

非定常スペクトル解析プログラム (NONSPA)

統計数理研究所 田 村 義 保

(1986年9月 受付)

1. はじめに

非定常スペクトル解析のためのプログラム NONSPA (NONstationary SPectrum Analysis by minimum ABIC procedure) の使用法を紹介する。非定常スペクトル解析のための諸手法については、田村(1984)のレビューがあり、NONSPA で用いている統計モデルの説明は Tamura (1987)に書かれている。ここでは、プログラムを使用する時の参考となる程度のモデルのアウトラインを書くに止める。

非定常時系列データ $\{x(1), x(2), \dots, x(N)\}$ が与えられ、また、この時サンプリング時間がデータの統計的性質の時間変化を検出できる程度に短くとられているものとする。このような条件の下で、測定区間を一定の長さを持った有限個のブロックにわけて、各ブロックごとに AR モデルをあてはめていくといったモデル化が可能になる。このような局所定常モデルは、Ozaki・Tong (1975), Kitagawa・Akaike (1978) によって論じられており、プログラムはそれぞれ TIMSAC-74 (1976), TIMSAC-78 (1979) に掲載されている。これらの方法はどちらも、異なったブロックの AR 係数は独立に求まるものとしている。AIC を用いて連続したブロックで同じ AR モデルをあてはめるべきか否かの判定をしているが、別々の AR モデルをあてはめる方が良いとなった場合、前のブロックのデータに含まれている統計情報を全て捨ててしまい、後のブロックのデータだけで、後のブロックの AR 係数を推定するようになっている。これに対して、NONSPA で用いている統計モデルでは、連続しているブロックの AR 係数間に後述する制約条件を課すことにより、AR 係数を推定したいブロックより前のブロックのデータの持つ統計情報を自然に取り込んで AR 係数を推定することが可能になっている。ブロック長を 1 にできない欠点はあるが、統計的性質がゆっくり変化する過程も、急激に変化する過程も取り扱うことができ、ある時点で統計的性質が全く変わるような場合には、自動的にその点を見出すことができるようになっている。

データ長 K の P 個のブロックに分け、各ブロックのデータにあてはめる AR モデルの次数を M とする。次の条件

$$N \geq N_0 = PK + M$$

を満足しなければならないことは言うまでもないのだが、NONSPA では、不適当なブロック数 P を入れた場合、この条件を満足する最大値を P として採るようになっている。計算の都合上、始めの $N - N_0$ 個の観測値を捨て、残りの N_0 個を $y(1-M), \dots, y(pK)$ に割付けることにする。

$$y(-M+i) = x(N - N_0 + i) \quad i=1, 2, \dots, N_0$$

この時 p 番目のブロック ($p=1, 2, \dots, P$) は

$$\{y((p-1)K+1), y((p-1)K+2), \dots, y(pK)\}$$

の K 個のデータからなる。第 0 ブロックは、次の M 個のデータからなっている。

$$\{y(-M+1), y(-M+2), \dots, y(0)\}$$

ここで、 M, K は条件

$$K \geq M+1+\alpha$$

を満足していなければならないことを注意しておく。ただし α はデータの平均値を 0 と考えてよい時は 0, そうでない時は 1 である。第 p ブロックにあてはめる AR モデルを

$$y(i) = \sum_{m=1}^M a_p(m)y(i-m) + a_p(0) + \varepsilon(i)$$

$$i = (p-1)K+1, (p-1)K+2, \dots, p * K$$

とする。データの平均値が 0 の時は $a_p(0)$ の項は不要である。ここで $\varepsilon(i)$ は平均 0, 分散 σ^2 の白色ガウス雑音である。隣合うブロックの AR 係数間に次の制約を課している。

$$c(p, i) = \sum_{m=1}^M (a_{p+1}(m) - a_p(m))y(i-m) + a_{p+1}(0) - a_p(0)$$

$$Ec(p, i)c(p', i') = (\sigma^2/\eta^2)\delta_{pp'}\delta_{ii'}$$

我々の制約条件は、直接にはひとつ前のブロックの AR 係数を用いた予測と注目しているブロックの AR 係数自身を用いた予測の差は小さくしなければならないということを表している。少し考えれば、スペクトルの時間変化の程度を制約していることと同等であることがわかると思う。パラメータ η の値によって時間変化の程度が変わる。このパラメータが赤池のベイズ模型に関する論文 (Akaike 1980) で “hyper parameter” と呼ばれているものにあたる。(以後、この小文で超パラメータと呼ぶ)

モデル、制約条件に含まれている未知パラメータ $a_p(m)$ ($m=0, 1, \dots, M, p=2, \dots, P$), σ^2 の推定、未知超パラメータ $a_1(m)$ ($m=0, 1, \dots, M$), η^2 の推定は、Akaike (1980) に従いベイズモデルの尤度が最大になるように行っている。(ABIC を最小にするものを推定値としている) η 以外は線形の連立方程式を用いて推定値を計算できるが、 η を推定するには非線形最適化法を用いなければならない。NONSPA では、計算を簡便にするためにグリッドサーチを用いている。

2. プログラム (NONSPA)

2.1 使用環境

プログラムは FORTRAN 77 を用いて書かれており、ステートメント数はコメント行も含めて 1752 である。実行に必要なメモリーは、後述の例では 2 メガバイトあれば十分であった。推定されたスペクトルを図示するために、Calcomp コンパチブルなコマンドで動作するグラフィック端末あるいは XY プロッタが使えることを想定している。これらの装置が使えないユーザや異なった作図言語の機械を用いているユーザは、プログラム中のコメントに従って作図部分を訂正する必要がある。

2.2 入力形式

次の変数の値を端末キーボードあるいはカードリーダー (実際はバッチファイル) から以下の

順に入力する必要がある。

[1] ITSS (並びに従った READ 文)

バッチジョブとするか (ITSS=0), TSS で実行するか (ITSS=1) を選択する。

[2] IAVE, ISRCH, JMODEL (並びに従った READ 文)

IAVE は前述のモデル $a_p(0)$ の項を含める時は 1, 含めない時は 0 にする。

ISRCH は超パラメータ η のグリッドサーチの方法を, JMODEL はグリッドサーチの範囲を規定し, 超パラメータの候補 EPSRN (I) は次の式で計算される。

$$\text{ISRCH}=1 \longrightarrow \text{EPSRN}(I) = \text{EPS0} * \text{DEL} * (I-1)$$

$$\text{ISRCH}=0 \longrightarrow \text{EPSRN}(I) = \text{EPS0} + \text{DEL} * (I-1)$$

$$I=1, 2, \dots, \text{JMODEL}$$

ただし, EPS0, DEL は 3 枚目のカードで入力される。

[3] EPS0, DEL (並びに従った READ 文)

[2] の説明にあるように, 超パラメータの初期値とサーチ間隔を決めるための変数である。

[4] DT, NDIV (並びに従った READ 文)

DT はデータのサンプリング時間である。(不明な時は, 1 としておけばよい。この時, スペクトルの最大周波数は 0.5 になる。)

NDIV はスペクトルを書く時の周波数軸の分割数である。(NDIV はサブルーチン PLTSPC の配列 XA, YA の寸法より 2 以上小さくしなければならない。実行例のように 100 と置くのがよいと思う。)

[5] IFIX, NMAX, NMIN (並びに従った READ 文)

IFIX は各ブロックのスペクトルを図示する時, y 軸の最大値, 最小値をすべてのスペクトルに共通した値を使うか, 各スペクトルごとに計算した最大値, 最小値を使うかの選択をする。

IFIX=1 の時は, NMAX, NMIN で与えた値 (単位 dB) がそれぞれ共通の最大値, 最小値となる。

IFIX=0 の時は, NMAX, NMIN はプログラム中では使用されないが, 適当な値をダミー値として与えておく必要がある。

[6] MT (フォーマット付き READ 文, 15)

データ本体を読み込むデバイス番号である。このデバイス番号をもつデータセット上には, データが次の形式で記録されていないといけない。

第 1 レコード: データ長 (フォーマット I5)

第 2 レコード: データ (フォーマット FORM)

以降 : (FORM については [8] で説明する)

[7] K, MKEEP, KEEP (フォーマット付き READ 文, 315)

K は各ブロックに含まれるデータ数である。

MKEEP は各ブロックのデータにあてはめる AR モデルの次数の最大値である。次数を 1 から MKEEP まで変えて, ABIC が最小になるモデルを選ぶ。

KEEP はブロック数の最大値で 296 行目を

DO 1000 IPP=2,KEEP

のようにプログラムを修正すると, $K * \text{KEEP}$ の約数となるブロック数 (IPP), ブロックのデータ数 (KKP) を選び出し, 異なるブロック数のモデルを比較することもできる。オリジナルプログラムでは,

DO 1000 IPP=KEEP, KEEP

となっており、ブロック数は固定している。

[8] FORM(I) (I=1, 2, ..., 20) (フォーマット付き READ 文, 20A4)

データが記録されているフォーマットを指定する。フォーマットは、両端の括弧を含めて、80文字以内で記述されていなければならない。

2.3 出力について

ラインプリンタ (バッチジョブ) あるいは端末画面 (TSS) に出力される内容を順に説明する。

- (1) ブロックに含まれるデータ数, AR モデルの最大次数, ブロック数, データ総数等が出力される。
- (2) 各ブロックに含まれるデータの値が出力される。
- (3) 次数 M の AR モデルで, 超パラメータを変えた時の ABIC と σ^2 の推定値が出力される。ABIC が, より低い次数のモデルでの ABIC の最小値より, 小さくなるような超パラメータの値が存在した場合は,

MINIMUM ABIC MODEL

以下の内容が, そうでない場合は,

NO MINIMUM EXISTS と

出力される。

- (4) 異なった次数の AR モデルの中から ABIC を最小とするモデルの各ブロックの AR 係数を出力する。
- (5) 標準偏差で規格化された原データ値 (+), 予測値 (P) が小数点第 1 位の精度で, ラインプリンタ (端末画面) に図示される。*印の意味を例を使って説明する。P が 0.6 のところがあり, P **** とあれば予測値 0.64 を, ** P とあれば予測値 0.58 を四捨五入して 0.6 になったことを意味する。
- (6) XY プロッタ (レーザプリンタ) あるいはグラフィック端末にスペクトルが 6 ブロック分ずつ出力される。

2.4 バッチによる実行例

統計数理研究所の HITAC M-280H (VOS3) における JCL の編成を示す。他の機種をご使用の方は適宜書き直して頂きたい。

//USERIDXB JOB PASSWORD, CLASS=C	(1)
//NONSPA EXEC FLGXYP	(2)
//FORT. SYSIN DD DSN=NONSPA. FORT, DISP=SHR	(3)
//GO. SYSIN DD *	(4)
0	[1]
0 __1__11	[2]
.5D-4 __10.0D00	[3]
1.D00 __100	[4]
1 __20__-20	[5]
__ __10	[6]
__ __20__ __10__ __25	[7]

```
(5D16.8) [8]
//GO. FT10F001 DD UNIT=DISK, DSN=LEC. DATA (GKM1), DISP=SHR (5)
// (6)
```

JCL の意味を説明する。

- (1) JOB の開始を意味する。JOB クラスは C である。
 - (2) カタログドプロシジャ FLGXYP の起動を意味する。統計数理研究所の M-280H システムで登録されており、FORT(コンパイル)、LKED(リンケージエディタ)、GO(ゴ-)、FOGINT(図形処理用のデータセットをレーザプリンタの出力様式に変換する)、FOGWTR(レーザプリンタへの出力準備) の 5 つのジョブステップからなっている。(大隅他, 1986)
 - (3) ソースプログラムの入っているデータセット名を指定する。
 - (4) データ入力の開始を意味する。
- [1] から [8] は、2.2 で示した変数の値を入力している。“_” は空白を意味している。
- (5) データの入っているデータセットを入力デバイス 10 番に割り当てる。(このデバイス番号を [6] で入力している。)
 - (6) JCL の終了を意味する。

2.5 TSS による実行例

計算を実行するのに必要な環境になっているものとする。統計数理研究所では、コマンド GALLOC を入力して、グラフィック出力が可能ないようにしておく必要がある。

```
E NONSPA. FORT (1)
ALLOC DD(FT10F001) DSN(LEC. DATA(GKM1)) SHR (2)
GRUN98 (3)
0 [1]
0 _1 _11 [2]
.5D-4 _10.0D00 [3]
1.D00 _100 [4]
1 _20 _-20 [5]
_ _ _10 [6]
_ _ _20 _ _ _10 _ _ _25 [7]
(5D16.8) [8]
Y (4)
```

各行の意味を説明する。

- (1) ソースプログラムのファイル NONSPA.FORT のエディットの開始を意味する。
- (2) 入力データセットをデバイス番号 10 に割り当てる。
- (3) プログラムの実行を開始する。GRUN98 は統計数理研究所固有のコマンドで、作図を行うのに必要なロードモジュールライブラリをリンクしてくれる。

[1] から [8] は 2.2 で示した変数の入力を行っている。“_” は空白であり、書かれている数だけ必要である。入力中に、並び違った READ 文を用いている場合はプロンプト “?” が計算機から送られて来る。確認のために、入力値の再表示も行われる。

- (4) 結果が出力される。最初にラインプリンタに出力されるのと同じ内容が出力される。続いてグラフィック出力が始まる。6 ブロック分ずつ表示され、次の画面に進むために “Y” を入力する必要がある。“PUSH CNTL+F7” と表示されるが、これは著者の使っている端末(SORD-

M68を用いた能動端末)でグラフィック画面のハードコピーをとるために必要なキー操作である。PC9801等,他の端末を使っている場合は,その端末に固有のキー操作をする必要がある。

3. おわりに

本稿では簡単な説明に止めたが, TIMSAC-84 (Akaike et al. (1985))にはソースプログラム中に使用方法の詳細な記述があり,またサンプル出力も掲載されている。本稿と併せて読んで頂ければ幸いである。

参 考 文 献

- Akaike, H. et al. (1976). TIMSAC-74 (A Time Series Analysis and Control Program Package), *Computer Science Monographs*, 5, 6, The Institute of Statistical Mathematics, Tokyo
- Akaike, H. et al. (1979). TIMSAC-78 (A Time Series Analysis and Control Program Package), *Computer Science Monographs*, 11, The Institute of Statistical Mathematics, Tokyo.
- Akaike, H. et al. (1985). TIMSAC-84 (A Time Series Analysis and Control Program Package), *Computer Science Monographs*, 22, 23, The Institute of Statistical Mathematics, Tokyo.
- Akaike, H. (1980). Likelihood and the Bayes procedure, *Bayesian Statistics* (eds. Bernardo J.S., Degroot M.H., Lindley D.V. and Smith A.F.M.)
- Kitagawa, G. and Akaike, H. (1978). A procedure for the modeling of non-stationary time series, *Ann. Inst. Statist. Math.*, 30, 351-363.
- 大隅 昇他 (1986). 電子計算機システムの手引 ——バッチ処理編——, 統計数理研究所.
- Ozaki, T. and Tong, H. (1975). On the fitting of non-stationary autoregressive models in time series analysis, *Proceeding of 8th Hawaii International Conference on System Sciences*, 224-226.
- 田村義保 (1984). 非定常スペクトル解析について, 統計数理研究所彙報, 32, 44-58.
- Tamura, Y.-H. (1987). An approach to the nonstationary process analysis, *Ann. Inst. Statist. Math.*, 39, 227-241.

NONSPA (NONstationary SPectrum Analysis by
minimum ABIC procedure)

Yoshiyasu-Hamada Tamura

(The Institute of Statistical Mathematics)

This paper describes NONSPA (NONstationary SPectrum Analysis by minimum ABIC procedure): a program for nonstationary spectrum analysis. Source program and a numerical example are printed in TIMSAC-84. For detail of this procedure, see Tamura ("An Approach to the Nonstationary Process Analysis", *Ann. Inst. Statist. Math.*, **39**, 1987).

In the paper (Tamura, 1978), a Bayesian approach to nonsationary process analysis is proposed. A nonstationary time series model (In our model a set of data is devided into some blocks with the same length.) with constraints on autoregressive coefficients of the successive blocks is developed. The smoothness parameters, which is concerned with our constraint, is determined by the minimum ABIC (Akaike Bayesian Information Criterion) which is attributed to Akaike ("Likelihood and the Bayes procedure", *Bayesian Statistics*, eds. Bernardo J.S. et al., 1980.). NONSPA is computer program for this nonstationary time series model.