АКАДЕМИЯ НАУК СССР
СИБИРСКОЕ ОТДЕЛЕНИЕ
АВТОМЕТРИЯ

№ 3 1970

УДК 621.391+519.2

В. С. КИРИЧУК
(*Новосибирск*)

ВЫБОР СТЕПЕНИ ПОЛИНОМА, СГЛАЖИВАЮЩЕГО РЕЗУЛЬТАТЫ ИЗМЕРЕНИЙ

При обработке результатов измерений обычно возникает задача выбора полинома, аппроксимирующего исследуемый процесс. Целью данной работы является получение оптимальной оценки для степени аппроксимирующего полинома в предположении, что границы для степени заданы. Существует ряд методов, которые используются для оценивания истинной степени полинома. Последние из них были предложены Андерсоном [1], Боганником [2] и Хоэлом [3]. В [2] на основе использования корреляционной матрицы строится асимптотически эффективная оценка степени, но алгоритм ее получения настолько сложен с вычислительной точки зрения, что его практическое применение затруднительно. Использование повторных выборок было предложено в [3], но этот метод непригоден при работе с единственной реализацией случайного процесса.

Наиболее приемлемый с практической точки зрения метод содержится в [1], где данная задача сводится к проверке множественной гипотезы и показывается, что оптимальную оценку степени дает последовательная процедура, проверяющая равенство нулю коэффициентов при степенях полинома, причем проверка начинается с коэффициента при высшей возможной степени.

Метод, рассмотренный в [1], дает оценку степени полинома, точность которой падает при подходе к низшей возможной степени q. В данной статье предлагается метод получения оценки, точность которой в области степеней q выше, чем точность оценки, полученной по методу Андерсона.

Пусть $z(t_1), z(t_2), \ldots, z(t_n)$ — истинные значения измеряемого сигнала в моменты времени $t_1, t_2, \ldots, t_n$. В результате измерения сигнала $z(t)$ получим величину $x(t)$, которая отличается от $z(t)$ на ошибку измерения $h(t)$:

$$x(t_i) = z(t_i) + h(t_i) \quad (i = \overline{1, n}). \tag{1}$$

Предполагается, что сигнал $z(t)$ — есть полином степени m, которая лежит в известных границах $q \leqslant m \leqslant s$, ошибки измерения $h(t_i)$ $(i = \overline{1, n})$ независимы, распределены нормально с одинаковой дисперсией σ^2 и имеют нулевое среднее.

Обычно на практике для оценки сигнала $z(t)$ используют метод наименьших квадратов, но оценки, полученные по этому методу, оптимальны лишь в том случае, когда известна степень аппроксимирующего полинома [4]. Поэтому для получения оптимальных линейных оценок измеряемого сигнала $z(t)$ необходимо предварительно по результатам измерений $x(t_1)$, $x(t_2),\ldots,$ $x(t_n)$ определить степень полинома m.

Разложим неслучайную компоненту сигнала $x(t)$ в ряд по ортонормированным полиномам Чебышева, определенным в точках $t_1, t_2,\ldots, t_n$:

$$x(t_i) = \sum_{j=1}^{s} T_j(t_i)\,a_j + h(t_i) \quad (i = 1,\ldots, n). \tag{2}$$

Коэффициенты a_j при полиномах Чебышева $T_j(t_i)$ степени выше s, в силу наших предположений об ограниченности степени, должны быть равны нулю. Уравнение (2) в матричной записи принимает вид

$$X=TA+H, \tag{3}$$

где X — вектор, составленный из значений функции $x(t)$; T — матрица $s\times n$, столбцы которой составлены из значений ортонормированных полиномов Чебышева в точках $t_1, t_2,\ldots, t_n$; A — вектор коэффициентов при полиномах Чебышева; H — вектор, составленный из случайных компонент измеренной величины.

Исследуемую задачу можно упростить, перейдя к новым переменным

$$Y=CX, \tag{4}$$

где C — матрица ортогонального преобразования, такая, что первые s строк матрицы составляются из значений выбранных полиномов Чебышева в точках $t_1, t_2,\ldots, t_n$, а другие строки выбираются взаимно ортонормированными к первым s строкам.

Компоненты y_i вектора Y независимы, распределены нормально с дисперсией σ^2 и вектором средних

$$\eta=M\{Y\}=CTA. \tag{5}$$

Из (5) следует, что компоненты η_i вектора η равны нулю при $i=s+1,\ldots, n$. Плотность распределения вектора Y:

$$P_Y = \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{s}(y_i-\eta_i)^2 + \sum_{i=s+1}^{n} y_i^2\right]\right\}. \tag{6}$$

Если истинная степень полинома равна m, то $\eta_i = 0$ $(i = \overline{m, s})$. Следовательно, задача сводится к определению того, какие η_i равны нулю при $i=q, q+1,\ldots, s$.

Если известно, что степень полинома либо s, либо m, то в этом случае проверяется гипотеза $\eta_{m+1} = \ldots \eta_s = 0$ при альтернативе $\eta_i \neq 0$ $(i = \overline{m+1, s})$.

Рассмотрим случай, когда $r=s-m=1$, т. е. процесс описывается полиномом степени s, и проверяется гипотеза о возможности описания процесса полиномом степени $s-1$ (гипотеза $\eta_s = 0$ при альтернативе $\eta_s \neq 0$). Остальные параметры распределения (6) σ^2, $\eta_1, \eta_2, \ldots, \eta_s$ будут «мешающими» [5]. Предполагая, что дисперсия σ^2 неизвестна, перейдем к достаточным статистикам. Переход к достаточным статистикам позволяет отбросить часть сведений, которая не несет информации об исследуемых параметрах. Для величин $l_i = \frac{1}{\sigma^2}\eta_i$ достаточными будут статистики $U_i = y_i$ $(i = \overline{1, s})$ [5, 6]. Переходя к статистикам $F = \sum_{i=1}^{n} y_i^2$, U_i, распределения P_Y запишем в виде

$$P_Y = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left\{-\frac{\sigma^2}{2}\sum_{i=1}^{n}\eta_i^2\right\} \exp\left\{vF + \sum_{i=1}^{s} l_i U_i\right\}. \tag{7}$$

В [5] показано, что для гипотезы $l_s = 0$ при альтернативе $l_s \neq 0$ существует равномерно наиболее мощный (РНМ) несмещенный критерий. Этот критерий принимает гипотезу, если

$$W_1 = \frac{|U_s|}{\sqrt{\frac{1}{n-s}\left(-\sum_{i=1}^{s} U_i^2 + F\right)}} < C_1, \tag{8}$$

где C_1 — корень уравнения $\alpha/2 = \int_{C_1}^{\infty} t_{n-s}(y)\,dy$; функция $t_{n-s}(y)$ распределена по Стьюденту с $n-s$ степенями свободы; α — выбранный уровень значимости.

Когда $r > 1$, то РНМ несмещенного критерия для гипотезы $l_i = 0$ $(i = \overline{m+1, s})$ не существует. Но в этом случае, как показано в [5], существует наиболее мощный инвариантный критерий. Наиболее мощный инвариантный критерий отвергает гипотезу, когда W_2 слишком велико, т. е.

$$W_2 = \frac{1}{r}\sum_{i=m+1}^{s} y_i^2 \Big/ \frac{1}{n-s}\sum_{i=s+1}^{n} y_i^2 > C_2. \tag{9}$$

Граничная точка C_2 при данном уровне значимости α определяется из уравнения

$$\int_{C_2}^{\infty} F_{r, n-s}(y)\,dy = \alpha,$$

где $F_{r, n-s}$ — распределение Фишера с r, $n-s$ степенями свободы. Мощность критерия β при альтернативе δ равна

$$\beta(\delta) = \int_{C_2}^{\infty} F_{r, n-s}^{\delta}(y)\,dy,$$

где $F_{r, n-s}^{\delta}$ — нецентральное распределение Фишера с параметром нецентральности [7]

$$\delta^2 = 1/\sigma^2 \sum_{i=1}^{n} \eta_i^2.$$

Критерий (8) совпадает с критерием (9) при $r=1$ и, следовательно, является не только РНМ несмещенным, но и наиболее мощным инвариантным.

В применении к конкретным задачам непосредственное выполнение преобразования (4) не всегда удобно. Учитывая то, что ортогональные преобразования оставляют расстояния неизменными, преобразуем критерии (8), (9):

$$W_1 = |y_s| / \sqrt{\frac{1}{n-s}\sum_{i=1}^{n}(x_i - \hat{z}_i)^2} < C_1; \tag{10}$$

$$W_2 = \frac{1}{r}\sum_{i=m+1}^{s}(\hat{z}_i - \hat{\hat{z}}_i)^2 / \frac{1}{n-s}\sum_{i=1}^{n}(x_i - \hat{z}_i)^2 < C_2, \tag{11}$$

где $\hat{z}$ — оценки, полученные по методу наименьших квадратов при условии, что процесс описывается полиномом порядка s; $\hat{\hat{z}}$ — аналогичные оценки при условии, что процесс описывается полиномом порядка m.

Статистика W_1 имеет распределение Стьюдента и поэтому некритична к распределению исходных величин X, т. е. с хорошей достоверностью можно применять этот критерий к величинам, имеющим распределение, отличающееся от нормального. Это существенно расширяет область применения критерия.

При наличии априорного распределения вероятностей возможных степеней полинома (т. е. если задан вектор P с компонентами p_i, обозначающими вероятность того, что в описании процесса присутствует полином Чебышева i-й степени) критерий (9) можно привести к виду

$$W_2' = \sum_{i=m+1}^{s} p_i y_i^2 / \frac{1}{n-s}\sum_{i=1}^{n}(x_i - \hat{z}_i)^2 < C_2'. \tag{9'}$$

Критерий (9) является частным случаем критерия (9′) с вероятностями $p_i = 1/r$ $(i=m+1, s)$.

Выше приведены критерии для выбора степени из двух возможных степеней: алгоритм (10) для выбора между s-й и $(s-1)$-й степенями, алгоритм (11) для выбора между s-й и $(s-m)$-й степенями. Теперь рассмотрим множественную гипотезу выбора степени из $q, q+1, \ldots, s$ возможных степеней. В рамках модели, предложенной Андерсоном [1], показано, что оптимальным критерием для решения данной множественной гипотезы является процедура, основанная на последовательном применении критерия (10). Этот критерий будет РНМ несмещенным.

Зададим вектор α с компонентами α_i, которые являются уровнями значимости для проверки гипотезы $\eta_i = 0$ при условии $\eta_j = 0$ $(j=\overline{i+1, s})$. При этом оптимальным будет критерий

$$|y_i| / \sqrt{\left(\sum_{j=1}^{n} x_j^2 - \sum_{j=1}^{i-1} y_j^2\right)\frac{1}{n-i}} < C_i, \tag{12}$$

где $\alpha_i/2 = \int_{C_i}^{\infty} t_{n-i}(y)\,dy$. Процедура начинается с проверки, удовлетворяет ли величина y_s соотношению (12) с $i=s$; если соотношение (12) выполняется, то аналогично строится критерий для проверки гипотезы $\eta_{s-1} = 0$ при условии $\eta_s = 0$ и т. д.

Процедура заканчивается в том случае, если y_i не будет удовлетворять соотношению (12). Тогда полагается, что степень сглаживающего полинома есть i. Уровень значимости множественной гипотезы (при условии, что в качестве степени выбрано i) равен $\sum_{j=s}^{i} p_j$, где p_j связано с α_j соотношением

$$\alpha_j = p_j/(1 - p_{j+1} + \ldots + p_s) \quad (j = \overline{i, s}).$$

Функция мощности этого критерия:

$$\beta_i = (1 - \prod_{j=s}^{i} \Delta_j),$$

где Δ_j — ошибка второго рода гипотезы $\eta_j = 0$ при уровне значимости α_j.

При сравнении мощности данного последовательного критерия с мощностью критерия (11) (т. е. сравнивается мощность, которая получается на i-м шаге процесса, с мощностью критерия (11) для гипотезы $\eta_s = \ldots = \eta_{s-i+1} = 0$ при одном уровне значимости) оказывается, что, пока проверяемая степень полинома близка к s, отличие этих мощностей несущественно (при проверке степени s они совпадают), но при значительном удалении проверяемой степени от s разница становится значимой. Это послужило стимулом к выработке критерия, который начинает проверку с низшей степени q и дает в этой области степеней приемлемую мощность. Таким критерием является последовательная процедура: сначала проверяется гипотеза H_0 $\eta_{q+1} = \ldots = \eta_s = 0$ [для такой гипотезы оптимальным будет критерий (11)]; если гипотеза отвергается, то строится аналогичный критерий с $C = C_2 r/(r-i)$ для гипотезы H_1 $\eta_{q+2} = \ldots = \eta_s = 0$ и т. д. Процедура прекращается, если H_i принимается, и за степень сглаживающего полинома выбирается величина $q+i$. Уровень значимости α_i для гипотезы H_i:

$$\alpha_i = \int_{C}^{\infty} F_{r-i,\, n-s}(y)\, dy, \tag{13}$$

где C_2 является корнем уравнения $\alpha_0 = \int_{C_2}^{\infty} F_{r,\, n-s}(y)\, dy$. Используя доказательство оптимальности для критерия (12), приведенное в [1], можно доказать, что данный критерий будет наиболее мощным в классе инвариантных критериев. Мощность такого критерия близка к мощности критерия (9) при степенях, близких к q (при проверке гипотезы $\eta_s = \ldots = \eta_{q+1} = 0$ они совпадают), а при увеличении проверяемой степени мощность падает. При сравнении мощностей критериев (12) и (13) можно показать, что при $r \geqslant 5$ критерий, основанный на распределении Фишера, дает лучший результат при степенях, близких к q, чем критерий (12). Это справедливо при практически приемлемых мощностях (начиная с 0,5 и выше).

Для реализации вышеприведенных алгоритмов на ЭВЦМ на языке АЛГОЛ-60 составлена и отлажена программа определения степени полинома (процедура 1). Перед пуском программы в нее необходимо ввести массив исходных данных X, длину этого массива n, высшую допустимую степень s и нижнюю степень q. Эта программа предназначена для работы как с алгоритмом (12), так и с (13). При работе с алгоритмом (12) в программе необходимо использовать в качестве degree процедуру 2. В процедуре 2 T — массив границ доверительных интервалов, который соответствует выбранному вектору α. При работе с алгоритмом, основанном на распределении Фишера, в программе polinom в качестве degree используется процедура 3. В этой процедуре T — константа, которая соответствует выбранному уровню значимости α_0. По окончании работы программы на печать выдается степень сглаживающего полинома m и массив оценок при полиномах Чебышева z. Программа была проверена на многих конкретных реализациях случайных процессов; счет велся на ЭВЦМ типа БЭСМ-6. Время счета $t=2 \div 3$ *сек* при $n=1200$ и $s=50$.

ПРОЦЕДУРА 1

```
procedure polinom (x, s, q, n, m, z); begin integer k, i;
real A, a, b, c, d, R; real array φ0, φ1, φ2 [0 : n — 1], y [1 : s];
procedure delta (A, k); begin a : = b : = 0; c : = k ↑ 2×(n ↑ 2—k ↑ 2)/(4×(4×k ↑ 2 —
1)×(n/3) ↑ 2); for i : = 0 step 1 until n — 1 do {φ0 [i] : = φ1 [i];
φ1 [i] : = φ2 [i]; φ2 [i] : = (i — R)×φ1 [i]/(n/3)— c×φ0 [i]; a : = a+ φ2 [i]×x [i];
b : = b+φ2 [i] ↑ 2}; A : = a/sqrt (b) end;
R : = (n — 1)/2; a : = b : = c : = d : =0; for i : =0 step 1 until n — 1 do {φ1 [i] : =1;
φ2[i] : = (i — R)/(n/3); a : =a+φ2[i] ↑ 2; c : = c+φ2 [i]×x [i];
b : = b+x [i]; d : = d+x [i] ↑ 2}; z[1] : =b/sqrt(n); z[2] : =c/sqrt(a);
for i : =3 step 1 until s do {delta (A, i — 2); z [i] : = A};
a : =0; for i : =1 step 1 until s do a : = a+z [i] ↑ 2; a : = a — d;
degree (T, m); output (m, z) end;
```

ПРОЦЕДУРА 2

```
procedure degree (T, m); begin for i : =s step —1 until q+1 do {y [i] : =z [i]/sqrt(a/
(n — i)); a : = a+z [i] ↑ 2; if abs (y [i]) ≥ T [i] then on tt}; tt: m : = i end;
```

ПРОЦЕДУРА 3

```
procedure degree (T, m); begin b : = 0; for i : = q+1 step 1 until s do b : = b+z [i] ↑ 2;
for i : = q+1 step 1 until s do {y [i] : = b/a; b : = b — z [i] ↑ 2; if abs (y[i]) ≤ T×(s—q)/
/(s — i) then on tt}; tt: m : =i — 1 end;
```

ЛИТЕРАТУРА

1. T. W. Anderson. The Choise of Degree of a Polinomial Regression as a Multiple Decision Problem.— Ann. Math. Statist., 1962, 33.
2. Г. Н. Боганик. Об установлении порядка уравнения параболической регрессии.— Теория вероятностей и ее применение, 1967, т. XII, № 4.
3. P. G. Hoel. On Testing for Degree of a Polynomial.— Techometrics, 1968, 10, № 4.
4. Ю. В. Линник. Метод наименьших квадратов и основы математической обработки наблюдений. М., Физматгиз, 1958.
5. Э. Леман. Проверка статистических гипотез. М., «Наука», 1964.
6. С. Уилкс. Математическая статистика. М., Физматгиз, 1958.
7. Г. Шеффе. Дисперсионный анализ. М., Физматгиз, 1963.

Поступила в редакцию
17 июня 1969 г.