第4卷 第4期
1987年12月

控制理论与应用
CONTROL THEORY AND APPLICATIONS

Vol. 4, No. 4
Dec., 1987

利用相关熵估计ARMA模型的参数

徐南荣 陈为余

（南京工学院）

摘 要

本文提出估计 ARMA 模型参数的一种线性方法．根据信息熵的概念，利用相关熵作为衡量残差序列与白噪声序列逼近程度的一个定量指标．借助于相关熵的变化规律来确定拟合观测序列的 AR 模型的阶次．估计出 ARMA 模型参数的初值，并运用简单的线性方法实现 ARMA 模型参数的精估计．数字仿真结果说明此方法优于其他类似的线性估计方法．

一、引 言

国内外学者估计平稳时间序列的 ARMA 模型参数的方法大都采用迭代算法，求解一个非线性最优化问题。非线性最优化问题的求解需花费繁重的计算量。人们进而谋求采用线性方法估计 ARMA 模型的参数，其中 Mayne 和 Firoozan 的线性方法[1]较好地提高了参数估计的精度。然而，该方法对于所拟合的 AR 模型阶次，未提出确切的选择依据，所采用的滤波方法的计算亦比较复杂。

本文根据信息熵的概念，建立了衡量随机变量序列整体的相关性的指标，以相关熵作为残差序列与白噪声之间逼近程度的定量尺度，解决了自回归模型阶次的选定问题。另一方面，本文通过将有色残差白噪声化的线性方法，实现了ARMA模型参数的精估计。

二、模型参数估计方法

当各态历经的平稳时间序列$\{y_k\}$的一个实现为$y_1, y_2, \cdots, y_N$时，设它由 ARMA(p,q) 模型描述

$$y_k + a_1 y_{k-1} + \cdots + a_p y_{k-p} = w_k + c_1 w_{k-1} + \cdots + c_q w_{k-q}, \tag{1}$$

或

$$A(z^{-1}) y_k = C(z^{-1}) w_k, \tag{2}$$

其中$\{w_k\}$为白噪声，z^{-1}是单位后向时移算子，

$$A(z^{-1}) = 1 + a_1 z^{-1} + \cdots + a_p z^{-p},$$

$$C(z^{-1}) = 1 + c_1 z^{-1} + \cdots + c_q z^{-q}。$$

要求在结构参数 p, q 已选定情况下，估计该 ARMA(p,q)模型的参数θ，

$$\theta^T = [a_1, a_2, \cdots, a_p, c_1, c_2, \cdots, c_q]。$$

本文于1985年12月10日收到。1986年9月28日收到修改稿。

(1)相关熵及其计算

利用信息熵的概念，笔者在文献〔2〕中提出了用相关熵作为度量随机变量序列整体相关性的指标。设有平稳随机变量序列$\{\varepsilon_i, i=1,2,\cdots,N\}$。该随机变量序列的相关熵$H_r$定义为序列的各分量$\varepsilon_i$的熵$H(\varepsilon_i)$之和与序列的联合熵$H(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_N)$之差

$$H_r=\sum_{i=1}^{N}H(\varepsilon_i)-H(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_N), \tag{3}$$

相关熵H_r是非负的。当随机变量序列是不相关的（如白噪声序列），则理论上应有$H_r=0$；当随机变量序列是相关的，则H_r值的大小综合地反映了各分量之间的相关程度。

在随机变量序列具有正态分布的情况下，对于均值为零、协方差阵为R的N维随机变量序列$\{\varepsilon_i, i=1,2,\cdots,N\}$，有

$$H_r=\frac{N}{2}\log r_0-\frac{1}{2}\log|R|, \tag{4}$$

式中r_0为各随机变量的方差。

计算相关熵H_r的过程中，认为随机变量序列具有正态分布是接近于实际情况的。（4）式中理论协方差阵R则可以用它的估值$\hat{R}$来代替。

$$\hat{R}=[\hat{r}_{ij}] \qquad i,j=1,2,\cdots,N。 \tag{5}$$

$\hat{R}$的各个元素$\hat{r}_{ij}$就是自协方差系数的估值。

$$\hat{r}_{ij}=\hat{r}_{|i-j|}=\hat{r}_\tau=\frac{1}{N}\sum_{k=1}^{N-\tau}\varepsilon_k\varepsilon_{k+\tau} \qquad \tau=1,2,\cdots。 \tag{6}$$

随着τ的增大，自协方差系数的估值$\hat{r}_\tau$的总趋势将逐渐衰减。在一定量的时延$\tau=s$以后，$\hat{r}_\tau$实际上已衰减为零，即$\hat{r}_{ij}\approx 0$，当$|i-j|>s$。实际计算$|\hat{R}|$时，可取$\tau\leq s$，而s一般可取为等于N值的$\frac{1}{30}$到$\frac{1}{50}$。这样，矩阵$\hat{R}$就形成一个宽为$(2s+1)$的带状矩阵。对称正定阵$\hat{R}$可分解为

$$\hat{R}=BDB^T, \tag{7}$$

其中B为下半带状矩阵，$(s+1)$为半带宽，而s就是主对角线以下的次对角线数目。当$|i-j|>s$时，B阵的元素$b_{ij}=0$；D为对角线元素$d_{ii}=1/b_{ii}$的对角阵。由（7）式，

$$b_{ij}=r_{ij}-\sum_{t=t_1}^{j-1}b_{it}b_{jt}/b_{tt}, \qquad \begin{array}{l} i=1,2,\cdots,N \\ j=t_1,\ t_1+1,\cdots i \end{array} \tag{8}$$

其中 $t_1=\begin{cases}1 & i\leq s+1, \\ i-s & i>s+1。\end{cases}$

为了节省存贮单元，对（8）式中的元素的下标进行变换，将原对角半带状矩阵变换成长方形矩阵，再将长方形阵分成L个$(s+1)\times(s+1)$方阵，加上页面控制和下标控制，就可以将b_{ij}的计算压缩在$(s+1)\times(s+1)$个工作单元中进行。这时只要计算下半

带状阵B的每一行元素，将其最后一个元素（即b_{ii}）相乘，就得到

$$|\hat{R}|=\prod_{i=1}^{N} b_{ii} \qquad \text{或} \ \log|\hat{R}|=\sum_{i=1}^{N}\log b_{ii}. \tag{9}$$

由于该协方差阵$\hat{R}$在$i>s+1$时，每行的元素均相同，所以，对下半带状阵B的每行元素计算到一定行数以后，其每行的元素b_{ij}趋于相同。利用这一收敛性质，可以大大减少计算量。例如，当$N=1000$，按常规，在计算1000×1000维协方差阵R的行列式时，对它的下半带状阵B需要计算1000行，才能得到$|R|$值。利用它的收敛性质后，只计算$l=66$行b_{ii}值就收敛于b_{ll}。于是

$$\log|\hat{R}|=\sum_{i=1}^{l}\log b_{ii}+(N-l)\log b_{ll}. \tag{10}$$

从而降低了十几倍的计算量。

（2）等价AR模型阶次的确定

在适当的正则性条件下（连续的非零谱密度函数），有限样本的平稳时间序列可以由下面的具有无限阶次的AR(∞)模型来描述[3]：

$$G(z^{-1})y_k=w_k, \tag{11}$$

其中

$$G(z^{-1})=1+g_1z^{-1}+g_2z^{-2}+\cdots$$

$G(z^{-1})$为次数无限的多项式。所以，根据给定的时间序列$\{y_k, k=1,2,\cdots,N\}$，估计其ARMA(p,q)模型的参数，就是去寻找一个有限参数的传递函数$A(z^{-1})/C(z^{-1})$，逼近参数无限的多项式$G(z^{-1})$的问题。为此，首先寻找出能够充分逼近$G(z^{-1})$的次数有限的多项式$G_m(z^{-1})$，

$$G_m(z^{-1})=1+\hat{g}_1z^{-1}+\hat{g}_2z^{-2}+\cdots+\hat{g}_mz^{-m}. \tag{12}$$

利用由公式（4）—（10）所计算出的随机向量的相关熵，确定$G_m(z^{-1})$的次数就有了定量的依据。显然，随着多项式$G_m(z^{-1})$的次数m的增高，$G_m(z^{-1})$将愈加逼近$G(z^{-1})$，白噪声w_k的估值序列$\{\hat{w}_k=G_m(z^{-1})y_k\}$的相关熵就会愈来愈小，序列$\{\hat{w}_k\}$就愈接近于白噪声。但是，$m$值增高意味着估计模型参数的计算量增大。显然，根据$\{\hat{w}_k\}$的相关熵随$m$而变化的规律，确定阶次$m$的方法是，应使$\{\hat{w}_k\}$最大限度地逼近白噪声，同时又不过多地增高阶次$m$。即，令$m\geq(p+q)$，且使对应于$m$的$\{\hat{w}_k\}$的相关熵比对应于$(m-1)$的相关熵有明显的下降。

（3）ARMA(p,q)模型参数的估计

确定$G_m(z^{-1})$的次数m后，按最小二乘原理，令

$$J_1=\sum_{k=m+1}^{N}[G_m(z^{-1})y_k]^2$$

为最小，可确定$G_m(z^{-1})$的各个系数值$\hat{g}_1,\cdots\hat{g}_2,\cdots,\hat{g}_m$，并由下式确定估值$\hat{w}_k$，

$$\hat{w}_k = G_m(z^{-1})y_k = y_k + \hat{g}_1 y_{k-1} + \cdots + \hat{g}_m y_{k-m} \qquad k = m+1, \cdots, N, \tag{13}$$

用 $\hat{w}_k$,替代（2）式中的 w_k，令

$$J_2 = \sum_{k=m+q+1}^{N} [A(z^{-1})y_k - C(z^{-1})\hat{w}_k]^2$$

为最小，就得到参数向量 θ 的第一次估值 $\hat{\theta}^{(1)}$.

为了提高参数 θ 的估计精度，令

$$\varepsilon_k^{(1)} = \hat{A}^{(1)}(z^{-1})y_k - \hat{C}^{(1)}(z^{-1})\,\hat{w}_k \qquad k = m+q+1, \cdots, N, \tag{14}$$

一般情况下，$\{\varepsilon_k^{(1)}\}$是有色残差序列，用自回归模型将它白噪声化，即令

$$D^{(1)}(z^{-1})\varepsilon_k^{(1)} = w_k^{(1)} \tag{15}$$

其中 $$D^{(1)}(z^{-1}) = 1 + d_1^{(1)}z^{-1} + \cdots + d_r^{(1)}z^{-r}.$$

在实际计算中，$D^{(1)}(z^{-1})$的次数 r 可定为 2 或 3. 由于$\{\varepsilon_k^{(1)}\}$已知，当令

$$J_3 = \sum_{k=m+q+r+1}^{N} [D^{(1)}(z^{-1})\,\varepsilon_k^{(1)}]^2$$

为最小，可得到 $D^{(1)}(z^{-1})$的系数的估值$\hat{d}_1$，$\hat{d}_2, \cdots, \hat{d}_r$，将（15）式代入（14）式，并进行滤波，即令

$$\left.\begin{aligned} \tilde{y}_k^{(1)} &= D^{(1)}(z^{-1})y_k \\ \tilde{\varepsilon}_k^{(1)} &= D^{(1)}(z^{-1})\,\hat{w}_k \end{aligned}\right\} \qquad k = m+r+1, \cdots, N, \tag{16}$$

令 $$J_4 = \sum_{k=m+q+r+1}^{N} [A(z^{-1})\tilde{y}_k^{(1)} - C(z^{-1})\tilde{\varepsilon}_k^{(1)}]^2$$

为最小，得到参数向量 θ 的第二次估值 $\hat{\theta}^{(2)}$. $\hat{\theta}^{(2)}$将比$\hat{\theta}^{(1)}$更接近真值. 为进一步提高参数 θ 的估计精度，还可以按上述方法继续求得参数 θ 的估值 $\hat{\theta}^{(3)}$、$\hat{\theta}^{(4)}$…直到θ值收敛为止.

三、数字仿真例子和结果

为比较起见，采用文献[1]中所举的三个例子，这三个系统是

① $y_k = w_k + 0.5w_{k-1}$,

② $y_k - 0.5y_{k-1} = w_k + 0.5w_{k-1}$,

③ $y_k + 0.64y_{k-2} = w_k - 0.25w_{k-2}$.

等价的 AR 模型阶次 m 为不同值下，这三个系统的残差序列的相关熵值H_r见表1，所相应的$H_r \sim m$曲线见图1。表1中*号所在行的 m 值表示用本方法所选定的等价 AR 模型阶次。本文所选取的 m 值小于文献[1]所取的值〔文献〔1〕中 m的取值分别为5、5、6）

表 1 三个系统的残差的相关熵

模型阶次 m	残差的相关熵 H_r		
	系统 ①	系统 ②	系统 ③
1	9.78	21.04	221.72
2	4.56	6.56	15.24
3	4.38*	4.70	8.02
4	4.31	4.45*	5.16
5	4.30	4.42	5.01*
6	4.16	4.32	4.27
7	4.00	4.26	4.21
8	3.95	3.91	3.82

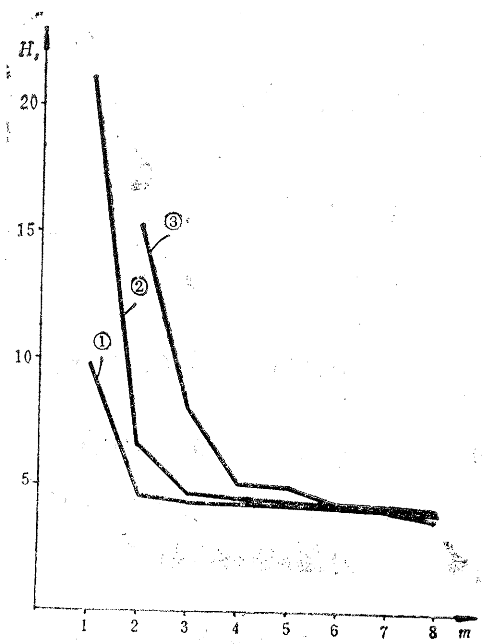

图 1

表2列出了利用本方法和 Mayne－Firoozan 方法[1]对这三个系统的模型参数进行估计的结果，除了系统③的模型中的一个参数 a_2(＝0.64)的估值稍高以外，本文方法所估计出的各个参数值 θ 均比 Mayne－Firoozan 方法所得到的参数估值更接近真值。

表 2 三个系统模型的参数估值

方法种类			本文方法				Mayne－Firoozan 方法			
系统① N＝500	真值 θ		0.500				0.500			
	估计值 $\hat{\theta}^{(i)}$	i＝1	0.501				0.551			
		2	0.496				0.438			
		3	0.492				0.451			
		4	0.489				0.449			
		5	0.487				0.449			
		6	0.485							
		7	0.484							
		8	0.484							
系统② N＝500	真值 θ		－0.500	0.500			－0.500	0.500		
	估计值 $\hat{\theta}^{(i)}$	i＝1	－0.489	0.481			－0.546	0.458		
		2	－0.507	0.463			－0.578	0.393		
		3	－0.506	0.458			－0.573	0.404		
		4	－0.508	0.454			－0.574	0.402		
		5	－0.508	0.454			－0.574	0.402		
系统③ N＝1000	真值 θ		0.000	0.640	0.000	－0.250	0.000	0.640	0.000	－0.250
	估计值 $\hat{\theta}^{(i)}$	i＝1	－0.020	0.675	－0.026	－0.237	－0.074	0.644	－0.045	－0.157
		2	－0.018	0.677	－0.021	－0.235	－0.077	0.641	－0.050	－0.163
		3	－0.018	0.677	－0.021	－0.234	－0.077	0.641	－0.051	－0.164
		4	－0.018	0.677	－0.021	－0.234	－0.077	0.641	－0.051	－0.164

四、结　论

本文提出的利用相关熵估计ARMA模型参数的线性方法，在算法上，计算工作量等方面均优于对参数的非线性函数迭代求解的各种方法。与其他类似的线性估计方法相比，本方法在选取等价AR模型的阶次上论据充分，并具有算法更为简便、估计精度高的优点。

参考文献

〔1〕 Mayne, D. Q. and Firoozan, F., Linear Identification of ARMA Processes, Automatica, **18**, 4, (1982), 461-466.

〔2〕 徐南荣，焦小澄，平稳时间序列模型的结构判定，南京工学院学报，1，(1984)，1-13.

〔3〕 Parzen, E., Some Recent Advances in Time Series Modelling, IEEE Trans. on Automatic Control, AC-**19**, 6, (1974), 723-730。

PARAMETER ESTIMATION FOR ARMA MODEL USING CORRELATION ENTROPY

Xu Nanrong, Chen Weiyu

(Nanjing Institute of Technology)

Abstract

In this paper a linear method for parameter estimation of ARMA model is proposed. According to the concepts of information entropy, tne correlation entropy is used as a quantitative performance index to measure the degree of closeness between the sequences of residues and white noise. The order of AR model fitted to observed data is selected according to the variation of the correlation entropy, and the initial values of the ARMA model parameters are estimated. These parameters are eventually identified by using simple linear method. Digital simulation for some examples shows that the method proposed is superior to other similar ones.