UDC 519.28: 519.242

分散を考慮した近似多項式次数選定の一方法

A Method of Polynominal Order Determination for Curve-Fitting for Data Contaminated with Random Errors

浜　崎　襄　二*
Joji HAMASAKI

ランダム誤差を含んだ等間隔観測データ群に最小二乗法を適用して低次の多項式を推定し，これから最良推定値を求める問題において，多項式次数と推定値の分散，次数不足に伴う系統的偏差との関係を解析し，これらを考慮した次数選定方法について解説する．

1. 序

等間隔観測データ群に最小二乗法を適用してランダム誤差を除去した低次多項式を推定し，これから最良推定値を求める場合には，多項式の次数を予め決めておかなければならない．この際，もし多項式次数が低すぎた場合にはデータ群と推定多項式の間に系統偏差を生ずるから，この多項式は適切な推定値を与えるものではない．もし多項式次数が高すぎた場合には，決定すべき係数の数が増すため限定された数のデータから求められた推定値分散の期待値が増大し，また高次項の係数はランダム誤差の強い影響を受けるため，意味の薄いものとなる．一例としてデータ数より高い次数の多項式を決めてみてもそれは観測されたランダム誤差を忠実に表現するに過ぎないから，観測ランダム誤差除去の役に立たないことは明らかであろう．

このような理由から，測定すべき量の予め知られた性質として決まる微係数の大きさや，最良推定値の使い方によって決まる許容偏差の大きさに伴って，適切な多項式次数を指定する必要が起こる．本文ではこのような際の多項式次数の選定に際して払うべき考慮について述べてみよう．次節以下では上記の問題に選点直交多項式，最小二乗法，線形仮設検定を応用して解析と考察を進める．これら基本事項については紙面の都合上十分な説明が許されないので，参考文献を挙げるにとどめる．

2. 選点直交多項式と推定多項式

変数 t を $t=hx$ (h は間隔を与える定数) なる変数 x で置き換えたとき，xが整数値 0, 1, 2,……, n-1, n をとる等間隔系列で観測値が得られたものとしよう．測定すべき量の真値を関数 $y(x)$ とし，その x における観測値を y_x としよう．また，真値 y は $0\leqq x\leqq n$ の範囲内では x の p 次多項式で，十分精密に表現可能であり，観測に伴うランダム誤差 ε_x の平均値はゼロ，分散期待値は σ^2 であるとしよう．

さて，選点直交多項式 $P_{m,n}(x)$ は周知のように[1]

$$P_{m,n}(x)=\sum_{r=0}^{m}(-1)^r\binom{m}{r}\binom{m+r}{r}\frac{x(x-1)\cdots(x-r+1)}{n(n-1)\cdots(n-r+1)} \tag{1}$$

$0\leqq m\leqq n$　　n, m: 整数

と表わされ，これは x の m 次多項式である．したがって真値 y は $P_{m,n}(x)$ (ただし $m=0, 1, 2\cdots, p$-1, p) で展開できるはずであり，その展開係数を β_m とおくならば，観測値 y_x は次式で表わされなければならない．

$$y_x=\sum_{m=0}^{p}\beta_m P_{m,n}(x)+\varepsilon_x \tag{2}$$

$$(0\leqq p<n,\ \ 0\leqq x\leqq n)$$

$(n+1)$ 個の観測値 y_x に最小二乗法[2]を適用して常法どおり β_m の推定値 b_m を求めよう．正規方程式の係数を要素とする $(n+1)\times(n+1)$ の正方行列を $\{a_{l+1,m+1}\}$ とするならば，$P_{m,n}(x)$ の選点直交性により $a_{l+1,m+1}$ は

$$a_{l+1,m+1}=\sum_{x=0}^{n}P_{l,n}(x)P_{m,n}(x)=\delta_{l,m}S_{m,n} \tag{3}$$

ただし

$$S_{m,n}=\sum_{x=0}^{n}\{P_{m,n}(x)\}^2=\frac{(n+m+1)!(n-m)!}{(2m+1)(n!)^2} \tag{4}$$

$$\delta_{l,m}=\begin{cases}1\,(l=m\ \text{のとき})\\0\,(l\neq m\ \text{のとき})\end{cases}$$

したがって $\{a_{l+1,m+1}\}$ の逆行列 $\{a^{l+1,m+1}\}$ の要素 $a^{l+1,m+1}$ は

$$a^{l+1,m+1}=\delta_{l,m}\frac{1}{S_{m,n}} \tag{5}$$

また，正規方程式の右辺の$(n+1)$行 1 列の行列を $\{g_{l+1}\}$ とすれば

$$g_{l+1}=\sum_{x=0}^{n}P_{l,n}(x)y_x \tag{6}$$

したがって，正規方程式の解 b_m は

$$b_m=\sum_{l=0}^{p}a^{m+1,l+1}g_{l+1}=\frac{1}{S_{m,n}}\sum_{x=0}^{n}P_{m,n}(x)y_x \tag{7}$$

となり，次のよく知られた推定多項式 $\hat{y}(x)$ が得られる．

* 東京大学生産技術研究所第 3 部

$$\hat{y}(x)=\sum_{m=0}^{p} b_m P_{m,n}(x) \tag{8}$$

この式は最小二乗法の意味で最良推定多項式である．

3. 中央付近における推定値分散の期待値

前節で求まった推定多項式を用いて x の範囲 $[0, n]$ の中央付近における関数 y の値とその t に関する微係数の値を推定できる．本節ではこれらの推定値の分散の期待値を求めてみよう．

関数 y の変数 x に関する微係数を y', $y''\cdots y^{(p)}$ でまた，変数 t に関する微係数を $\dot{y}$, $\ddot{y}\cdots \overset{(p)}{y}$ で表わすことにすれば，

$$\overset{(p)}{y}=\frac{1}{h^p}y^{(p)} \tag{9}$$

範囲の中央 $x=\dfrac{n}{2}$ から微小量 $\Delta x=\dfrac{1}{h}\Delta t$ だけ離れた点 $x=\dfrac{n}{2}+\Delta x$ における選点直交多項式 $P_{m,n}(x)$ とその微係数の値はテーラー展開により

$$P_{m,n}\left(\frac{n}{2}+\Delta x\right)\cong P_{m,n}\left(\frac{n}{2}\right)+P_{m,n}'\left(\frac{n}{2}\right)\cdot\Delta x \tag{10}$$

$$P_{m,n}'\left(\frac{n}{2}+\Delta x\right)\cong P_{m,n}'\left(\frac{n}{2}\right)+P_{m,n}''\left(\frac{n}{2}\right)\cdot\Delta x \tag{11}$$

$P_{m,n}(x)$ の性質

$$P_{m,n}(x)=(-1)^m P_{m,n}(n-x) \tag{12}$$

より，$p+m$ が奇数となるような p に対しては，$x=\dfrac{n}{2}$ において

$$P_{m,n}^{(p)}\left(\frac{n}{2}\right)=0 \qquad p+m:\ \text{奇数} \tag{13}$$

となる．この関係を用いて $\hat{y}\left(\dfrac{n}{2}+\Delta x\right)$, $\hat{\dot{y}}\left(\dfrac{n}{2}+\Delta x\right)$ の値を求めると

$$\hat{y}\left(\frac{n}{2}+\Delta x\right)=\sum_{\substack{m=0,2\cdots\\ \text{even}}}^{p} b_m P_{m,n}\left(\frac{n}{2}\right)+(\Delta x)\times \\ \times\sum_{\substack{m=1,3\cdots\\ \text{odd}}}^{p} b_m P_{m,n}'\left(\frac{n}{2}\right)+\frac{(\Delta x)^2}{2}\sum_{\substack{m=0,2\cdots\\ \text{even}}}^{p} b_m P_{m,n}''\left(\frac{n}{2}\right) \\ +\cdots\cdots \tag{14}$$

$$\hat{\dot{y}}\left(\frac{n}{2}+\Delta x\right)=\frac{1}{h}\left\{\sum_{\substack{m=1,3\cdots\\ \text{odd}}}^{p} b_m P_{m,n}'\left(\frac{n}{2}\right)+(\Delta x)\times\right. \\ \times\sum_{\substack{m=0,2\cdots\\ \text{even}}}^{p} b_m P_{m,n}''\left(\frac{n}{2}\right)+\frac{(\Delta x)^2}{2}\sum_{\substack{m=1,3\cdots\\ \text{even}}}^{p} b_m P_{m,n}'''\left(\frac{n}{2}\right) \\ \left.+\cdots\cdots\right\} \tag{15}$$

また b_m の分散・共分散行列は $\sigma^2\{a^{l+1,m+1}\}$ となるから，$\hat{y}$, $\hat{\dot{y}}$ の分散・共分散の期待値は次のように求められる．

$$\text{var }\hat{y}\left(\frac{n}{2}+\Delta x\right)\cong\sigma^2\{Q_{p,n}+(\Delta x)^2(Q_{p,n}'+R_{p,n})\} \tag{16}$$

$$\text{var }\hat{\dot{y}}\left(\frac{n}{2}+\Delta x\right)\cong\frac{\sigma^2}{h^2}\{Q'_{p,n}+(\Delta x)^2(Q''_{p,n}+R_{p,n}')\} \tag{17}$$

$$\text{cov }\hat{y}\left(\frac{n}{2}+\Delta x\right)\hat{\dot{y}}\left(\frac{n}{2}+\Delta x\right)\cong\frac{\sigma^2}{h}(\Delta x)(Q_{p,n}'+R_{p,n}) \tag{18}$$

ただし

$$Q_{p,n}=\sum_{\substack{m=0,2\cdots\\ \text{even}}}^{p}\frac{1}{S_{m,n}}\left\{P_{m,n}\left(\frac{n}{2}\right)\right\}^2$$

$$Q_{p,n}'=\sum_{\substack{m=1,3\cdots\\ \text{odd}}}^{p}\frac{1}{S_{m,n}}\left\{P_{m,n}'\left(\frac{n}{2}\right)\right\}^2$$

$$Q_{p,n}''=\sum_{\substack{m=0,2\cdots\\ \text{even}}}^{p}\frac{1}{S_{m,n}}\left\{P_{m,n}''\left(\frac{n}{2}\right)\right\}^2$$

$$R_{p,n}=\sum_{\substack{m=0,2\cdots\\ \text{even}}}^{p}\frac{1}{S_{m,n}}P_{m,n}\left(\frac{n}{2}\right)P_{m,n}''\left(\frac{n}{2}\right)$$

$$R_{p,n}'=\sum_{\substack{m=1,3\cdots\\ \text{odd}}}^{p}\frac{1}{S_{m,n}}P_{m,n}'\left(\frac{n}{2}\right)P_{m,n}'''\left(\frac{n}{2}\right)$$

式(18)によれば，$\hat{y}$, $\hat{\dot{y}}$ の共分散は $\Delta x=0$ のときゼロとなるから，$\hat{y}$, $\hat{\dot{y}}$ を互に独立として処理したい場合には $\Delta x=0$ の値を最良推定値としてよいであろう．しかし，式(16)，(17)，(18) の係数の中 $Q_{p,n}$, $Q_{p,n}'$, $Q_{p,n}''$ は正項の和故に常に正であり，次数 p とともに単調に増大するが，$R_{p,n}$, $R_{p,n}'$, は負項の和故に常に負で単調減少するので，式(16)，(17) より $\hat{y}$, $\hat{\dot{y}}$ の分散は $\Delta x=0$ のとき極小になるとは限らない．しかし，分散の値は次数 p とともに単調増大するから，必要以上に高い次数の多項式を用いるのは推定値分散の増大を招くので不利となる．

4. 低すぎる次数の多項式使用に伴う分散の増大

前節においては推定多項式の次数が高いほど最良推定値の分散が大きくなることを定量的に説明した．本節では低すぎる次数の多項式を当てはめようとした場合におこる分散の増大を求めてみよう．

関数 y は前述の如く x の p 次式を用いるならば十分精密に表わされると仮定する．この場合，最小二乗法の適用にあたり x の $(p-q)$ 次式を用いたならば分散期待値はいかほど増大するだろうか．$(p-q)$ 次多項式を用いて最小二乗法を適用した場合の残差平方和とその自由度とをそれぞれ SSe', DFe' とするならば

$$SSe'=\sum_{x=0}^{n}y_x^2-\sum_{m=0}^{p-q}b_m g_{m+1} \\ =\sum_{x=0}^{n}y_x^2-\sum_{m=0}^{p-q}\frac{1}{S_{m,n}}\left(\sum_{x=0}^{n}P_{m,n}(x)y_x\right)^2 \tag{19}$$

$$DFe'=n-p+q \qquad p-q\geqq 0 \tag{20}$$

式(19) の y_x に式(2) を代入し，$P_{m,n}(x)$ の選点直交性と ave $\varepsilon_x^2=\sigma^2$ を用いて整理するならば，この場合の

分散期待値 σ'^2 は次式のように求められる.

$$\sigma'^2=\mathrm{ave}\frac{SSe'}{DFe'}$$
$$=\frac{1}{n-p+q}\mathrm{ave}\Big[\sum_{x=0}^{n}\Big(\sum_{m=0}^{p}\beta_m P_{m,n}(x)+\varepsilon_x\Big)^2 -\sum_{m=0}^{p-q}\frac{1}{S_{m,n}}\Big\{\sum_{x=0}^{n}\sum_{m'=0}^{p}\beta_{m'}P_{m,n}(x)P_{m',n}(x)+\sum_{x=0}^{n}P_{m,n}(x)\varepsilon_x\Big\}^2\Big]$$
$$=\frac{1}{n-p+q}\Big[\sum_{m=0}^{p}\beta_m{}^2S_{m,n}+(n+1)\sigma^2-\sum_{m=0}^{p-q}\frac{1}{S_{m,n}}\cdot(\beta_m{}^2S_{m,n}{}^2+\sigma^2S_{m,n})\Big]$$
$$=\sigma^2+\frac{1}{n-p+q}\sum_{m=p-q+1}^{p}\beta_m{}^2S_{m,n} \qquad (21)$$

式(21)によれば，$q=0$ のときはランダム誤差の分散 σ^2 のみが残るが $q\geqq1$ ならば $P_{m,n}(x)\,(p\geqq m\geqq p-q+1)$ の展開係数の二乗 $\beta_m{}^2$ に依存する式 (21) 第二項だけ分散が増大する．この増大分は使用した多項式次数が低すぎたゆえに生じたものだから，ランダム誤差ではなく系統偏差の目安を与える．すなわち，前節の通り次数がいたずらに高いのは推定値分散の増大を招くのみだから必要最小限度の次数を用いたいが，もし次数が低すぎると偏差の増大を招く．したがって，次のような次数の決定方法が可能であろう．まず代表的ないくつかの y に関し，$m=p$, $m=p-1$, $m=p-2,\cdots$について $\frac{1}{n-p}\beta_m{}^2S_{m,n}$ を計算し，それらの $m=p-q+1$ より $m=p$ までの和がランダム誤差分 σ^2 に比較して無視できる程度に小さいならば $(p-q)$ 次式を使用してさしつかえない．無視可能か否かは推定値の使用法で決まる系統偏差の許容値による.

特に $q=1$ の場合には最高次微係数の大きさから $(p-1)$ 次多項式を用いた時の偏差の程度が判定される．すなわち，選点直交多項式 $P_{p,n}(x)$ の最高次 x^p の係数は式 (1) より

$$(-1)^p\frac{(2p)!(n-p)!}{(p!)^2n!}$$

であり，さらに，関数 y の変数 t に関する p 次微係数の値（範囲 [0, n] で一定）を $\overset{(p)}{y_0}$ とすれば，y を表わす多項式の x^p の係数は $\frac{h^{p(p)}}{p!}y_0$ となるから，y を展開した時の $P_{p,m}(x)$ の係数 β_p の値は次式で与えられる.

$$\beta_p=(-1)^p\frac{(p!)^2n!}{(2p)!(n-p)!}\cdot\frac{h^{p(p)}}{p!}y_0 \qquad (22)$$

$q=1$ を考慮して式(21) に式(4) と式(22) を代入すれば

$$\sigma'^2=\sigma^2+\frac{h^{2p}(\overset{(p)}{y_0})^2}{n-p+1}\cdot\frac{(p!)^2(n+p+1)!}{(2p+1)!(2p)!(n-p)!} \qquad (23)$$

すなわち，$(p-1)$ 次多項式を用いた場合の系統偏差は $(\overset{(p)}{y_0})^2$ の大きさから見当をつけることができる.

5. 多項式次数に関する線形仮設検定

本節では多少観点を変えて多項式次数選定の問題を考えてみる．すなわち，関数 y は変数 x の p 次式により十分精密に表わされるという前提のもとで，y が $(p-q)$ 次式（ただし $p-q\geqq0$）で表わされるという仮設 (H_0) を設け，この仮説の検定を行なう．検定の結果，この仮設 H_0 が危検率 α にて棄却されると期待される状況ならば，y は x の $(p-q+1)$ 次以上の多項式で表わされると考えなければならない.

線形仮設検定論によれば[3]，前提のもとで計算された残差平方和とその自由度とをそれぞれ SSe, DFe とし前提と仮設の双方の条件のもとで計算された残差平方和と自由度とをそれぞれ SSe^0, DFe^0 とするとき

$$SSH\equiv SSe^0-SSe \qquad (24)$$
$$DFH\equiv DFe^0-DFe \qquad (25)$$

なる量を計算する．また危険率 α, 分子の自由度 DFH, 分母の自由度 DFe なる F 分布関数の値を $F^{DFH}_{DFe}(\alpha)$ とする．然るとき，

$$\frac{SSH/DFH}{SS_e/DFe}\geqq F^{DFH}_{DFe}(\alpha) \qquad (26)$$

が成立するならば，仮設 H_0 は α の危険率を認めた上で棄却されなければならない.

この定式を前述の多項式次数検定に適用するならば，式(19)，(20)と同様にして

$$SSe=\sum_{x=0}^{n}y_x{}^2-\sum_{m=0}^{p}\frac{1}{S_{m,n}}\Big(\sum_{x=0}^{n}P_{m,n}(x)y_x\Big)^2 \qquad (27)$$
$$DFe=n-p \qquad (28)$$
$$SSe^0=\sum_{x=0}^{n}y_x{}^2-\sum_{m=0}^{p-q}\frac{1}{S_{m,n}}\Big(\sum_{x=0}^{n}P_{m,n}(x)y_x\Big)^2 \qquad (29)$$
$$DFe^0=n-p+q \qquad (30)$$

式(27)～(30)を式(24)，(25)に代入すれば

$$SSH=\sum_{m=p-q+1}^{p}\frac{1}{S_{m,n}}\Big(\sum_{x=0}^{n}P_{m,n}(x)y_x\Big)^2 \qquad (31)$$
$$DFH=q \qquad (32)$$

観測値 y_x はランダムな観測誤差を伴うので式(26)の分子，分母の期待値を計算してみよう．計算は式(21)の場合と同様であるので，結果のみ記すと

$$\mathrm{ave}\frac{SSH}{DFH}=\sigma^2+\frac{1}{q}\sum_{m=p-q+1}^{p}\beta_m{}^2S_{m,n} \qquad (33)$$
$$\mathrm{ave}\frac{SSe}{DFe}=\sigma^2 \qquad (34)$$

ゆえに，式(26)の左辺の分子，分母をそれぞれの期待値で置き換えれば，次の仮設棄却の条件式が得られる.

$$1+\frac{1}{q\sigma^2}\sum_{m=p-q+1}^{p}\beta_m{}^2S_{m,n}\geqq F^q_{n-p}(\alpha) \qquad (35)$$

この式によれば，$\frac{1}{q}\sum_{m=p-q+1}^{p}\beta_m{}^2S_{m,n}$ なる量が σ^2 と比べてどの程度の大きさとなるかにより仮設棄却の可否がきまる．また，式(21)を参照すれば，棄却限界では分散増分相対値と F 分布関数の値とには簡単な関係があることがわかる．すなわち，

$$\frac{\sigma'^2-\sigma^2}{\sigma^2}=\frac{q}{n-p+q}(F^q_{n-p}(\alpha)-1) \quad (36)$$

したがって，式(21)に基づき低すぎる次数の多項式の使用に伴う系統的偏差の目安 $(\sigma'^2-\sigma^2)/\sigma^2$ から次数を決定する方法と，本節の式(35)に基づき次数に関する仮設検定から決定する方法とは本質的な矛盾を含むものではない．ただ，前節では推定値の使用法できまる系統偏差許容値を基準にして $(\sigma'^2-\sigma^2)/\sigma^2$ を決めなければならなかったが，そのときランダム誤差に埋れた系統偏差の存在の指摘に伴う危険率は，式(36)を満足する如き α であることがわかる．最良推定値をそのまま使用する場合には α は小さく（すなわちほとんど間違いなく偏差指摘可能なほどに次数も低く）してもかまわないが，多数組のデータから多数の最良推定値を求め，それらをさらに統計的に処理して一層正確な値を求めたい場合には，系統的偏差をきわめて小さくするため α はかなり大きく（系統偏差存在の疑いがほとんどないように次数を高く）なければならない．

特に $q=1$ の場合には，y の最高次微係数の大きさの関係から次数を決定できる．すなわち，式(22)を式(35)に代入整理するならば，仮設棄却の条件は

$$|\overset{(p)}{y_0}|\geq\frac{\sigma}{h^p}\sqrt{\frac{(2p)!(2p+1)!(n-p)!}{(p!)^2(n+p+1)!}(F^1_{n-p}(\alpha)-1)} \quad (37)$$

関数 y が変数 x の p 次多項式では十分精密に表わせないとしても，m がある値以上に増すにつれて $\beta_m{}^2$ が著しく小さくなり，式(35)の左辺第2項に相当する級数が実質的に一項と考えてよいような，なめらかな関数である場合には，p を減らして行くとき式(37)が成立しなくなるような最大の p を求め，これを推定多項式の次数とすればよいであろう．

6. 間隔 0.1 秒毎の 20 点のデータを対象とした数値例

等時間間隔 $h=0.1$ 秒毎に $n+1=20$ 点の観測データが得られた場合の数値例を挙げてみよう．関数 y は式(37)の所で述べたような単純な関数であるとし，推定多項式の次数が3次，2次，1次の場合を考えてみよう．また観測値のランダム誤差の標準偏差は σ メートルとする．

必要な選点直交多項式に関する数値を表1に示す．表中の $Q_{p,n}$, $Q_{p,n}'$, $Q_{p,n}''$ および $R_{p,n}$, $R_{p,n}'$ は式(16)(17)，(18)の下に現われた和を表わす．

表 1-1

$n=19$	$m=0$	$m=1$	$m=2$	$m=3$
$S_{m,n}$	20	140/19 =7. 3684210	308/57 =5. 4035087	5060/969 =5. 2218782
$P_{m,n}\left(\frac{n}{2}\right)$	1	0	$-7/12$ $=-0.58\dot{3}$	0
$P_{m,n}'\left(\frac{n}{2}\right)$	0	$-2/19$ $=-0.10526315$	0	23/114 =0. 20175438
$P_{m,n}''\left(\frac{n}{2}\right)$	0	0	2/57 =0. 035087719	0
$P_{m,n}'''\left(\frac{n}{2}\right)$	0	0	0	$-20/969$ $=-0.020639835$

表 1-2

$n=19$	$p=0$	$p=1$	$p=2$	$p=3$
$Q_{p,n}$	0. 05	0. 05	0. 11297348	0. 11297348
$Q_{p,n}'$	0	0. 0015037593	0. 0015037593	0. 0092988151
$Q_{p,n}''$	0	0	0. 00022784232	0. 00022784232
$R_{p,n}$	0	0	-0.0037878787	-0.0037878787
$R_{p,n}'$	0	0	0	-0.00079744817

式(16)，(17)，(18)と表1の数値を用いるならば

a） $p=3$（3次式）の場合の最良推定値の分散・共分散

$$\mathrm{var}\ \hat{y}\left(\frac{n}{2}+\Delta x\right)\cong\sigma^2\{0.1130+0.00551(\Delta x)^2\}\ (\mathrm{m})^2$$

$$\mathrm{var}\ \hat{\dot{y}}\left(\frac{n}{2}+\Delta x\right)\cong\sigma^2\{0.930-0.0570(\Delta x)^2\}\ (\mathrm{m/sec})^2$$

$$\mathrm{cov}\ \hat{y}\left(\frac{n}{2}+\Delta x\right)\hat{\dot{y}}\left(\frac{n}{2}+\Delta x\right)\cong\sigma^2\times0.0551\,\Delta x\quad(\mathrm{m})^2/\mathrm{sec}$$

b） $p=2$（2次式）の場合の最良推定値の分散・共分散

$$\mathrm{var}\ \hat{y}\left(\frac{n}{2}+\Delta x\right)\cong\sigma^2\{0.1130-0.0228(\Delta x)^2\}\quad(\mathrm{m})^2$$

$$\mathrm{var}\ \hat{\dot{y}}\left(\frac{n}{2}+\Delta x\right)\cong\sigma^2\{0.150+0.0228\,\Delta x)^2\}\ (\mathrm{m/sec})^2$$

$$\mathrm{cov}\ \hat{y}\left(\frac{n}{2}+\Delta x\right)\hat{\dot{y}}\left(\frac{n}{2}+\Delta x\right)\cong\sigma^2\times(-0.0228\,\Delta x)\quad(\mathrm{m})^2/\mathrm{sec}$$

c） $p=1$（1次式）の場合の最良推定値の分散・共分散

$$\mathrm{var}\ \hat{y}\left(\frac{n}{2}+\Delta x\right)\cong\sigma^2\{0.05+0.00150(\Delta x)^2\}\quad(\mathrm{m})^2$$

$$\mathrm{var}\ \hat{\dot{y}}\left(\frac{n}{2}+\Delta x\right)\cong\sigma^2\{0.150+0\}\quad(\mathrm{m/sec})^2$$

$$\mathrm{cov}\ \hat{y}\left(\frac{n}{2}+\Delta x\right)\hat{\dot{y}}\left(\frac{n}{2}+\Delta x\right)\cong\sigma^2\times(0.0150\,\Delta x)\quad(\mathrm{m})^2/\mathrm{sec}$$

範囲〔0, n〕の中央 $x=\frac{n}{2}$ の場合について考えてみよう．b)，c)の var $\hat{\dot{y}}\left(\frac{n}{2}\right)$ の値を比較するならば，p

$=2$ の方が $0.1130/0.05=2.26$ 倍だけ大きい．したがって，もし y が1次式でも十分精密に表わされるのにもかかわらず，2次式によって推定値を求めたならば，$\hat{y}$ の分散期待値は1次式によって求めた場合の 2.26 倍となってしまう．同様に，a)，b) の $\mathrm{var}\,\hat{y}\left(\frac{n}{2}\right)$ の値の比較より，もし y が2次式でも十分精密に表わされるのにもかかわらず，3次式より推定値を求めたならば $\hat{y}$ の分散期待値は2次式により求めた場合の $0.930/0.150=6.20$ 倍となることがわかる．すなわち，不必要に高い次数の多項式によって推定するならば，最良推定値の分散期待値が著しく増大するので好ましくない．

表 2

$n=19$	$p=0$	$p=1$	$p=2$	$p=3$
$\dfrac{(2p)!(2p+1)!(n-p)!}{(p!)^2(n+p+1)!}$	1/20 $=0.05$	1/665 $=0.001504$	1/4389 $=0.000228$	20/245157 $=0.0000816$
$F^1_{n-p}(0.05)$	4.38	4.41	4.45	4.49

次に式(37)の条件を求めてみよう．すなわち，観測に伴うランダム誤差に埋もれた p 次微係数 $\overset{(p)}{y_0}$ を検出し，$(p-1)$ 次多項式では不十分なことを危険率 $\alpha=0.05$ にて指摘できるために必要な $\overset{(p)}{y_0}$ の大きさを求めてみよう．同時に，式(36)よりその限界状態における系統偏差の目安 $(\sigma'^2-\sigma^2)/\sigma^2$ を求めておこう．表2に示された $\{(2p)!(2p+1)!(n-p)!\}/\{(p!)^2(n+p+1)!\}$ と F 分布関数の数値とを用いるならば，

d) $p=3$ の場合

$\dddot{y}_0 \geqq 16.8\,\sigma$　　m/(sec)^3

$(\sigma'^2-\sigma^2)/\sigma^2=0.205$

e) $p=2$ の場合

$\ddot{y}_0 \geqq 2.80\,\sigma$　　m/(sec)^2

$(\sigma'^2-\sigma^2)/\sigma^2=0.192$

f) $p=1$ の場合

$\dot{y}_0 \geqq 0.773\,\sigma$　　m/sec

$(\sigma'^2-\sigma^2)/\sigma^2=0.189$

$\alpha=0.05$ のとき，上述の例では $(\sigma'^2-\sigma^2)/\sigma^2=0.2$ 位の値となっている．多数の最良推定値をさらに統計的に処理してランダム誤差の影響を減少させようとする場合には，この値は必ずしも無視できるほど小さくはない．α をもっと大きい値に選んだならば，系統偏差の大きさは小さくなる．

7. む す び

等間隔観測データ群に最小二乗法を適用してランダム誤差を除去した多項式を求め，これから最良推定値を求める問題において，推定多項式次数の選定に際して払うべき考慮について述べた．独立変数可変範囲の中央付近の関数値および微係数を推定した場合の推定値分散の期待値が推定多項式次数の増大とともに増す様子，関数値を正しく表わせないほどに低すぎる次数の推定多項式を用いた際の分散に含まれる系統偏差成分の大きさを示す公式を求め最良推定値を求めようとする際の多項式の次数が不適切な場合の影響を述べた．また，次数が低すぎるか否かの判定に線形仮設検定を用いた場合の公式を求め，前記の方法との関連について述べた．

多数の等間隔観測データ群の数値的処理の第一段階として最小二乗法による多項式当てはめは，よく利用される所である．本文の考察が多少とも参考になれば幸いである．

謝辞：本文に取上げた問題は，レーダ・データ処理グループ打合せ会で提起されたものであり，本所渡辺教授はじめ日本電気株式会社原一行氏ほかのご教示に負う所が多い．また線形仮設検定論は東京大学工学部森口教授講義ノート（昭和 29 年度）によった．深謝の意を表明する．

(1968 年 1 月 13 日受理)

文 献

1) 森口・他：数学公式 III §27, pp. 99～102, 岩波全書 (1964).
詳しくは Milne; Numerical calculus Princeton (1949).

2) 日本応用力学会：応用統計学 第6章 克誠堂出版（昭和 24 年）.

3) 森口：応用統計学講義（昭和 29 年）.