国防科技大学学报
JOURNAL OF NATIONAL UNIVERSITY OF DEFENSE TECHNOLOGY
第18卷第2期 1996年6月 Vol. 18 No. 2

测量误差平稳性检验的数据预处理研究*

吴 翊 易东云
（国防科技大学系统工程与数学系 长沙 410073）

摘 要 在航天飞行器测量中，测量噪声是不可观测的。这就需要对含有误差的测量数据进行预处理以获得测量噪声的信息。本文从理论上对广泛应用的两种模型进行了系统性的研究，为处理方法的合理性提供了理论依据，并用仿真结果进行了验证。

关键词 数据处理，平稳性检验，差分方法，最小二乘法
分类号 O212.1

A Pre-process Study of Stationary Test for Measure Error

Wu Yu Yi Dongyun
(Department of Systems Engineering and Mathematics, NUDT, Changsha, 410073)

Abstract Measuring noise in aircraft tracking and measuring can't be observed and hence it must be pre-processed to obtain information about measuring noise. In this paper, we conduct a systematic study of the two models that are applied extensively and present theoretic bases for processing method.

Key words data processing, stationary test, difference method, LS method.

在航天飞行器数据的测量中，往往需要了解测量噪声的统计特性，如平稳性，但是一个最大的问题是噪声的数据是不可观测的，总是依附在某种真实信号之中。通常要对含有误差的信号数据进行预处理，其常用手段是通过差分方法或建立线性模型的方法来获得相应的统计量，以对噪声进行平稳性检验。问题的一般提法是：设测量数据 $y(t_1)$，$y(t_2)$，…，$y(t_n)$可表成如下形式

$$y(t) = f(t) + e(t) \tag{1}$$

其中 $f(t)$是测量数据中的确定性部分，$e(t)$是不可观测的测量噪声。需要检验 $e(t)$是否平稳。这里的难题是如何设法获得能够“替代”$e(t)$的数据。考虑两个模型，模型Ⅰ：$f(t)$为

* 湖南省自然科学基金资助项目
1995年5月19日收稿

阶数不超过某个 n 的 t 的未知多项式；模型Ⅱ：$f(t)=f(t,\theta)=\sum_{i=1}^{M}x_i(t)\theta_i,\theta$ 为未知参数。对模型Ⅰ，可采用差分方法进行预处理，此方法略嫌粗糙；对模型Ⅱ，则可用最小二乘方法进行更为精细的处理，但模型Ⅱ的建立需要对测量数据有更多的信息。对模型Ⅰ的差分处理方法在现有文献中只是作为一种实用方法，尚未见有较系统的理论研究。对模型Ⅱ的最小二乘处理方法，虽有一些文献([4])讨论，但通常条件都较强，在实用中难以进行验证。本文对上述两模型及各自的处理方法分别从理论上进行了研究。对模型Ⅰ的差分处理方法，从理论上刻划了处理结果的内涵性；对模型Ⅱ，在相当弱的条件下证明了最小二乘残差数据"替代"噪声数据的合理性，该条件在实用中通常是容易满足的。本文最后的仿真计算对上述结果进行了说明和验证。

1 模型Ⅰ的研究

如果测量数据 $y(t)$ 中确定因素项 $f(t)$ 是 t 的多项式，从理论上说，总可以通过等距差分将 $f(t)$ 从 $y(t)$ 中剔除。事实上，设

$$f(t)=a_0+a_1t+\cdots+a_nt^n \tag{2}$$

取等间隔 $h=t_k-t_{k-1}$ 的采样点，$t_1<t_2<\cdots<t_k<\cdots$，经过 $n+1$ 次差分后，取得

$$\nabla^{n+1}f(t_k)=0\qquad k=n+2,n+3,\cdots \tag{3}$$

于是由(1)

$$\nabla^{n+1}y(t_k)=\nabla^{n+1}e(t_k)\qquad k=n+2,n+3,\cdots \tag{4}$$

序列(4)便可用来作为待检数据，或者说可以看作是检验问题

$$H_0:y(t_k)\text{ 中的噪声 }e(t_k)\text{ 是平稳的} \tag{5}$$

的一组统计量。显然，当 H_0 成立时，$\nabla^{n+1}e(t_k)$ 是平稳的。因此可通过检验"$\nabla^{n+1}y(t_k)$ 是平稳的"这一结论是否成立来证实问题(5)。这即是差分法的原理。

这一方法过于粗糙，而且有几方面的问题。首先是差分的阶数不宜过高，过高则精度很差。一般认为，差分阶数 $n+1$ 最好不要超过 3。其次，用 $\{\nabla^{n+1}y(t_k)\}$ 作为检验数据，若其平稳性被否定，显然，(5)式的 H_0 也被否定；而当它的平稳性被接受时，并不能完全导致 H_0 以同样的概率被接受。进一步的困难还在于，由于假设 H_0 过于一般，无法得出一个什么分布，因而所涉及的概率实际上是无法计算的。当然，作为一种实用方法，还是可以做一些半定量的分析。

引理 1 差分方程 $\nabla^{n+1}y(t)=0$ 的解为一不超过 n 阶的多项式，它的系数由 $n+1$ 个初等条件唯一确定。

证明类似于高阶微分方程，由差分方程理论即知。

定理 2 设 $e(t)$ 为一个二阶矩有限过程，$e(t_1),e(t_2),\cdots,e(t_m)$ 是 $e(t)$ 的 m 个等距采样点。如果 $\{\nabla^{n+1}e(t_k),t=n+2,\cdots,m\}(m\gg n)$ 为具有有理谱密度的零均值平稳序列，则 $\{e(t_k)\}$ 或者是平稳的，或者仅含有次数不超过 n 的多项式(在采样点分辨的意义下)。

证明 考虑所有由二阶矩过程 $z(t)$ 的 $t_1<t_2<\cdots<t_m$ 处的采样点 $Z=\{z(t_k),k=1,2,\cdots,m\}$ 构成的 Hilbert 空间 $\mathscr{H}$，其中的内积定义为

$$(Z,Z')=\sum_{k=1}^{m}\mathrm{E}[z(t_k)-\mathrm{E}z(t_k)][z'(t_k)-\mathrm{E}z'(t_k)]$$

此处 $Z=\{z(t_k)\},Z'=\{z'(t_k)\}\in\mathscr{H}$。又设 $\mathscr{H}_0=\{Z\in\mathscr{H},Z$ 平稳，具有有理谱密度，$\mathrm{E}z(t_1)=0\}$，则 $\mathscr{H}_0$ 是 $\mathscr{H}$ 的闭子空间。

为证此事实，设 $Z,Z'\in\mathscr{H}$，则 $\mathrm{E}(z(t_1)+z'(t_1))=0$，又它们有 Wald 展式

$$z_k=\sum_{j=0}^{\infty}c_j\varepsilon_{k-j},z'_k=\sum_{j=0}^{\infty}c'_j\varepsilon_{k-j}\tag{6}$$

其中 $z_k=z(t_k)$，$z'_k=z'(t_k)$，ε_k 为标准白噪声。因此对于 $s>0$

$$\begin{aligned}\mathrm{E}[z_{k+s}+z'_{k+s})(z_k+z'_k)&=r_Z(s)+r_{Z'}(s)+\mathrm{E}z_{k+s}z'_k+\mathrm{E}z'_{k+s}z_k\\&=r_Z(s)+r_{Z'}(s)+\sum_{i=0}^{\infty}c_{i+s}c'_i+\sum_{i=0}^{\infty}c'_{i+s}c_i\end{aligned}\tag{7}$$

式(7)仅与 s 有关，即 $Z+Z'\in\mathscr{H}_0$。显然，$aZ\in\mathscr{H}_0$。最后，若 $Z^{(1)},\cdots,Z^{(n)},\cdots$ 是一列 $\mathscr{H}_0$ 中的元素，且依 $\mathscr{H}$ 收敛的极限为 Z，则

$$\mathrm{E}z_{k+s}z=\lim_n\mathrm{E}z^{(n)}_{k+s}z^{(n)}_k=\lim_n r_{z^{(n)}}(s)$$

即 Z 的 s 步相关函数仅与步长 s 有关，因此 $Z\in\mathscr{H}_0$。

于是记 $\mathscr{H}_0^{\perp}$ 为 $\mathscr{H}_0$ 的正交补，可得 $E=\{e(t_k)\}$ 的正交分解

$$E=U+Z\tag{8}$$

其中 $U=\{u(t_k)\}\in\mathscr{H}_0^{\perp},Z=\{z(t_k)\}\in\mathscr{H}_0$。如果 $u(t)$ 不是阶不超过 n 的多项式，则由引理 1，$\nabla^{n+1}u(t)\neq0$，从而 $\nabla^{n+1}U\neq0$

$$\nabla^{n+1}E=\nabla^{n+1}U+\nabla^{n+1}Z\notin\mathscr{H}_0$$

矛盾。

定理证毕。

定理 2 清楚地表明了差分数据平稳时原数据的特性。

2 模型Ⅱ的研究

当对确定性因素 $f(t)$ 有较多的了解时，可以用更精确的方法。例如 $f(t)$ 满足某个线性方程

$$f(t)=\beta_1C_1(t)+\beta_2C_2(t)+\cdots+\beta_rC_r(t)\tag{9}$$

则可以得到统计中的线性模型

$$y(t)=\beta_1C_1(t)+\cdots+\beta_rC_r(t)+e(t)\tag{10}$$

其向量形式为

$$Y=Z\beta+e\tag{11}$$

其中

$$Y=\begin{bmatrix}y_1\\y_2\\\vdots\\y_3\end{bmatrix},e=\begin{bmatrix}e_1\\e_2\\\vdots\\e_3\end{bmatrix},\beta=\begin{bmatrix}\beta_1\\\beta_2\\\vdots\\\beta_r\end{bmatrix},X=\begin{bmatrix}C_1(t_1)&\cdots&C_1(t_1)\\\vdots&&\vdots\\C_1(t_m)&\cdots&C_r(t_m)\end{bmatrix}$$

$y_k=y(t_k),e_k=e(t_k)$。线性模型(11)的最小二乘残差为

$$\hat{e} = (\hat{e}_1, \cdots, \hat{e}_m)^T = Y - \hat{Y} = (I - X(X^TX)^{-1}X^T)Y \tag{12}$$

可以证明，利用残差序列作为待检数据有其合理性。

引理 3 设$\{e_k\}$是 ARMA 序列，$\mathrm{E}e_k=0$，其相关函数

$$r_k = \mathrm{E}e_{t+k}e_t \qquad k = 0, \pm 1, \cdots$$

则$\sum_{k=-\infty}^{+\infty}|r_k|<\infty$。

引理 4 设 R_m 为 ARMA 序列 $e_1,\cdots,e_m$ 的相关阵，则存在一个与 m 无关的整数 a，使 R_m 的最大特征值 $\lambda^*\leqslant a$。

引理 5 设 A，B 为 m 阶非负定对称阵，C 为 m 阶矩阵，且 $A\geqslant B$，则 $\mathrm{tr}(C^TAC)\geqslant \mathrm{tr}(C^TBC)$

定理 6 对线性模型(11)，e 为 ARMA 序列，$\hat{e}$ 为其最小二乘残差(12)，则

$$\frac{1}{m}\sum_{j=1}^{m}\mathrm{E}(\hat{e}_j - e_j)^2 \to 0 \qquad (m\to\infty) \tag{13}$$

即残差向量与误差向量之差依各分量平均的意义下均方趋于零。

证明 $\mathrm{E}\|\hat{e}-e\|^2=\mathrm{E}\|X(\hat{\beta}-\beta)\|^2=\mathrm{tr}\mathrm{E}[X(\hat{\beta}-\beta)(\hat{\beta}-\beta)^TX^T]$

$$=\mathrm{tr}[X(X^TX)^{-1}X^TR_m(X^TX)^{-1}X^T] \tag{14}$$

取引理 4 中的 a 时，有

$$aI_m - R_m \geqslant 0 \tag{15}$$

事实上，取正交阵 P，使 $P^TR_mP=\mathrm{diag}(\lambda_1,\cdots,\lambda_m)$，则

$$P^T(aI_m - R_m)P = aI_m - \mathrm{diag}(\lambda_1,\cdots,\lambda_m) = \mathrm{diag}(a-\lambda_1,\cdots,a-\lambda_m) \geqslant 0$$

从而由引理 5 及式(13)知

$$\begin{aligned}\mathrm{E}\|\hat{e} - e\|^2 &\leqslant a\mathrm{tr}[X(X^TX)^{-1}X^TI_mX(X^TX)^{-1}X^T] = a\mathrm{tr}(X(X^TX)^{-1}X^T)\\ &= a\mathrm{tr}[(X^TX)^{-1}X^TX] = a\mathrm{tr}I_p = ap\end{aligned} \tag{16}$$

式(16)即是

$$\frac{1}{m}\sum_{j=1}^{m}E(\hat{e}_j - e_j)^2 \leqslant \frac{ap}{m} \to 0 \qquad (m\to\infty)$$

定理证毕。

定理 6 说明了当样本容量 m 取得充分大时，线性模型（11）的残差序列在某种意义下接近测量噪声序列。只要噪声是平稳的，利用这一点，就可以近似地以残差序列来替代测量噪声序列，作为待检序列。

3 仿真计算

设有如下模型

$$\begin{cases} y(t) = 3.4 - 5.6t + 0.8t^2 + e(t) \\ e(t) = 0.2e(t-1) - 0.3e(t-2) + 0.1e(t-3) \\ \qquad - 0.5e(t-4) + \varepsilon(t) + 0.9\varepsilon(t-1) - 0.7\varepsilon(t-2), \\ \varepsilon(t) \sim \mathcal{N}(0,1) \end{cases} \qquad t = 1,2,\cdots,m$$

即测量噪声设为平稳 ARMA(4,2)序列，表 1 给出了测量数据数 m=50,100,300,500 时

$e(t)$的真实数据及 $\Delta e(t)=e(t)-\hat{e}(t)$。表 1 结果说明，当样本容量 m 取得充分大时、$\Delta e(t)\approx 0$，$\hat{e}(t)$可作为 $e(t)$的"替代"数据。这正是定理 6 的结论。

表 1

	t	1	10	15	20	25	30	35	40
m	$e(t)$	−1.974	−1.119	−1.042	−0.438	−0.800	−2.789	2.127	0.166
50	$\Delta e(t)$	−0.616	−0.354	−0.147	0.050	0.102	0.144	0.130	0.062
100	$\Delta e(t)$	−0.253	−0.206	−0.163	−0.043	−0.087	−0.055	−0.026	−0.028
300	$\Delta e(t)$	−0.072	−0.065	−0.057	−0.017	0.043	−0.037	−0.011	−0.026
500	$\Delta e(t)$	0.016	0.013	0.009	0.006	0.003	0.000	−0.002	−0.005

参 考 文 献

1 张金明．平稳正态序列谱函数估计的收敛速度．应用概率统计，1988(2)：171～176

2 易东云．平稳增广混合回归模型参数估计的一种新方法及其应用．数学的实践与认识，1995(3)：1～4.

3 杨位钦，顾岚．时间序列分析与动态数据建模．北京理工大学，1988

4 陈希孺等．线性模型参数的估计理论．北京：科学出版社，1985

（责任编辑　潘　生）

（上接第 123 页）

3 Kang L S, Asynchronous parallel algorithm for mathematical physics problems, Acta Mathematica Scientia, 1983(3): 483～494

4 Bramble J H, Pasciak J E, A Domain Decomposition Technique for Stokes Problems, Applied Numer. Math., 1990(6): 251～261

5 Ma H P, Guo B Y. The Chebyshev Spectral Method for Burgers-like Equations, J. Comput. Math, 1988(6): 48～53

6 Xiong Y S. The errors estimation of the Chebyshev spectral-difference method for two-dimensional vorticity equation, Appl Math-JCU, 1994(9B): 153～167

7 Adams R A. Sobolev Space, New York Academic Press, 1975

8 Delves L M, Hall C A. An implicit matching procedure for global element calculations, J inst Math Appl, 1979(23): 223～234

椭圆型方程 Chebyshev 拟谱区域分解格式的等价形式

熊岳山　李晓梅

（国防科技大学计算机系　长沙　410073）

摘　要　本文对一维、二维椭圆型方程建立了 Chebyshev 拟谱区域分解格式，对这种拟谱区域分解格式给出了一种等价的广义变分形式。

关键词　等价性，Chebyshev 拟谱方法，区域分解，椭圆型方程。

分类号　O241.82

（责任编辑　潘　生）