第 19 卷 第 3 期
1997 年 5 月

电子科学学刊
JOURNAL OF ELECTRONICS

Vol.19 No.3
May 1997

分段平稳随机过程的参数估计方法 [1]

王文华　　王宏禹

(大连理工大学电子工程系 大连 116023)

摘　要　P. M. Djuric 等人 (1992) 用贝叶斯法对非平稳随机信号化为分段平稳随机信号的问题作了研究，导出了一个关于分段数、各段自回归 (AR) 模型阶数和各段之间分界点的优化方程，但未给出具体的解法．由于此方程非常繁琐，故本文对其解法进行了研究，推导出其中一些递归关系，使求解问题变得简化，大大减少了计算量，同时对非平稳随机信号的分段问题作了进一步讨论．

关键词　非平稳随机信号，平稳随机信号，AR 模型，递归关系

中图号　TN911.23

1 前　言

实际中通常遇到非平稳随机信号，其中一类特殊的非平稳随机信号可视为准平稳或局部平稳随机信号，近年来引起了广泛注意 [1]．这类非平稳随机信号的统计特性在一些未知时刻突变，而在这些时刻之间保持平稳特性．这些未知时刻称为分界点，各平稳段可以用平稳随机信号参数模型来描述，因此可把这类非平稳随机信号称为分段平稳随机信号．

分析这类信号需要解决的问题是：(1) 分段的段数；(2) 各段之间的分界点；(3) 选择各段最佳的平稳随机信号参数模型．文献 [2] 中用贝叶斯法对此问题作了研究，导出了一个关于分段数、各段 AR 模型阶数和各段之间分界点的优化方程．此法不需要设置门限，可适用于许多应用领域，但由于优化方程的解法十分复杂，计算量很大，影响了其实际应用，故本文对优化方程的解法进行了研究，给出其中一些递归关系式，使求解问题简化，减少了计算量，同时对如何使用搜索法求解优化方程作了阐明．

2 非平稳随机信号的平稳分段

设非平稳随机信号是分段平稳的，观测所得的数据 $\boldsymbol{Y}=[y(1)y(2)\cdots y(N)]^T$，各分段 $S_1,S_2,\cdots,S_q$ 统计独立，各分段对应的模型分别为 $M^{(S_1)},M^{(S_2)},\cdots,M^{(S_q)}$，这里 $M^{(S_m)}(m=1,2,\cdots,q)$ 是模型阶数为 p_m 的 AR 模型，即服从下列递归关系：

$$y[n]=-a_1y(n-1)-a_2y(n-2)-\cdots-a_{p_m}y[n-p_m]+e(n)$$

式中 $e(n)$ 是零均值．方差为 σ_m^2 的正态白噪声，分界点为 $j_1,j_2,\cdots j_{q-1}$，$j_1<j_2<\cdots<j_{q-1}$。设最大可能的段数为 Q，$Q\geq q$。各段 AR 模型的最大阶数为 $P(P\geq p_m,m=1,2,\cdots,q)$。现要求解的是 q，$J_{q-1}=[j_1,j_2,\cdots,j_{q-1}]^T$ 和 $P_q=[p_1,p_2,\cdots,p_q]^T$。

[1] 1995-07-24 收到，1996-06-05 定稿
国家自然科学基金资助项目

令 H_k 为有 k 个分段的假设，在此假设之下，由贝叶斯定理，后验概率 $P(H_k, J_{k-1}, P_k/Y)$ 为

$$P(H_k, \boldsymbol{J}_{k-1}, \boldsymbol{P}_k/\boldsymbol{Y}) = f(\boldsymbol{Y}/H_k, \boldsymbol{J}_{k-1}, \boldsymbol{P}_k) \cdot P(H_k, \boldsymbol{J}_{k-1}, \boldsymbol{P}_k)/f(\boldsymbol{Y}), \tag{1}$$

式中 $f(\boldsymbol{Y})$ 是数据的边缘概率密度，$P(H_k, \boldsymbol{J}_{k-1}, \boldsymbol{P}_k)$ 是先验概率密度，$f(\boldsymbol{Y}/H_k, \boldsymbol{J}_{k-1}, \boldsymbol{P}_k)$ 是数据的条件概率密度．当假设为 H_k 时，若 $\boldsymbol{J}_{k-1}$ 和 $\boldsymbol{P}_k$ 独立，则

$$P(H_k, \boldsymbol{J}_{k-1}, \boldsymbol{P}_k) = P(\boldsymbol{J}_{k-1}/H_k) \cdot P(\boldsymbol{P}_k/H_k) \cdot P(H_k). \tag{2}$$

假设 H_k 的先验概率为

$$P(H_k) = 1/Q, \qquad k = 1, 2, \cdots, Q. \tag{3}$$

如果最短的段仅由两个采样构成，各段之间分界点的可能位置总数为

$$L_k = \Pi_{l=1}^{k-1}(N-k-l)/(k-1)!, \qquad 1 < k \le Q \le N/2. \tag{4}$$

如果它们的出现是等可能的，那么

$$P(\boldsymbol{J}_{k-1}/H_k) = 1/L_k = (k-1)!/\Pi_{l=1}^{k-1}(N-k-l), \tag{5}$$

$$P(\boldsymbol{P}_k/H_k) = 1/(P+1)^k. \tag{6}$$

由假设各段统计独立，有

$$f(\boldsymbol{Y}/H_k, \boldsymbol{J}_{k-1}, \boldsymbol{P}_k) = \Pi_{i=1}^{k} f(\boldsymbol{Y}_{j_{i-1}, j_i-1})/p_i), \tag{7}$$

式中 $j_0 = 1$，而

$$f(\boldsymbol{Y}_{j_{i-1}, j_i-1}/p_i) = \int_\theta f(\boldsymbol{Y}_{j_{i-1}, j_i-1}/\boldsymbol{\theta}_i, p_i) \cdot f(\boldsymbol{\theta}_i/p_i) \mathrm{d}\boldsymbol{\theta}_i, \tag{8}$$

式中 $\boldsymbol{Y}_{j_{i-1}, j_i-1} = [y(j_{i-1}), y(j_{i-1}+1), \cdots, y(j_i-1)]^T$，$\boldsymbol{\theta}_i = [\sigma_i^2, a_1^{(i)}, a_2^{(i)}, \cdots, a_{p_i}^{(i)}]^T$ 是第 i 段 AR 模型参数组，$f(\boldsymbol{\theta}_i/p_i)$ 是 AR 模型参数的先验概率密度 [3]．

将 (2)、(3)、(5)、(6) 式代入 (1) 式中，按最大后验估计最优准则，文献 [2] 给出了下列关于分段数、各段 AR 模型阶数和各段之间分界点估计值 $(\hat{q}, \hat{\boldsymbol{J}}_{q-1}, \hat{\boldsymbol{P}}_q)$ 的优化方程:

$$\begin{aligned}(\hat{q}, \hat{\boldsymbol{J}}_{q-1}, \hat{\boldsymbol{P}}_q) = \arg\min_{k, J_{k-1}, P_k} \Bigg\{ \Bigg[-\sum_{i=1}^{k} \sum_{l=j_{i-1}+1}^{j_i-1} \ln f[y(l)/(\boldsymbol{Y}_{j_{i-1}, l-1}, p_k)\Bigg] \\ + \ln(\Pi_{l=1}^{k-1}(N-k-l)/(k-1)!) + k \cdot \ln(P+1)\Bigg]/(N-k) \Bigg\},\end{aligned} \tag{9}$$

式中 $k \in (1, 2, \cdots, Q)$, arg 表示自变量．

3 分段平稳优化方程的求解

分段平稳优化方程 (9) 式虽然可用动态规划法求解，但是非常繁琐，其中困难主要在于求解

$$-\sum_{i=1}^{k}\sum_{l=j_{i-1}+1}^{j_i-1}\ln f(y(l)/\boldsymbol{Y}_{j_{i-1},l-1}p_k).$$

文献 [4] 给出了 $f(y(l)/\boldsymbol{Y}_{j_{i-1},l-1},p_k)$ 的推导方法和结果，其结果为

$$f[y(l)/\boldsymbol{Y}_{j_{i-1},l-1},p_k=0]=\frac{\Gamma[(l-j_{i-1}+1)/2]\cdot[(\boldsymbol{Y}^T_{j_{i-1},l-1}\boldsymbol{Y}_{j_{i-1},l-1})/2]^{(l-j_{i-1})/2}}{\sqrt{2\pi}\cdot\Gamma[(l-j_{i-1})/2]\cdot[(\boldsymbol{Y}^T_{j_{i-1},l}\boldsymbol{Y}_{j_{i-1},l}/2]^{(l-j_{i-1}+1)/2}} \tag{10a}$$

$$f[y(l)/(\boldsymbol{Y}_{j_{i-1},l-1},p_k)>0]=\frac{1}{\sqrt{2\pi}}\cdot\frac{|\boldsymbol{H}^T_{p_k}(l-2)\boldsymbol{H}_{p_k}(l-2)|^{1/2}}{|\boldsymbol{H}^T_{p_k}(l-1)\boldsymbol{H}_{p_k}(l-1)|^{1/2}}$$

$$\cdot\frac{\Gamma[(l-2p_k-j_{i-1}+1)/2]\cdot[(\boldsymbol{Y}^T_{p_k+j_{i-1},l-1}\boldsymbol{P}_{p_k}(l-2)\boldsymbol{Y}_{p_k+j_{i-1},l-1})/2]^{(l-2p_k-j_{i-1})/2}}{\Gamma[(l-2p_k-j_{i-1})/2]\cdot[(\boldsymbol{Y}^T_{p_k+j_{i-1},l}\boldsymbol{P}_{p_k}(l-1)\boldsymbol{Y}_{p_k+j_{i-1},l})/2]^{(l-2p_k-j_{i-1}+1)/2}}, \tag{10b}$$

式中

$$l\geq 2p_k+1,$$

$$\boldsymbol{P}_{p_k}(l)=\boldsymbol{I}-\boldsymbol{H}_{p_k}(l)(\boldsymbol{H}^T_{p_k}(l)\boldsymbol{H}_{p_k}(l))^{-1}\boldsymbol{H}^T_{p_k}(l),$$

$$\Gamma(x)=\int_0^\infty t^{(x-1)}\exp\{-t\}\mathrm{d}t,\qquad x>0,$$

$$\boldsymbol{H}_{p_k}(l)=\begin{bmatrix} y(p_k+j_{i-1}-1) & y(p_k+j_{i-1}-2) & \cdots & y(j_{i-1}) \\ y(p_k+j_{i-1}) & y(p_k+j_{i-1}-1) & \cdots & y(j_{i-1}+1) \\ \vdots & \vdots & \vdots & \\ y(l) & y(l-1) & \cdots & y(l-p_k+1)\end{bmatrix}$$

令

$$G_{p_k}=-\sum_{l=j_{i-1}+1}^{j_i-1}\ln f[y(l)/\boldsymbol{Y}_{j_{i-1},l-1},p_k], \tag{11}$$

对于不同的 p_k,(11) 式是不同的．当 p_k=0 时

$$G_0=-\sum_{l=j_{i-1}+1}^{j_i-1}\ln f[y(l)/\boldsymbol{Y}_{j_{i-1},l-1},p_k=0],$$

当 $p_k=1$ 时

$$G_1=-\sum_{l=j_{i-1}+1}^{j_{i-1}+2}\ln f[y(l)/\boldsymbol{Y}_{j_{i-1},l-1},p_k=0]-\sum_{l=j_{i-1}+3}^{j_i-1}\ln f[y(l)/\boldsymbol{Y}_{j_{i-1},l-1},p_k=1],\cdots$$

当 $p_k=m(m>0)$ 时

$$G_m=-\sum_{l=j_{i-1}+1}^{j_{i-1}+2}\ln f[y(l)/\boldsymbol{Y}_{j_{i-1},l-1},p_k=0]-\sum_{l=j_{i-1}+3}^{j_{i-1}+4}\ln f[y(l)/\boldsymbol{Y}_{j_{i-1},l-1},p_k=1],$$

$$-\cdots-\sum_{l=j_{i-1}+2m+1}^{j_i-1}\ln f[y(l)/\boldsymbol{Y}_{j_{i-1},l-1},p_k=m].$$

由此可知，(9) 式的求解是十分复杂的，其简化关键在于简化 (11) 式，对于第 i 段，令

$$S_{i0}=-\sum_{l=j_{i-1}+1}^{j_i-1}\ln f[y(l)/\boldsymbol{Y}_{j_{i-1},l-1},p_k=0].\tag{12}$$

将 (10a) 式代入 (12) 式中，经化简，有

$$\begin{aligned}S_{i0}=&(j_i-j_{i-1}-1)\ln\sqrt{2\pi}+\ln\Gamma(1/2)-ln\Gamma[(j_i-j_{i-1})/2]\\&+[(j_i-j_{i-1})/2]\ln[(\boldsymbol{Y}^T_{j_{i-1},j_i-1}\boldsymbol{Y}_{j_{i-1},j_i-1})/2]-(1/2)\ln(y_{j_{i-1}}/2).\end{aligned}\tag{13}$$

当分界点 j_i 向后位移一点时，有如下递归关系：

$$\begin{aligned}S'_{i0}=&S_{i0}+\ln\sqrt{2\pi}+\ln\Gamma[(j_i-j_{i-1})/2]-\ln\Gamma[(j_i-j_{i-1}+1)/2]\\&-[(j_i-j_{i-1})/2]\ln[(\boldsymbol{Y}^T_{j_{i-1},j_i-1}\boldsymbol{Y}_{j_{i-1},j_i-1})/2]+[(j_i-j_{i-1}+1)/2]\ln[(\boldsymbol{Y}^T_{j_{i-1},j_i}\boldsymbol{Y}_{j_{i-1},j_i})/2].\end{aligned}\tag{14}$$

令

$$\begin{aligned}S_{im}=&-\sum_{l=j_{i-1}+1}^{j_{i-1}+2}\ln f[y(l)/\boldsymbol{Y}_{j_{i-1},l-1},p_k=0]-\sum_{l=j_{i-1}+3}^{j_{i-1}+4}\ln f[y(l)/\boldsymbol{Y}_{j_{i-1},l-1},p_k=1]\\&-\cdots-\sum_{l=j_{i-1}+2m+1}^{j_i-1}\ln f[y(l)/\boldsymbol{Y}_{j_{i-1},l-1},p_k=m],\end{aligned}\tag{15}$$

$$\begin{aligned}A_{im}=&-\sum_{l=j_{i-1}+1}^{j_{i-1}+2}\ln f[y(l)/\boldsymbol{Y}_{j_{i-1},l-1},p_k=0]-\sum_{l=j_{i-1}+3}^{j_{i-1}+4}\ln f[y(l)/\boldsymbol{Y}_{j_{i-1},l-1},p_k=1]\\&-\cdots-\sum_{l=j_{i-1}+2(m-1)+1}^{j_{i-1}+2m}\ln f[y(l)/\boldsymbol{Y}_{j_{i-1},l-1},p_k=m-1],\end{aligned}\tag{16}$$

$$B_{im}=-\sum_{l=j_{i-1}+2m+1}^{j_i-1}\ln f[y(l)/\boldsymbol{Y}_{j_{i-1},l-1},p_k=m],\tag{17}$$

则

$$S_{im}=A_{im}+B_{im}.\tag{18}$$

将 (10b) 式代入 (17) 式，经化简，有

$$\begin{aligned}B_{im} = &[j_i - (j_{i-1} + 2m + 1)] \ln \sqrt{2\pi} + (1/2) \ln |\boldsymbol{H}_m^T(j_i - 2)\boldsymbol{H}_m(j_i - 2)| \\ &- (1/2) \ln |\boldsymbol{H}_m^T(j_{i-1} + 2m - 1)\boldsymbol{H}_m(j_{i-1} + 2m - 1)| + \ln \Gamma(1/2) - \ln \Gamma[(j_i - 2m - 1)/2] \\ &+ [(j_i - 2m - j_{i-1})/2] \ln[(\boldsymbol{Y}_{m+j_{i-1},j_i-1}^T \boldsymbol{P}_m(j_i - 2)\boldsymbol{Y}_{m+j_{i-1},j_i-1})/2] \\ &- (1/2) \ln[(\boldsymbol{Y}_{m+j_{i-1},j_{i-1}+2m}^T \boldsymbol{P}_m(j_{i-1} + 2m - 1)\boldsymbol{Y}_{m+j_{i-1},j_{i-1}+2m})/2].\end{aligned} \tag{19}$$

当分界点 j_i 向后位移一点时，有如下递归关系：

$$\begin{aligned}S'_{im} = &S_{im} + \ln \sqrt{2\pi} - (1/2) \ln |\boldsymbol{H}_m^T(j_i - 2)\boldsymbol{H}_m(j_i - 2)| \\ &+ (1/2) \ln |\boldsymbol{H}_m^T(j_i - 1)\boldsymbol{H}_m(j_i - 1)| + \ln \Gamma[(j_i - 2m - j_{i-1})/2] - \ln \Gamma[(j_i - 2m - j_{i-1} + 1)/2] \\ &- [(j_i - 2m - j_{i-1})/2] \ln[(\boldsymbol{Y}_{m+j_{i-1},j_i-1}^T \boldsymbol{P}_m(j_i - 2)\boldsymbol{Y}_{m+j_{i-1},j_i-1})/2] \\ &+ [(j_i - 2m - j_{i-1} + 1)/2] \ln[(\boldsymbol{Y}_{m+j_{i-1},j_i}^T \boldsymbol{P}_m(j_i - 1)\boldsymbol{Y}_{m+j_{i-1},j_i})/2].\end{aligned} \tag{20}$$

这样就得到了关于 (11) 式的简化式 (13) 、 (19) 式和递归关系 (14) 、 (20) 式，采用简化式和递归关系之后，将大大减少求解 (9) 式的计算量.

由于 (9) 式的特殊性和复杂性，较好的方法是用搜索法来求解，该法简述如下：

(1) 首先将整个数据序列视为一段 (k=1), 改变模型阶数，求 (9) 式的最小值；

(2) 将整个数据序列视为两段 (k=2), 改变两段的模型阶数 p_1, p_2 ($p_i = 0, 1, \cdots, P, i = 1, 2$) 和分界点 j_1 的位置，($j_1 = j_0 + 2, j_0 + 3, \cdots, N - 2$), 逐点进行搜索，分别计算 (9) 式，并与 (1) 中所得的最小值比较，得到一个最小值；

(3) 依次类推，让 $k = 3, 4, \cdots, Q$, 搜索方法同上，分别计算 (9) 式，经比较，得到一个最小值，这样就完成了搜索，得到的最小值所对应的分段数、各段模型阶数和分界点即为所求.

4 仿真和结果讨论

这里采用蒙特卡罗法进行仿真实验，所选 AR 模型如下： AR(1): $y(n) = 0.8y(n-1) + e(n)$, AR(2): $y(n) = 1.37y(n-1) - 0.56y(n-2) + e(n)$, AR(4): $y(n) = 2.7607y(n-1) - 3.8106y(n-2) + 2.6536y(n-3) - 0.9237y(n-4) + e(n)$. 式中 $e(n)$ 如前说所述 (方差为 1) . 仿真均在 486 微机上实现，每种实验进行 100 次，采样点数在 50 ~ 120 之间变化.

实验 1 所用数据序列包括两段 AR 过程： AR(2) 、 AR(1) . 分界点 $j_1 = 12$, 采样点数 N=50, 模型的最大阶数 P=3, 最大可能的分段数 Q=3 . 实验结果如表 1 所示.

实验 2 条件与实验 1 相同，但在开始搜索时让两个相邻的分界点间隔 10 个点，实验结果如表 2 所示.

实验 3 所用数据序列包括两段 AR 过程： AR(4) 、 AR(1) . 分界点 j_1=31, 采样点数 N=50, 模型的最大阶数 P=5, 最大可能的分段数 Q=3 . 搜索方法与实验 2 相同，实验结果如表 3 所示.

实验 4 所用数据序列包括两段 AR 过程： AR(2) 、 AR(1). 分界点 j_1=81, 采样点数 N=120, 模型的最大阶数 P=3, 最大可能的分段数 Q=3 . 搜索方法与实验 2 相同，实验结果如表 4 所示.

表 1a

分段数 k	1	2	3
实验 100 次出现次数	2	50	48

表 1b 出现次数

段	模型阶数			
	0	1	2	3
S_1	2	10	18	20
S_2	0	28	17	5

表 1c

点区间	1-5	6-8	9-15	16-25	26-35	36-45	46-50
出现次数	0	15	34	0	1	0	0

表 2a

分段数 k	1	2	3
实验 100 次出现次数	4	91	5

表 2b 出现次数

段	模型阶数			
	0	1	2	3
S_1	3	18	30	40
S_2	0	73	14	4

表 2c

点区间	11-15	16-25	26-35	36-45	46-50
出现次数	88	2	1	0	0

表 3a

分段数 k	1	2	3
实验 100 次出现次数	0	91	9

表 3b 出现次数

段	模型阶数					
	0	1	2	3	4	5
S_1	0	0	0	0	49	42
S_2	6	56	20	6	2	1

表 3c

点区间	11-15	16-25	26-35	36-45	46-50
出现次数	0	0	91	0	0

表 4a

分段数 k	1	2	3
实验 100 次出现次数	0	95	5

表 4b 出现次数

段	模型阶数			
	0	1	2	3
S_1	0	2	81	12
S_2	0	74	19	2

表 4c

点区间	11-15	16-25	26-35	36-45	46-55	56-65	66-75	76-85	86-95	96-105	106-115	116-120
出现次数	0	0	0	0	0	0	2	92	1	0	0	0

分析和讨论

(1) 采用了简化和递归关系式之后，大大减少了求解 (9) 式所需的计算量．对于实验 6, 微机进行一次仿真仅需 6.5min, 而在相同条件下，若直接进行求解，一次搜索将超过 5h;

(2) 数据序列越长，求解 (9) 式得到的关于分段数、各段 AR 模型阶数和分界点的正确率越高．搜索时让两个相邻的分界点至少间隔 10 个点，将会得到较好的实验结果．这是由于 (9) 式中 $f(y(l)/Y_{j_{i-1},l-1},p_k)$ 的影响所致，它实际上是 $y(l)$ 的预测概率密度，其中 $Y_{j_{i-1},l-1}$ 是过去数据向量，p_k 是这一段数据所假设的 AR 模型阶数．可以看出，当数据序列很短时，将会影响求解 (9) 式的正确率，而在搜索时让两个相邻的分界点至少间隔 10 个点，正是避免了这种影响，提高了求解的正确率，同时，又减少了计算量．

参 考 文 献

[1] Basseville M. Detecting changes in singles and system: A survey. Automatica,1988, 24: 309–326.

[2] Djuric P M, Kay S M, Faye Boudreaux-Bartels G. Segmentation of nonstationary signals. Proceedings of the IEEE ICASSP, San Francisco,USA: 1992, 5, 161–164.

[3] Djuric P M, Kay S M. Predictive probability as a criterion for model selection. Proceedings of the IEEE ICASSP, Aouquerque, New Mexico: 1990, 2415–2418.

[4] Djuric P M, Kay S M. Order selection of autoregressive models. IEEE Trans on SP, 1992, SP-40(11), 2829–2833.

A RESEARCH ON SEGMENTATION OF NONSTATIONARY STOCHASTIC PROCESS INTO PIECEWISE STATIONARY STOCHASTIC PROCESS

Wang Wenhua Wang Hongyu

(*Department of Electronic Engineering, Dalian University of Technology, Dalian* 116023)

Abstract P. M. Djuric, et al.(1992) researched on the segmentation of nonstationary stochastic process into piecewise stationary stochastic process by Bayesian theory, and gave a dynamic equation about the number of segments, their boundaries and AR model orders for each segment, but did not give details of solution for the equation. Because the solution for the equation is very complex, this paper investigates the solution, derives some recursive relations, simplifies the problem, saves computation time and goes further into the segmentation of nonstationary stochastic process into piecewise stationary stochastic process.

Key words Stochastic process, Stationary stochastic process, AR models, Recursive relation

王文华： 女，1969 年生，博士生，从事非平稳随机信号分析和处理的研究．

王宏禹： 男，1929 年生，教授，博士生导师，从事随机、时变和自适应数字信号处理理论、方法和应用研究．