

УДК 519.65, 519.245

О возможности применения методов Монте-Карло в анализе нелинейных регрессионных моделей

Г.И. Рудой

Вычислительный центр Российской академии наук, ул. Вавилова, 40, Москва, 119333

E-mail: 0xd34df00d@gmail.com

Рудой Г.И. О возможности применения методов Монте-Карло в анализе нелинейных регрессионных моделей // Сиб. журн. вычисл. математики / РАН. Сиб. отд-ние. — Новосибирск, 2015. — Т. 18, № 4. — С. 425–434.

Предлагается понятие устойчивости коэффициентов существенно нелинейных суперпозиций, а также метод оценки устойчивости решения задачи восстановления регрессионной зависимости. Данный метод иллюстрируется вычислительным экспериментом на данных, полученных при измерении зависимости показателя преломления полимера от длины волны в области 400–1000 нм, соответствующей области прозрачности полимера, а также исследуется сходимость предложенного метода к известному аналитическому решению для линейной зависимости.

DOI: 10.15372/SJNM20150407

Ключевые слова: *символьная регрессия, нелинейные модели, устойчивость решений, дисперсия прозрачной среды, методы класса Монте-Карло.*

Rudoy G.I. On applying Monte Carlo methods to analysis of nonlinear regression models // Siberian J. Num. Math. / Sib. Branch of Russ. Acad. of Sci. — Novosibirsk, 2015. — Vol. 18, № 4. — P. 425–434.

This paper presents a criterium, called the coefficients stability for inaccuracy in determining the coefficients of nonlinear regression models describing inexact data. A method for the coefficients stability estimation is also described. The proposed criterium is illustrated by a computational experiment with the data obtained by measurements of a refractive index dependence on the wavelength in 400–1000 nm band for a transparent polymer. The convergence of the proposed criterium to the known analytical solution for the case of linear regression is also studied.

Keywords: *symbolic regression, nonlinear models, solution stability, transparent medium dispersion, Monte Carlo methods.*

1. Введение

Символьная регрессия часто используется для построения экспертно интерпретируемых моделей [1–5]. В приложении к естественно-научным экспериментам речь идет о восстановлении функциональной зависимости между измеряемыми и задаваемыми с некоторой точностью параметрами, например зависимость термоэмиссионного тока электронной лампы от температуры катода $I_k(T)$ при неизменной геометрии системы и разности потенциалов, зависимость мощности излучения непрерывного лазера от коэффициента отражения выходного зеркала $W_l(R)$ при постоянных модовой структуре излучения и мощности возбуждения активной среды, зависимость показателя преломления материала от длины волны $n(\lambda)$ при постоянной температуре и т.п., далее мы более подробно рассмотрим именно последний случай.

При регрессионном анализе такого рода экспериментов необходимо учитывать следующие обстоятельства:

1. Все измеряемые (и контролируемые) параметры в каждой экспериментальной точке определяются с некоторой (обычно известной) точностью, причем абсолютная погрешность σ_i соответствующего параметра может существенно изменяться в исследуемом диапазоне. Например, если в качестве спектрального прибора, выделяющего конкретную длину волны λ_i при измерении $n_i(\lambda_i)$, используется дифракционная решетка, то $\frac{\sigma_i}{\lambda_i} \approx \text{const}$, и считать погрешность определения длины волны постоянной некорректно для измерений в достаточно широком спектральном диапазоне.
2. Как правило, эксперимент ставится так, что измеряется функциональная зависимость от одной переменной x , т. е. строится зависимость вида $y(x, \omega)$, где ω — набор параметров, которые поддерживаются неизменными. Как отмечалось выше, параметры поддерживаются постоянными с конечной точностью и в ряде случаев при построении модели это обстоятельство необходимо учитывать. Однако обычно эксперт заранее может оценить влияние вариаций условий эксперимента и обеспечить необходимую стабильность проведения измерений. В противном случае необходимо прямо учитывать зависимость измеряемой характеристики от нескольких переменных, что для целей настоящей работы неприципиально.
3. В большинстве случаев эксперт заранее знает вид искомой функциональной зависимости или же требуется провести выбор между несколькими возможными вариантами, что упрощает задачу регрессии. В то же время для эксперта важнейшее значение имеют не только оптимальные значения коэффициентов регрессионной формулы, но и дисперсия этих коэффициентов, а также связь их дисперсии с точностью определения измеряемых (контролируемых) в эксперименте величин. Это особенно существенно в тех случаях, когда коэффициенты регрессионной модели прямо связаны с фундаментальными характеристиками исследуемого процесса и по ним рассчитывается, например, эффективная масса электронов в полупроводнике, температура Дебая, резонансная частота и затухание оптического перехода и т. д. Соответственно, точность измерения соответствующих материальных констант определяется точностью вычисления коэффициентов регрессионной модели.

В такой постановке, когда требуется определить не только оптимальные коэффициенты регрессионной модели, но и их погрешность, насколько нам известно, задача нелинейной регрессии не рассматривалась. Известны теоретические результаты для случая линейной регрессии

$$y = ax + b;$$

в случае, когда дисперсия всех экспериментально измеренных значений y_i зависимой переменной y одна и та же: $D(y_i) = \sigma^2$, а значения независимой переменной x_i известны точно: $D(x) = 0$. Тогда при переходе к представлению

$$y_i = a(x_i - \bar{x}) + b + \xi_i \quad | \quad i \in \{1, \dots, \ell\},$$

где $\bar{x} = \frac{\sum_{i=1}^{\ell} x_i}{\ell}$, а $\xi_i \sim \mathcal{N}(0, \sigma^2)$, согласно [6], случайные величины a и b независимы и нормально распределены, и, кроме того, их дисперсии выражаются известными соотношениями:

$$D(a) = \frac{\sigma^2}{\sum_{i=1}^{\ell} (x_i - \bar{x})^2}, \quad (1)$$

$$D(b) = \frac{\sigma^2}{\ell}. \quad (2)$$

В настоящей работе предложен общий метод определения погрешности коэффициентов нелинейной регрессии, и на примере зависимости $n(\lambda)$ для прозрачного полимера определена зависимость погрешности параметров регрессии от точности определения длины волны и показателя преломления. Здесь мы ограничиваемся одной независимой переменной λ . Обобщение предлагаемого метода на случай нескольких переменных проводится очевидным образом.

2. Основная гипотеза

Пусть имеется набор из ℓ экспериментальных точек, в окрестности каждой из которых задана вероятность отклонения экспериментальных значений измеряемых переменных от этой точки (например, из-за случайных ошибок измерений). Иными словами, пусть имеется выборка $D = \{(x_i, y_i)\}$, $i = 1, \dots, \ell$, причем для каждой пары (x_i, y_i) известно распределение вероятности отклонений $\Delta_i x$ и $\Delta_i y$ независимой и зависимой переменных $P(\Delta_i x) = P_i^x(x - x_i)$ и $P(\Delta_i y) = P_i^y(y - y_i)$ от значений x_i и y_i соответственно. Считается, что математическое ожидание отклонения равно нулю, и, таким образом, реализуемые в эксперименте значения x_i и y_i принимаются средними. Вероятности $P_i^x(x - x_i)$ и $P_i^y(y - y_i)$ обычно принимаются гауссовыми, и для них считаются известными значения дисперсий σ_i^x, σ_i^y .

Рассматривается параметрическая регрессионная модель $y(x, \omega)$ (например индуктивно порожденная [5] или предложенная экспертом), вектор параметров ω которой минимизирует некоторый функционал S , например среднеквадратичное отклонение:

$$S = \sum_{i=1}^{\ell} (y(x_i, \omega) - y_i)^2 \xrightarrow{\omega} \min. \quad (3)$$

Для таким образом определенного функционала, а также для его модификаций, учитывающих сложность регрессионной модели [5], процедура минимизации эффективно проводится с помощью алгоритма Левенберга–Марквардта (АЛМ) [7, 8].

Далее фиксируем структурный вид полученной зависимости $y(x, \omega)$ и многократно повторяем следующую вычислительную процедуру:

1. На k -м шаге генерируется случайная выборка $D_k = \{(x_i^k, y_i^k)\}$, $i = 1, \dots, \ell$. Значение x_i^k получается из соответствующего значения x_i исходной выборки D путем добавления случайного шума, распределенного согласно P_i^x :

$$x_i^k = x_i + \xi_i^k, \quad \xi_i^k \sim P_i^x.$$

Аналогично получается y_i^k .

2. Для таким образом построенного набора данных D_k (далее реализация), используя один и тот же алгоритм оптимизации, находим оптимальный (минимизирующий выбранный функционал) набор ω^k коэффициентов регрессии $y(x, \omega)$ для k -й реализации.

Таким образом, для каждого конкретного коэффициента регрессии ω_p получаем совокупность его значений в сгенерированных реализациях $\{\omega_p^k\}$.

3. Для достаточно большого числа реализаций M обычным образом определим среднее значение и стандартное отклонение соответствующего коэффициента регрессии ω_p :

$$\bar{\omega}_p = \frac{\sum_{i=1}^M \omega_p^i}{M}, \quad (4)$$

$$D(\omega_p) = \sigma_{\omega_p}^2 = \frac{1}{M-1} \sum_{i=1}^M (\omega_p^i - \bar{\omega}_p)^2. \quad (5)$$

Наша гипотеза состоит в том, что полученные согласно (4) и (5) значения соответствуют реальности. Предлагаемый подход к определению погрешности регрессионных коэффициентов, очевидно, представляет собой фактически применение метода типа Монте-Карло к задаче регрессии.

Из предложенной интерпретации также следует очевидный критерий останова вычислительной процедуры, когда с ростом числа реализаций M вариация значений $\bar{\omega}_p$ и $D(\omega_p)$ становится меньше экспертно выбранного значения.

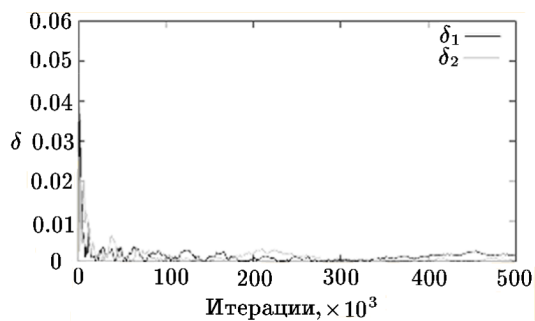
Необходимо заметить, что в общем случае пределы выражений типа (4) и (5) при $M \rightarrow \infty$ могут и не существовать, что делает предложенную вычислительную схему некорректной. Однако для достаточно гладких функций, которые собственно и представляют практический интерес, корректность предложенной процедуры можно строго доказать, что, однако, выходит за рамки данной работы.

3. Модельный случай

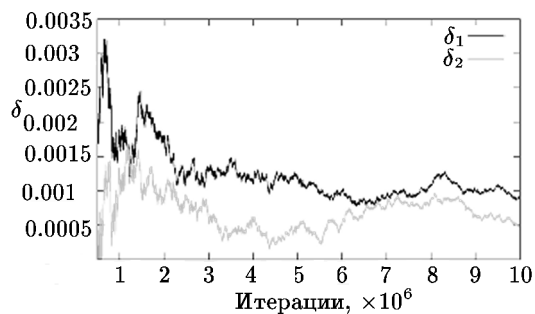
Проверим предлагаемый метод на случае с известным аналитическим решением (1), (2), рассчитав погрешности коэффициентов линейной зависимости с точно известной независимой переменной и гауссовым распределением ошибки зависимой переменной, при этом параметры распределения погрешности зависимой переменной одинаковы для каждой экспериментальной точки. В вычислительном эксперименте моделируемая зависимость имела вид $y = ax + b$, $a = 3$, $b = 10$. Независимая переменная определена в $\ell = 10$ (в другом эксперименте $\ell = 50$) точках отрезка $[0, 10]$, количество реализаций составляло 100 миллионов. Оптимизация коэффициентов регрессии проводилась с помощью алгоритма Левенберга–Марквардта по аналогии с более общим случаем нелинейной регрессии.

На рис. 1–3 по оси абсцисс указано количество реализаций N , а по оси ординат для каждого из коэффициентов a , b — отношение $\delta = \frac{\sigma_{\text{в.э.}} - \sigma_{\text{т}}}{\sigma_{\text{т}}}$, где $\sigma_{\text{в.э.}}$ — значение дисперсии, полученное в вычислительном эксперименте, а $\sigma_{\text{т}}$ — точное теоретическое значение дисперсии согласно (1), (2).

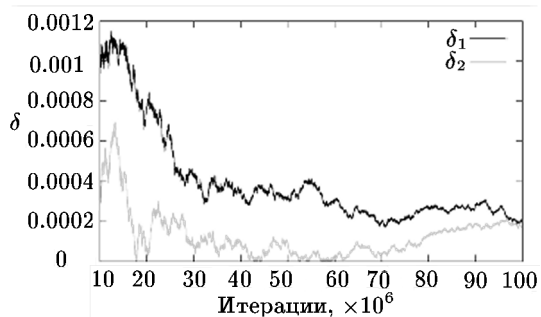
Как видно, уже для $N \approx 2.5 \cdot 10^7$ относительное различие $\delta = \frac{\sigma_{\text{в.э.}} - \sigma_{\text{т}}}{\sigma_{\text{т}}}$ не превышает 0.05% вне зависимости от количества точек, в которых определена независимая переменная, и значения дисперсии случайной величины. Это, на наш взгляд, является хорошим результатом и свидетельствует в пользу корректности обсуждаемого подхода и основной гипотезы.



(a) $N \in [0; 5 \cdot 10^5]$

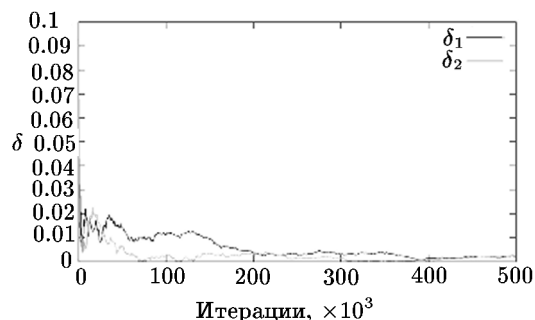


(б) $N \in [5 \cdot 10^5; 10^7]$

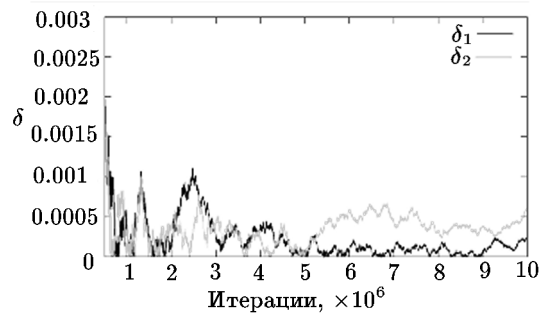


(в) $N \in [10^7; 10^8]$

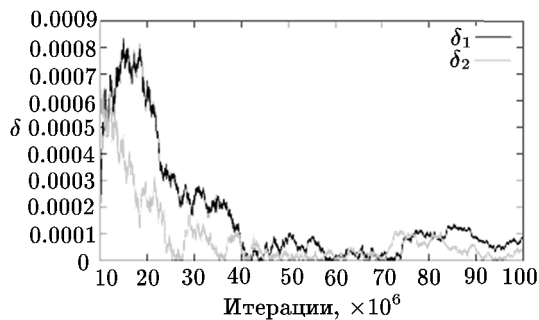
Рис. 1. Зависимость δ от числа итераций N при $D(\xi) = 10$ и $\ell = 10$



(a) $N \in [0; 5 \cdot 10^5]$



(б) $N \in [5 \cdot 10^5; 10^7]$



(в) $N \in [10^7; 10^8]$

Рис. 2. Зависимость δ от числа итераций N при $D(\xi) = 1$ и $\ell = 10$

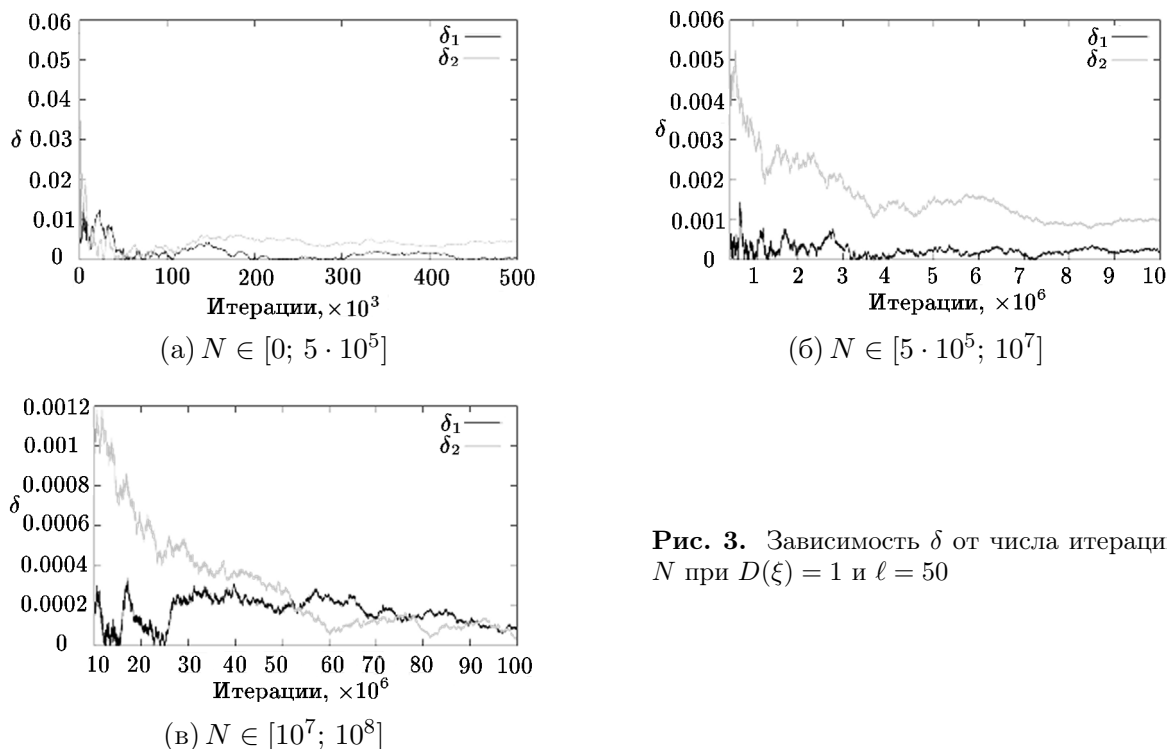


Рис. 3. Зависимость δ от числа итераций N при $D(\xi) = 1$ и $\ell = 50$

4. Погрешность регрессионных коэффициентов зависимости $n(\lambda)$

В эксперименте при 17 значениях длины волны был измерен показатель преломления прозрачного (в исследуемом спектральном диапазоне) полимера, результаты представлены в табл. 1.

Таблица 1. Экспериментальные значения коэффициентов преломления

λ , нм	$n(\lambda)$	λ , нм	$n(\lambda)$	λ , нм	$n(\lambda)$
435.8	1.35715	577.0	1.34968	750	1.34607
447.1	1.35625	587.6	1.34946	800	1.34544
471.3	1.35449	589.3	1.34938	850	1.34487
486.1	1.35349	656.3	1.34768	900	1.34437
501.6	1.35275	667.8	1.34740	950	1.34407
546.1	1.35083	706.5	1.34664		

Абсолютная погрешность измерения показателя преломления σ_n составляет обычно $(3 \div 10) \cdot 10^{-5}$ во всем диапазоне длин волн (λ , нм), относительная погрешность определения длины волны σ_λ/λ для используемых дифракционных приборов составляет $(3 \div 30) \cdot 10^{-4}$.

Для предложенной экспертом модели зависимости $n(\lambda)$ требуется определить погрешность коэффициентов регрессии, обусловленную экспериментальными погрешностями измерения длины волны и показателя преломления.

Кроме того, целесообразно найти зависимость погрешности коэффициентов регрессии от точности эксперимента, что имеет не только теоретический, но и практический интерес. Например, улучшение точности измерения показателя преломления n до $2 \cdot 10^{-5}$ требует значительных усилий (и затрат), но если при этом погрешность определения коэффициентов регрессионной модели практически не изменится, то соответствующая

работа по модернизации экспериментальной установки не имеет смысла. Точно так же при улучшении точности определения показателя преломления имеет не только теоретический, но и практический смысл вопрос о том, насколько требуется улучшить точность определения длины волны, чтобы обеспечить значительное снижение погрешности определения коэффициентов регрессии.

Экспертом была предложена функциональная зависимость следующего вида:

$$n(\lambda) = a + \frac{b}{\lambda^2} + \frac{c}{\lambda^4}. \quad (6)$$

Для указанного вида нелинейной регрессии с помощью АЛМ были определены оптимальные коэффициенты регрессии, обеспечивающие среднеквадратичную ошибку $\approx 3.97 \cdot 10^{-9}$ при $a = 1.33344$; $b = 2841.63$; $c = 1599.27$; длина волны измеряется в нанометрах.

Вычисление погрешностей коэффициентов регрессионной модели было проведено для ряда комбинаций погрешностей определения длины волны (в относительных единицах) и показателя преломления (также в относительных единицах), в каждом случае проводилась статистическая обработка миллиона реализаций. Некоторые результаты вычислительного эксперимента приведены в табл. 2–4.

Таблица 2. Относительная погрешность определения коэффициента регрессии a

$\frac{\sigma_\lambda}{\lambda} \backslash \frac{\sigma_n}{n}$	$2 \cdot 10^{-5}$	$1 \cdot 10^{-4}$	$3 \cdot 10^{-4}$
10^{-3}	$1.255 \cdot 10^{-5}$	$5.54 \cdot 10^{-5}$	$1.656 \cdot 10^{-4}$
10^{-2}	$6.32 \cdot 10^{-5}$	$8.24 \cdot 10^{-5}$	$1.753 \cdot 10^{-4}$

Таблица 3. Относительная погрешность определения коэффициента регрессии b

$\frac{\sigma_\lambda}{\lambda} \backslash \frac{\sigma_n}{n}$	$2 \cdot 10^{-5}$	$1 \cdot 10^{-4}$	$3 \cdot 10^{-4}$
10^{-3}	$2.14 \cdot 10^{-3}$	$8.68 \cdot 10^{-3}$	$2.54 \cdot 10^{-2}$
10^{-2}	$1.36 \cdot 10^{-2}$	$1.59 \cdot 10^{-2}$	$2.85 \cdot 10^{-2}$

Таблица 4. Относительная погрешность определения коэффициента регрессии c

$\frac{\sigma_\lambda}{\lambda} \backslash \frac{\sigma_n}{n}$	$2 \cdot 10^{-5}$	$1 \cdot 10^{-4}$	$3 \cdot 10^{-4}$
10^{-3}	1.15	1.426	1.454
10^{-2}	1.45	1.456	1.47

Для $\frac{\sigma_\lambda}{\lambda} < 10^{-4}$ погрешности первого и второго коэффициентов регрессии очень слабо зависят от дальнейшего увеличения точности определения длины волны и почти линейно зависят от точности измерения показателя преломления. Абсолютная погрешность определения третьего коэффициента регрессии высока и сравнима по порядку величины с самим коэффициентом, что, очевидно, связано с незначительным вкладом третьего слагаемого в (6) в показатель преломления: изменение этого слагаемого слабо влияет на общую величину среднеквадратичного отклонения, следовательно, этот компонент определяется с низкой точностью.

5. Заключение

1. Предложенный в настоящей работе метод определения погрешностей коэффициентов регрессии может использоваться при любом распределении вероятности ошибки зависимых и независимых переменных, включая различные виды распределения ошибки для разных переменных, а также для одной переменной в различных экспериментальных точках.

2. Предложенный метод будет давать несколько разные значения погрешностей при использовании различных функционалов ошибки S . Такая ситуация не является необычной [5], и, например, коэффициенты регрессии будут разными, если по стандартному методу наименьших квадратов (МНК) минимизируется сумма квадратов расстояний по оси, соответствующей зависимой переменной, от экспериментальных точек до аппроксимирующей регрессионной модели, согласно (3), или же минимизируется сумма евклидовых расстояний от экспериментальных точек до аппроксимирующей модели аналогично известной работе Пирсона [9].

С нашей точки зрения, выбор оптимального функционала G проводится экспертом, исходя, прежде всего, из анализа погрешностей определения зависимых и независимых переменных. В самом деле, если независимая переменная определяется точно, то МНК предпочтителен, а при близких погрешностях переменных подход Пирсона представляется предпочтительным.

3. Для оптимизации коэффициентов регрессии использовался АЛМ в стандартной форме, когда минимизируется функционал типа (3). Как указано выше, такой подход корректен, когда погрешность определения независимой переменной мала. В общем случае корректнее минимизировать суммы квадратов регрессионных остатков, соответствующих расстоянию до регрессионной кривой (см. [9]) с учетом погрешностей измеряемых переменных, а не расстоянию до точки на кривой с фиксированной координатой x , что корректно при отсутствии погрешности измерения независимой переменной. Для достаточно гладких регрессионных зависимостей, когда в окрестности (x_i, y_i) зависимость можно линеаризовать, соответствующее расстояние приобретает вид $\frac{(y_i - f(x_i, \omega))^2}{(\sigma_i^y)^2 + k^2(\sigma_i^x)^2}$, где $k = \frac{\partial f}{\partial x}|_{x=x_i}$. Иными словами, “стандартный” регрессионный остаток нормируется на некоторую величину, зависящую от погрешностей измерений σ_i^x и σ_i^y и наклона регрессионной кривой в окрестности рассматриваемой точки (x_i, y_i) . Соответствующая модификация АЛМ является предметом дальнейшей работы.

В то же время для конкретных вычислительных экспериментов, результаты которых представлены в настоящей работе, использование немодифицированного АЛМ корректно. В самом деле, для рассмотренного в п. 3 модельного случая заранее предполагается, что независимая переменная известна точно, $\forall i : \sigma_i^x = 0$, а погрешности определения зависимой переменной постоянны: $\sigma_i^y = \text{const}$. Тогда знаменатель приведенного выше выражения для расстояния вырождается в одну и ту же константу, на которую нормируется каждый регрессионный остаток, что полностью соответствует условию применимости точного результата (1), (2).

В рассмотренном в четвертом пункте случае анализа регрессионной зависимости дисперсии полимера для реализуемых погрешностей σ_n и σ_λ (σ^y и σ^x соответственно) во всем спектральном диапазоне $k = \frac{\partial n}{\partial \lambda} \leq \frac{1}{5} \frac{\sigma_n}{\sigma_\lambda}$, что позволяет использовать стандартный АЛМ с достаточной точностью.

Таким образом, в настоящей работе предложен метод определения погрешностей коэффициентов регрессионной модели, включая случай нелинейной регрессии. В вычислительном эксперименте получено хорошее соответствие теоретически описанному случаю, когда стандартное отклонение коэффициентов регрессии определяется точно. Показана целесообразность применения предложенного вычислительного алгоритма при анализе конкретных физических экспериментов.

Литература

1. **Davidson J.W., Savic D.A., and Walters G.A.** Symbolic and numerical regression: experiments and applications // *Developments in Soft Computing* / John Robert, Birkenhead Ralph. — Heidelberg, Germany: Physica Verlag, 2001. — Vol. 9. — P. 175–182.
2. **Sammut C., Webb G.I.** Symbolic Regression // *Encyclopedia of Machine Learning* / C. Sammut, G.I. Webb — Berlin: Springer-Verlag, 2010. — <http://dx.doi.org/10.1007/978-0-387-30164-8>
3. **Strijov V., Weber G.W.** Nonlinear regression model generation using hyperparameter optimization // *Mathematics with Applications*. — 2010. — Vol. 60, № 4. — P. 981–988. — <http://dx.doi.org/10.1016/j.camwa.2010.03.021>
4. **Стрижов В.В.** Методы индуктивного порождения регрессионных моделей. — Москва, 2008. — (Препринт / ВЦ РАН им. А.А. Дородницына).
5. **Рудой Г.И., Стрижов В.В.** Алгоритмы индуктивного порождения суперпозиций для аппроксимации измеряемых данных // *Информатика и ее применения*. — 2013. — Vol. 7, № 1. — С. 44–53.
6. **Ватутин В.А., Ивченко Г.И., Медведев Ю.И., Чистяков В.П.** Теория вероятностей и математическая статистика в задачах. Изд-ние 3. — М.: Дрофа, 2005.
7. **Marquardt D.W.** An algorithm for least-squares estimation of non-linear parameters. // *J. of the Society of Industrial and Applied Mathematics*. — 1963. — Vol. 11, № 2. — P. 431–441.
8. **More J.J.** The Levenberg-Marquardt algorithm: Implementation and theory // *Numerical Analysis* / G.A. Watson. — Berlin: Springer-Verlag, 1978. — P. 105–116. — (Lect. Notes in Mathematics; 630).
9. **Pearson K.** On lines and planes of closest fit to systems of points in space // *Philosophical Magazine*. — 1901. — Vol. 2, № 6. — P. 559–572.

*Поступила в редакцию 4 июня 2014 г.,
в окончательном варианте 24 марта 2015 г.*

Литература в транслитерации

1. **Davidson J.W., Savic D.A., and Walters G.A.** Symbolic and numerical regression: experiments and applications // *Developments in Soft Computing* / John Robert, Birkenhead Ralph. — Heidelberg, Germany: Physica Verlag, 2001. — Vol. 9. — P. 175–182.
2. **Sammut C., Webb G.I.** Symbolic Regression // *Encyclopedia of Machine Learning* / C. Sammut, G.I. Webb — Berlin: Springer-Verlag, 2010. — <http://dx.doi.org/10.1007/978-0-387-30164-8>
3. **Strijov V., Weber G.W.** Nonlinear regression model generation using hyperparameter optimization // *Mathematics with Applications*. — 2010. — Vol. 60, № 4. — P. 981–988. — <http://dx.doi.org/10.1016/j.camwa.2010.03.021>
4. **Strijov V.V.** Metody induktivnogo porozhdeniya regressionnykh modelej. — Moskva, 2008. — (Preprint / VTs RAN im. A.A. Dorodnitsyna).
5. **Rudoy G.I., Strijov V.V.** Algoritmy induktivnogo porozhdeniya superpozitsij dlya approksimatsii izmeryaemykh dannykh // *Informatika i ee primeneniya*. — 2013. — Vol. 7, № 1. — С. 44–53.
6. **Vatutin V.A., Ivchenko G.I., Medvedev Yu.I., Chistyakov V.P.** Teoriya veroyatnostej i matematicheskaya statistika v zadachakh. Izd-nie 3. — М.: Дрофа, 2005.

7. **Marquardt D.W.** An algorithm for least-squares estimation of non-linear parameters. // J. of the Society of Industrial and Applied Mathematics. — 1963. — Т. 11, № 2. — С. 431–441.
8. **More J.J.** The Levenberg-Marquardt algorithm: Implementation and theory // Numerical Analysis / G.A. Watson. — Berlin: Springer-Verlag, 1978. — P. 105–116. — (Lect. Notes in Mathematics; 630).
9. **Pearson K.** On lines and planes of closest fit to systems of points in space // Philosophical Magazine. — 1901. — Vol. 2, № 6. — P. 559–572.