УДК 614.841.345

А.Б. Исаев, Р.М. Мархиев

АЛГОРИТМ ПОСТРОЕНИЯ ПАРАБОЛИЧЕСКОЙ РЕГРЕССИОННОЙ ЗАВИСИМОСТИ ПО ВЫБОРКЕ РЕЗУЛЬТАТОВ СОВМЕСТНЫХ ИЗМЕРЕНИЙ ОТЯГОЩЕННЫХ СЛУЧАЙНЫМИ ПОГРЕШНОСТЯМИ

Рассмотрена возможность разработки алгоритма, который позволит значительно повысить точность найденной модели (ее адекватность к истиной) по сравнению с традиционным алгоритмом МНК.

Ключевые слова: итерационный процесс, алгоритм конфлюэнтного анализа, МНК, случайные погрешности, регрессионная зависимость.

Обрабатывать результаты совместных измерений в условиях отягощения погрешностями как зависимой, так и независимой переменной, необходимо при помощи алгоритмов конфлюентного анализа, обеспечивающих в этих условиях наибольшую близость рассчитанной с их помощью аппроксимирующей зависимости к неизвестной, истинной зависимости. Как известно, оценки [3], полученные традиционным методом наименьших квадратов (МНК), в данном случае оказываются смещенными и неэффективными.

Рассмотрим схему итерационного процесса, учитывающую априорную информацию о совместной, условной и частных плотностях распределений случайных переменных, составляющих экспериментальную выборку $(x_1, y_1), (x_2, y_2, \ldots, (x_n, y_n)$.

Итерационный процесс нахождения оценок $\tilde{a}_i$ параметров a_i зависимости

$$\eta(\xi, a) = a_0 + a_1\xi + a_2\xi^2 + \ldots, \tag{1}$$

где $Mx_i = \xi_i; \quad My_i = \eta_i; \quad i = \overline{1}, n;$

$$M(y / x) = \eta(x, a) = a_0 + a_1 x + a_2 x^2 + \ldots; \tag{2}$$

(x является случайной переменной, имеющей одинаковые одномерные распределения $p_i(x) = p(x, \xi_i, \sigma_{xi}^2))$ с центрами в точках $\xi_i (i = 1, n)$ и дисперсиями σ_{xi}^2) заключается в пошаговой процедуре отыскания минимума квадратичной формы:

$$S_{(k)}^2 = \sum_{i=1}^{n} \omega_i^{(k)} \left[y_i - M(x / y) \right]^2 = \sum_{i=1}^{n} \omega_i^{(k)} \left[y_i - \widehat{\eta}(x_i, a) \right]^2, \tag{3}$$

где $\omega_i^{(k)}$ — вес k-го приближения x_i; $\omega_i^{-1(k)} = \tilde{\sigma}^2(y, x)$;

$$M(y / x) = \int_Y \int_{X_i} y p(y / x_i) p_i(x) dx dy \quad — \tag{4}$$

соответствует значению величины y, усредненному по допустимой области изменения X_i величины x вокруг точки ξ_i; Y-область изменения y; $\tilde{\eta}(x_i, a)$— — оценка для i-го шага (1); $\tilde{\sigma}^2(y, x_i)$ - оценка k-го шага для дисперсии $\sigma^2(y, x_i)$:

$$\sigma^2(y, x_i) = M\left\{\left[y - M(y / x_i)\right]^2\right\} \quad — \tag{5}$$

условная дисперсия у относительно регрессии $M(y / x_i)$.

Целесообразность подобного выбора весов будет видна из численных расчетов, подтвердивших корректность этого далеко не единственного способа назначения весов, интуитивно примыкающего к выбору весов в случае взвешенного МНК.

Чтобы реализовать на ЭВМ итерационный процесс, порожденный минимизацией (3), надо получить аналитические выражения для оценок $\tilde{\eta}(x_i, a)$ и $\tilde{\sigma}^2(y, x_i)$, определяющих скорость сходимости и точку сходимости итерационного процесса.

Перейдем к расчету оценок $\tilde{\eta}(x_i, a)$ и $\tilde{\sigma}^2(y, x_i)$. Согласно определению регрессии как условного математического ожидания

$$M(y / x) = \int\limits_Y y p(y / x) dy = \eta(x, a), \tag{6}$$

где $x \in X_i$, $x \sim p_i(x)$, видим

$$M(y / x_i) = \int\limits_Y \int\limits_{X_i} y p(y / x) p_i(x) dy dx = \int\limits_{X_i} \eta(x, a) p_i(x) dx. \tag{7}$$

Заметим, что для (7) справедлива эквивалентная запись $M(y / x_i) = \int\limits_Y y p(y / x \in X_i) dy$, где $p(y, x_i) = p(y / x \in X_i)$ — условная плотность вероятности для y, если $x \in X_i$ — допустимой области значений содержащей точку ξ_i

$$p(y / x \in X_i) \equiv p(y / x_i) = \int\limits_{X_i} p(y / x) p_i(x) dx. \tag{8}$$

Если функция $\eta(x, a)$ в окрестности точки x_i допускает разложение

$$\eta(x, a) = \eta(x_i, a) + (x - x_i)\frac{\partial \eta}{\partial x}\bigg| + \frac{1}{2}(x - x_i)^2 \frac{\partial^2 \eta}{\partial x^2}\bigg|, \tag{9}$$

Которое справедливо с точностью до членов выше второго порядка степеней $(x - x_i)$, то подставив (9) в (7), получим $M(y / x_i) \cong \eta(x_i, a) + \frac{1}{2}\eta''_{x = x_i}\sigma^2_{xi}$, поскольку $\int\limits_{Xi} p_i(x) dx = 1$, $\int\limits_{X_i} (x - x_i)\eta'_{x = x_i} p_i(x) dx = 0$.

Введя обозначения $\nabla_1 = \partial\eta / \partial x \Big|_{x=x_i} = \eta'_{x=x_i}, \nabla_2 = \partial^2\eta / \partial x^2 \Big|_{x=x_i} = \eta''_{x=x_i}$, можем переписать (7) в виде $M(y / x_i) = \eta(x_i, a) + \dfrac{1}{2}\nabla_2\sigma^2_{xi}$. (11)

После несложный выкладок

$$\begin{aligned}\sigma^2(y - x_i) &= \int\limits_Y \left[y - M(y / x_i)\right]^2 p(y / x \in X_i)dy = \\ &= \int\limits_Y \left\{\left[y - \eta(x, a\right]^2 + \left[\eta(x, a) - M(y / x_i)\right]^2\right\} p(y / x_i)dy\end{aligned} \qquad (12)$$

Подставив (8) и (11) в (12) и опустив длинные выкладки, придем к

$$\sigma^2(y, x_i) = \sigma^2_i + \sigma^2_{xi}\nabla^2_{1i}; \qquad (13)$$

$$\sigma^2_i \int\limits_{X_i}\int\limits_Y \left[y - \eta(x, a)\right]^2 p(y / x)p_i(x)dxdy = \int\limits_{X_i} \sigma^2_y p_i(x)dx, \qquad (14)$$

где σ^2_y — условная дисперсия величины y относительно линии регрессии, полученной в отсутствие погрешностей измерений y независимой переменной $\sigma^2_{xi} = \int\limits_{xi} (x - \xi_i)^2 p_i(x)dx$, $\nabla^2_{1i} = \eta'_{x=x_i}{}^2$, а (14) дает «исправленное» на распределение $p_i(x)$, т.е. усредненное по области X_i значение σ^2_y.

При получении (13) использованы уравнения (4)-(9), а члены $\theta(\sigma^2_{xi})$ отброшены.

На основании (11), (13) и (14) приходим к следующей схеме итерационного процесса.

1. Составляем сумму $S^2_{(0)} = \sum\limits_{i=1}^{n} \omega_0^{(0)} \left[y_i - \eta(x_i, a) - \dfrac{1}{2}\nabla_{2i}\sigma^2_{xi}\right]^2$, где $\omega^{(0)}_i = \sigma^{-2}_{yi} = \left(\int\limits_{Xi} \sigma^2_y p_i(x)dx\right)^{-2}$ — веса нулевого приближения.

2. Подсчитываем величины $\nabla^2_{1i}(\tilde{a}_{(0)})\sigma^2_{xi}$, $\omega^{(1)}_i = \left[\sigma^2_{yi} + \nabla^2_{1i}(\tilde{a}_{(0)})\sigma^2_{xi}\right]^{-1}$.

3. Составляем сумму $S^2_{(1)} = \sum\limits_{i=1}^{n} \left[y_i - \eta(x_i, a) - \dfrac{1}{2}\nabla_{2i}(\tilde{a}_{(0)})\sigma^2_{xi}\right]^2$ и отыскание значения $\tilde{a}_{(1)}$, минимизруеещее $S^2_{(1)}$ и т.д. При достижении на k-м шаге $\left|(\tilde{a}_{(k)} - \tilde{a}_{(k-1)}) / \tilde{a}_{(k-1)}\right| \le \varepsilon$, где ε — наперед заданное малое число, процесс останавливается.

В качестве оценок параметров принимаем численные значения параметров $\tilde{a}_{(k)}$ в точке остановки итерационного процесса. Для иллюстрации корректности оценок, получаемых с помощью этого алгоритма проведен численный модельный эксперимент. Выбрана функция вида $\eta = -2\xi^2 + 10\xi - 8$, где $\xi_i = M_{xi}$, и для значений $\xi_i = 2i/n, \quad i = 0, n-1$.

Таким образом, результаты измерений у независимой и зависимой переменных оказываются случайными величинами, распределенными соответственно:

$x_i \sim N(\xi_i; 0{,}04)$,

$y_i \sim N(\eta_i; 0{,}04)$;

Составлена целевая функция $S = \sum_{i=1}^{n} \omega_i \left[y_i - \eta(x_i, a) - \frac{1}{2} \nabla_{2i} \sigma_{xi}^2 \right]^2$ (15)

Для предложенной итерационной процедуры. Итерационный процесс останавливался при $\varepsilon = 0.001$. Определены также оценки для функции (15) по МНК.

Для итерационного алгоритма конфлюентного анализа получены оценки параметров:

$\tilde{a}_0 = -8{,}23; \quad \tilde{a}_1 = 10{,}06; \quad \tilde{a}_2 = -2{,}00;$

Оценки МНК следующие:

$\tilde{a}_0 = -8{,}01; \quad \tilde{a}_1 = 10{,}17; \quad \tilde{a}_2 = -2{,}17;$

Таким образом, предложенный метод определения оценок зависимостей дает более точное значение искомых параметров; получаемая с его помощью кривая расположена ближе к истинной, чем кривая, определенная МНК.

СПИСОК ЛИТЕРАТУРЫ

1. *Исаев А.Б.* Измерительная техника, 1982. — №10 — С. 13.

2. *Федоров В.В.* Теория оптимального эксперимента. — М.: Наука, 1976.

3. *Демиденко Е.З.* Линейная и нелинейная регрессии. — М.: Финансы и статистика, 1981. ГИАБ

КОРОТКО ОБ АВТОРАХ

Исаев Андрей Борисович — кандидат технических наук, доцент,
Мархиев Рамазан Мусаевич — аспирант,
Российский университет дружбы народов, rector@rudn.ru