k+1}) = \frac{\rho_{(k+1)1} - \rho_{(k+1)2} \rho_{k1}}{1 - \rho_{k1}^2} \quad (6)$$

$$\eta_2(\tau_{k+1}) = \frac{\rho_{(k+1)2} - \rho_{(k+1)1} \rho_{k1}}{1 - \rho_{k1}^2} \quad (7)$$

式中:

$$\rho_{(k+1)i} = \frac{1}{\alpha_i - \alpha_i} (\alpha_i e^{\alpha_1 \tau_{(k+1)i}} - \alpha_i e^{\alpha_2 \tau_{(k+1)i}}) \quad (8)$$

$$\rho_{k1} = \frac{1}{\alpha_i - \alpha_i} (\alpha_i e^{\alpha_1 \tau_{k1}} - \alpha_i e^{\alpha_2 \tau_{k1}}) \quad (9)$$

预测的均方误差为:

$$MSE(X_{k+1}|k) = \sigma^2 I^2(b, t_{k+1}) \Phi(\alpha, \tau_{k+1}) \quad (10)$$

式中 $a, b, \alpha = (\alpha_1, \alpha_2)'$; σ 由文献[1]求得。 $\Phi(\alpha, \tau_{k+1})$ 由下式给出:

$$\Phi(\alpha, \tau_{k+1}) = \frac{1 - \rho_{(k+1)1}^2 - \rho_{(k+1)2}^2 - \rho_{k1}^2 + 2\rho_{(k+1)1}\rho_{(k+1)2}\rho_{k1}}{1 - \rho_{k1}^2} \quad (11)$$

3 非等间距相关系数 $AR(p)$ 序列 预测公式

首先根据文献[1]方法求得非等间距相关系数平稳序列模型参数 $a = (a_0, a_1, \dots, a_r)'$, $b = (b_0, b_1, \dots, b_s)'$, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_p)$ 和 σ , 然后按如下方法进行预测。

对于非等间距相关系数平稳序列 $X = (X_1, X_2, \dots, X_n)$, 根据前 p 个观测值 $X_{k-p+1}, X_{k-p+2}, \dots, X_k$ 对 $k+1$ 时刻观测值 X_{k+1} 进行最小方差预测:

$$X_{k+1}|k = E(X_{k+1} | X_k, X_{k-1}, \dots, X_{k-p+1}) \quad (12)$$

X_{k+1} 的预测值 $X_{k+1}|k$ 及预测均方误差分别为:

$$X_{k+1}|k = \mu(a, t_{k+1}) + I(b, t_{k+1}) \sum_{i=1}^p \eta_i(\tau_{k+1}) y_{k+1-i} \quad (13)$$

$$MSE(X_{k+1}|k) = I^2(b, t_{k+1}) \sigma_e^2 \quad (14)$$

式中:

$$\sigma_e^2 = \sigma^2 \Phi(\alpha, \tau_{k+1}) \quad (15)$$

$$\Phi(\alpha, \tau_{k+1}) = 1 - 2 \sum_{i=1}^p \eta_i(\tau_{k+1}) \rho_{(k+1)i} + \sum_{i,s=1}^p \eta_i(\tau_{k+1}) \eta_s(\tau_{k+1}) \rho_{k+1-i} \rho_{k+1-s} \quad (16)$$

其中 $\tau_{k+1} = \{\tau_{(k+1)1}, \tau_{(k+1)2}, \dots, \tau_{(k+1)p}\}'$; $\tau_{(k+1)i} = t_{k+1} - t_{k+1-i}$ ($i = 1, 2, \dots, p$), $k \leq k+1 - m \in (l, s)$ 。自回归系数 $\eta_{k+1} = \{\eta_1(\tau_{k+1}), \eta_2(\tau_{k+1}), \dots, \eta_p(\tau_{k+1})\}$ 可由尤尔-瓦耳克方程求得:

$$\eta_{k+1} = V_{k+1}^{-1} \rho_{k+1} \quad (17)$$

式中:

$$V_{k+1} = \begin{bmatrix} 1 & \rho_{k1} & \rho_{k2} & \dots & \rho_{k(p-1)} \\ \rho_{k1} & 1 & \rho_{(k-1)1} & \dots & \rho_{(k-1)(p-2)} \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{k(p-1)} & \rho_{(k-1)(p-2)} & \rho_{(k-2)(p-3)} & \dots & 1 \end{bmatrix} \quad (18)$$

向量 $\rho_{k+1} = \{\rho_{(k+1)1}, \rho_{(k+1)2}, \dots, \rho_{(k+1)p}\}'$ 。相关系数函数 ρ_{uv} 由下式给出^[2,6]:

$$\rho_{uv} = \frac{\sum_{i=1}^p \beta_i e^{\alpha_i \tau_{uv}}}{\sum_{i=1}^p \beta_i} \tag{19}$$

其中:
$$\beta_i = \lambda_i \sum_{j=1}^p \frac{\lambda_j}{\alpha_i + \alpha_j} \tag{20}$$

$$\lambda_i = \frac{1}{\prod_{j=1, j \neq i}^p (\alpha_i - \alpha_j)} \tag{21}$$

4 应用举例

例1: 表1给出了一组由随机缺失("—"处即表示数据缺失)造成的非等间距相关系数平稳序列观测数据, 共90个观测值, 其一维分布函数为正态分布, 并且均值和方差都随观测时间呈线性变化。根据文献[1]方法对该序列进行分析, 可得非等间距相关系数AR (1) 模型如下:

表1 非等间距相关系数AR (1) 序列观测数据

Table 1 Measured data of correlation coefficient AR (1) series with unequally spaced data

序号 观测值		序号 观测值		序号 观测值		序号 观测值	
k	x _k	k	x _k	k	x _k	k	x _k
1	100.13	25	107.64	49	129.69	73	117.32
2	102.23	26	108.69	50	127.65	74	119.50
3	104.07	27	112.51	51	127.48	75	119.88
4	104.01	28	113.79	52	124.93	76	—
5	104.17	29	117.32	53	124.41	77	118.52
6	104.14	30	122.01	54	125.08	78	122.26
7	104.30	31	125.82	55	127.15	79	123.45
8	104.96	32	123.84	56	126.22	80	116.00
9	103.79	33	127.21	57	125.00	81	116.71
10	104.63	34	128.30	58	122.50	82	122.23
11	104.64	35	127.80	59	—	83	114.62
12	105.44	36	128.41	60	—	84	118.78
13	101.52	37	124.00	61	119.86	85	126.30
14	104.54	38	119.69	62	120.03	86	127.61
15	103.57	39	—	63	117.42	87	127.00
16	102.50	40	114.02	64	120.51	88	128.92
17	102.63	41	120.81	65	127.76	89	133.27
18	103.48	42	125.08	66	128.00	90	132.83
19	101.58	43	130.72	67	—	91	143.36
20	105.64	44	127.73	68	120.80	92	146.07
21	105.35	45	126.62	69	120.23	93	147.68
22	104.48	46	126.47	70	117.47	94	147.02
23	106.28	47	125.97	71	115.84	95	152.64
24	109.72	48	128.89	72	112.54		

$$\frac{x_k - 103.15 - 0.379k}{1 + 0.0289k} = e^{-0.084\tau_k} \times \frac{x_{k-1} - 103.15 - 0.379(k-1)}{1 + 0.0289(k-1)} + \epsilon \tag{22}$$

根据式(2)和式(22)进行一步预测, 步长分别取 $\tau=1, 2, 3$, 预测结果列于表2。从中可以看到预

表2 观测值与预测值对比结果

Table 2 Comparison between the measured value and the predicted value

序号 k	步长 τ	观测值 x_k	预测值 $x_{k 95}$	相对 误差/%
96	1	152.53	152.03	0.328
97	2	148.51	151.49	2.01
98	3	147.96	151.02	2.07

表3 非等间距相关系数AR (2) 序列观测数据

Table 3 Measured data of correlation coefficient AR (2) series with unequally spaced data

序号 观测值		序号 观测值		序号 观测值		序号 观测值	
k	x _k	k	x _k	k	x _k	k	x _k
1	196.48	26	231.08	51	225.15	76	270.43
2	198.50	27	234.64	52	223.84	77	268.73
3	203.94	28	240.92	53	226.50	78	270.81
4	203.94	29	245.03	54	228.92	79	267.22
5	200.95	30	248.20	55	233.63	80	256.11
6	194.62	31	249.40	56	238.94	81	259.81
7	189.30	32	257.52	57	235.38	82	260.17
8	192.38	33	257.99	58	229.12	83	—
9	197.90	34	254.12	59	238.80	84	270.95
10	202.72	35	252.94	60	—	85	273.35
11	213.16	36	—	61	—	86	279.08
12	218.41	37	242.13	62	251.14	87	283.17
13	224.76	38	247.00	63	255.46	88	286.38
14	227.78	39	248.99	64	263.80	89	288.42
15	232.28	40	249.72	65	—	90	292.53
16	236.97	41	247.00	66	—	91	290.87
17	240.21	42	243.43	67	251.87	92	290.25
18	242.35	43	240.21	68	248.37	93	—
19	241.49	44	241.48	69	256.35	94	298.96
20	239.42	45	—	70	257.70	95	304.80
21	241.18	46	228.21	71	258.56	96	311.37
22	237.65	47	216.32	72	260.40	97	312.55
23	229.29	48	—	73	263.40		
24	234.35	49	212.84	74	—		
25	230.87	50	219.34	75	276.33		

测值与观测值的相对误差较小。

例2: 表3给出了一组由随机缺失造成的非等间距相关系数平稳序列观测数据, 共87个观测值, 其一维分布函数为正态分布。其均值随观测时间线性变化, 方差为常数。根据文献[1]方法对该序列进行分析, 可得非等间距相关系数AR(2)模型如下:

$$x_k = 204.7 - 0.9k =$$
$$\eta_1(\tau_k)[x_{k-1} - 204.7 - 0.9(k-1)] +$$
$$\eta_2(\tau_k)[x_{k-2} - 204.7 - 0.9(k-2)] + \epsilon_k$$

(23)

表4 观测值与预测值对比结果

Table 4 Comparison between the measured value and the predicted value

序号 <i>k</i>	步长 <i>τ</i>	观测值 <i>x_k</i>	预测值 <i>x_{k 97}</i>	相 对 误差/%
98	1	318.47	312.17	1.98
99	2	319.80	311.15	2.70
100	3	312.89	310.43	0.786

根据式(5)和式(23)进行一步预测, 步长分别取 $\tau = 1, 2, 3$ 。预测结果列于表4。可以看到预测值与观测值的相对误差较小。式中 $\eta_1(\tau_k)$ 和 $\eta_2(\tau_k)$ 分别由式(6)和式(7)计算, 其中:

$$\rho_{k^*i} = 0.781[-0.109\exp(-1.39\tau_{k^*i})$$
$$+ 1.39\exp(-0.109\tau_{k^*i})]$$
$$k^* = k, k-1 \quad i = 1, 2 \quad (24)$$

5 结 论

(1) 建立了非等间距相关系数AR(*p*)序列的预测方法。大量计算表明, 与通过插值将非等间距序列转换为等间距序列进行预测的传统方法相比, 本文方法具有更高的预测精度。

(2) 详细讨论了工程上常见的非等间距相关系数AR(1)、AR(2)序列预测问题, 给出了相应的公式及其均方误差估计。

(3) 对于因数据缺失造成的非等间距序列, 当缺失的数据较少时, 插值方法尚可应用, 当缺失的数据稍多时, 则误差较大。对于非等间距采样得到的非等间距序列, 则插值方法不能应用, 必须采用本文方法进行预测。

(4) 对非等间距序列进行预测时, 预测步长通常应小于该序列中的最大间距, 建议采用该序列的平均间距。

参考文献

[1] 马小兵, 傅惠民 非等间距相关系数平稳序列分析方法[J] 航空动力学报, 2003, 18(4): 470- 476

[2] 傅惠民, 刘成瑞, 马小兵 相关系数平稳过程时频分析方法[J] 航空动力学报, 2002, 17(5): 505- 512

[3] 傅惠民, 马小兵 相关系数MA(*q*)序列分析方法[J] 机械强度, 2003, 25(2): 183- 186

[4] 傅惠民, 刘成瑞 相关系数ARMA(*p, q*)序列分析方法[J] 航空动力学报, 2003, 18(2): 161- 166

[5] 刘成瑞, 傅惠民 相关系数平稳序列滤波、预测和平滑[J] 机械强度, 2003, 25(5): 531- 536

[6] 杨叔子, 吴雅等 时间序列分析的工程应用(上册)[M] 武汉: 华中理工大学出版社, 1991