

Моделирование процессов методом конфлюэнтного анализа

Н.М. Оскорбин, М.А. Сапаров, Ш.А. Сафарова
АлтГУ, г. Барнаул

В данной статье рассмотрена методика сравнительного исследования точности оценок параметров в задачах конфлюэнтного и регрессионного анализов. С использованием технологии имитационного моделирования разработана компьютерная модель в предположениях регрессионного и конфлюэнтного анализа, в результате обработаны модельные данные с ошибками измерения всех переменных и произведен анализ полученных результатов исследования с точными значениями коэффициентов линейной модели.

Ключевые слова: *математическое моделирование процессов, метод наименьших квадратов, регрессионный анализ, конфлюэнтный анализ, компьютерное моделирование.*

В данной работе решается проблема создания компьютерной модели для исследования точности в задачах прогноза и оценки параметров методами наименьших квадратов (МНК) и конфлюэнтного анализов на основе результатов вычислительных экспериментов в среде Excel. В нашей работе ошибки наблюдения входных и выходных переменных являются случайными, поэтому для решения задачи моделирования процессов необходимо применять статистические методы, использующие метод наименьших квадратов.

В литературе наиболее известным методом статистического моделирования является регрессионный анализ. Теоретически этот метод обоснован для случая, когда входные переменные процесса измеряются с ошибками существенно меньшими ошибок измерения выходной переменной. Для линейного процесса при достаточно большом числе наблюдения N в литературе получены функции распределения оценок вектора параметров α и оценок выходной переменной моделируемого процесса. В этих условиях методами математической статистики можно при фиксированном значении входных переменных и заданном уровне доверительной вероятности

найти доверительный интервал для среднего значения выходной переменной исследуемого процесса:

Задача математического моделирования процессов в общем случае при наличии ошибок наблюдений за входными переменными и выходной переменной формулируется следующим образом: найти функцию $y = F(x, a)$ причинно-следственной связи выходной переменной от входных с точность до A – множества возможных значений параметров a ($a \in A$).

Тогда можно оценить ожидаемые значения выходной переменной при известном значении вектора x_0 решением двух задач математического программирования:

$$y_0^H = \min_{a \in A} F(x_0, a); \quad y_0^V = \max_{a \in A} F(x_0, a). \quad (1)$$

Конфлюэнтный анализ определяется как совокупность методов математического моделирования процессов в условиях, когда входные переменные $x \in R^n$ являются количественными (случайными или неслучайными) и наблюдаются в каждом опыте с ошибками [1, 2]. По этому условию метод конфлюэнтного анализа отличается от методов классического регрессионного анализа.

С учетом предположений относительно ошибок, таблица наблюдений $\{Z_1^j, \dots, Z_n^j, V^j\}, j = 1, \dots, N$ за всеми переменными процесса определяется так:

$$\begin{aligned} Z_i^j &= x_i^j + \varepsilon_{xi}^j; \quad V_i^j = y + \varepsilon_{yi}^j; \quad M[\varepsilon_{xi}^j] = M[\varepsilon_{yi}^j] = 0; \quad D[\varepsilon_{xi}^j] = d_i; \quad D[\varepsilon_{yi}^j] \\ &= d_y. \end{aligned}$$

Связь входных и выходной переменных предполагаем заданной: $y = a_1 x_1, \dots, a_n x_n$, а задачу конфлюэнтного анализа ограничиваем оценкой истинных значений коэффициентов приведенного уравнения.

Метод наименьших квадратов в данном случае позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y от расчетных (теоретических) значений минимальна. При числе входных переменных равном 2 эта задача записывается в виде:

$$S = \sum_{i=1}^N (y_i + \varepsilon_{yi} - a_1(x_{1i} + \varepsilon_{x_{1i}}) - a_2(x_{2i} + \varepsilon_{x_{2i}}))^2 \rightarrow \min. \quad (2)$$

Для определения минимума функции двух переменных вычисляются частные производные этой функции по каждому из оцениваемых параметров и приравниваются к нулю, т.е. $S'_{a_1} = 0$; $S'_{a_2} = 0$:

$$\begin{cases} \frac{\partial S}{\partial a_1} = 2 \sum_{i=1}^N \left(y_i + \varepsilon_{y_i} - a_1(x_{1i} + \varepsilon_{x_{1i}}) - a_2(x_{2i} + \varepsilon_{x_{2i}}) \right) (x_{1i} + \varepsilon_{x_{1i}}) = 0 \\ \frac{\partial S}{\partial a_2} = 2 \sum_{i=1}^N \left(y_i + \varepsilon_{y_i} - a_1(x_{1i} + \varepsilon_{x_{1i}}) - a_2(x_{2i} + \varepsilon_{x_{2i}}) \right) (x_{2i} + \varepsilon_{x_{2i}}) = 0 \end{cases}.$$

Рассмотрим метод оценки истинных значений параметров моделируемого процесса. Пусть без потери общности средние значения переменных в таблице наблюдений являются нулевыми. Тогда согласно [2] оценки истинных коэффициентов моделируемого процесса находятся решением следующей системы уравнений:

$$a_i(R_{ii} - d_i) + \sum_{k=1, k \neq i}^n a_k R_{ik} = R_{iy}; \quad i = 1, \dots, n. \quad (3)$$

В приведенной системе n уравнений коэффициентами выступают ковариации соответствующих переменных, вычисленные по таблице наблюдений $\{Z_1^j, \dots, Z_n^j, V^j\}, j = 1, \dots, N$.

Вычислительные эксперименты по исследованию средней точности оценок методом конфлюэнтного анализа и сравнению с точностью оценок МНК проведены для числа наблюдений N, равных 20 при 15 повторности и числа переменных n, равных 2. Входные переменные моделируемого процесса генерировались функцией СЛЧИС() в интервале $[-100; 100]$ и предварительно центрировались относительно нулевого значения. Значение выходной переменной у вычислялось с коэффициентами $a_1 = 1; a_2 = 2$. Переменными при исследовании выступали уровни дисперсий нормально распределенных ошибок наблюдений, приведенные в следующей таблице. Погрешности оценок параметров уравнения определялись как модули разности их истинных значений. В приведенной таблице указаны средние значения ошибок по 15 вычислительных экспериментов.

№ п/п	Дисперсии ошибок			Ошибки МНК		Ошибки КА	
	d_1	d_2	d_y	$Mo[a_1]$	$Mo[a_2]$	$Mo[a_1]$	$Mo[a_2]$
1	0	0	0,25	0,73%	0,83%	0,73%	0,83%
2	0,5	0,5	0,25	3,40%	4,10%	2,56%	2,79%
3	1,0	1,0	0,5	6,29%	5,75%	4,86%	3,00%
4	1,5	1,5	1,0	6,87%	7,87%	4,18%	2,99%

Как результат экспериментов можно сделать вывод о том, что метод конфлюэнтного анализа эффективнее метода наименьших квадратов, причем эффективность становится выше при увеличении

дисперсии ошибок. Следовательно, метод конфлюэнтного анализа является полезным для решения задач моделирования процессов по реальным данным.

Библиографический список

1. Кубланов М.С. Математическое моделирование. Методология и методы разработки математических моделей механических систем и процессов. – М.: МГТУ ГА, 2013. – 125с.
2. Максимов А.В., Оскорбин Н.М. Многопользовательские информационные системы: основы теории и методы исследования. – Барнаул: Изд-во Алт. ун-та, 2013. – 264с.