ВЕСТНИК ТОМСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА

2008 Математика и механика № 2(3)

УДК 519.6

О.П. Федорова

ОБ ОДНОМ ПОДХОДЕ К ПРИБЛИЖЕНИЮ ФУНКЦИИ СПЛАЙНАМИ

В настоящей работе построен сплайн, который аппроксимирует функцию одной переменной. Коэффициенты сплайна выбраны так, что интегралы по области задания функции и сплайна совпадают. Оценена погрешность сплайна. Приведены результаты численного эксперимента.

Ключевые слова: *Аппроксимация сплайнами, приближение функции, оценивание погрешности приближения.*

Пусть задана функция $f(x)$, $x \in D$, обладающая свойством

$$\int_D f(x)\,dx = 1 .$$

Приблизим функцию сплайнами так, чтобы значения определенного интеграла по области задания приближаемой функции и сплайна совпадали. Будем использовать аппроксимацию сплайнами дефекта один.

На сетке $\Delta: -\infty < x_{-m} < \ldots < x_0 < x_1 < \ldots < x_n < x_{n+1} < \ldots < x_{n+m} < \infty$, сплайн $S(x)$ степени m на отрезке $[x_0; x_n]$ выражается линейной комбинацией нормализованных B-сплайнов:

$$S(x) = \sum_i z_i(f) B_i(x), \tag{1}$$

где $z_i(f)$ – последовательность линейных функционалов, вид которых определяет вид приближения. Выберем $z_i(f)$, так чтобы значения определенного интеграла по области задания приближаемой функции и сплайна были равными:

$$\int_D S(x)\,dx = \int_D f(x)\,dx .$$

Пусть область D совпадает с отрезком $[x_0; x_n]$, т.е. $D = [x_0; x_n]$. Тогда

$$\int_D S(x)\,dx = \sum_{i=0}^{n} \int_{x_i}^{x_{i+1}} S(x)\,dx, \text{ а } \int_D f(x)\,dx = \sum_{i=0}^{n} \int_{x_i}^{x_{i+1}} f(x)\,dx .$$

Исходя из целей построения сплайна, потребуем, чтобы

$$\int_{x_i}^{x_{i+1}} S(x)\,dx = \int_{x_i}^{x_{i+1}} f(x)\,dx . \tag{2}$$

На промежутке $[x_i; x_{i+1}]$ кубический сплайн $S(x)$ дефекта 1 (1) запишется в виде

$$S(x) = \sum_{p=i-1}^{i+2} z_p B_p(x). \tag{3}$$

Подставим $S(x)$ в виде (3) в формулу (2) и получим для нахождения коэффициентов z_i систему линейных уравнений

$$a_i z_{i-1} + b_i z_i + c_i z_{i+1} + d_i z_{i+2} = \int\limits_{x_i}^{x_{i+1}} f(x)\,dx, \quad i = \overline{0, n-1}. \tag{5}$$

Здесь

$$a_i = \int\limits_{x_i}^{x_{i+1}} B_{i-1}(x)\,dx,\ b_i = \int\limits_{x_i}^{x_{i+1}} B_i(x)\,dx,\ c_i = \int\limits_{x_i}^{x_{i+1}} B_{i+1}(x)\,dx,\ d_i = \int\limits_{x_i}^{x_{i+1}} B_{i+2}(x)\,dx.$$

Для определения n+3 коэффициентов сплайна имеется n условий вида (2). Дополнительные условия можно задать различным способом, исходя из знаний о свойствах приближаемой функции $f(x)$. Например, потребуем, чтобы значение сплайна или его r-й производной и соответственно значения функции или ее производной совпадали в трех точках из четырех: x_0, x_1, x_{n-1}, x_n.

$$\sum_{p=-1}^{1} z_p B_p^{(r)}(x_k) = f^{(r)}(x_k), \quad k = 0,\ 1,\ n-1,\ n. \tag{6}$$

Система (5), (6) имеет четырехдиагональную матрицу коэффициентов, решение которой будем искать методом прогонки в виде

$$z_i = W_i z_{i+1} + Q_i z_{i+2} + R_i.$$

Здесь $$W_i = -\frac{c_i + Q_{i-1}\cdot a_i}{b_i + W_{i-1}\cdot a_i},\ Q_i = -\frac{d_i}{b_i + W_{i-1}\cdot a_i},\ R_i = \frac{p_i - R_{i-1}\cdot a_i}{b_i + W_{i-1}\cdot a_i}.$$

Рассмотрим функцию $R(x) = S(x) - f(x)$ – погрешность приближения. Если функция $f(x) \in C^2[D]$ и сплайн – кубический дефекта 1, то погрешность $R(x)$ есть дважды непрерывно дифференцируемая функция. По условию (2)

$$\int\limits_{x_i}^{x_{i+1}} S(x)\,dx - \int\limits_{x_i}^{x_{i+1}} f(x)\,dx = \int\limits_{x_i}^{x_{i+1}} R(x)\,dx = 0.$$

Или $$\int\limits_{x_i}^{x_{i+1}} R(x)\,dx = R(\tilde{x})\cdot h_i = 0.$$

Из этого равенства следует, что $\exists\, \tilde{x}_i \in [x_i; x_{i+1}]$ такая, что $R(\tilde{x}_i)$=0, т.е. на каждом промежутке существует хотя бы одна точка, в которой значение сплайна $S(x)$ и функции $f(x)$ совпадают. Запишем для $\forall x \in [x_i; x_{i+1}]$ формулу Тейлора для погрешности $R(x)$= $R(\tilde{x}_i + \delta_i)$:

$$R(x) = R(\tilde{x}_i) + R'(\theta_i)\delta_i, \tag{7}$$

где $\theta_i \in [x_i; x_{i+1}]$. Формулу (7), учитывая что $R(\tilde{x}_i)$=0, каждого промежутка можно записать в виде $R(x)$= $R'(\theta_i)\delta_i$. Таким образом, для остаточного члена аппроксимации в области D выполняется оценка

$$|R(x)| \le M\delta, \tag{8}$$

здесь $M = \max\limits_i |\theta_i|$, $\delta = \max\limits_i |\delta_i|$. Для построения оценки погрешности приближения можно использовать следующий подход. Построим кубический сплайн $\tilde{S}(x)$ по точкам $\tilde{x}_i$, этот сплайн будет интерполирующим как для функции $f(x)$, так и для сплайна $S(x)$. При выбранных граничных условиях выполняются следующие неравенства [1]:

$$\left\| f(x)-\tilde{S}(x) \right\| \le \frac{19}{96}\bar{h}^2 \omega(f''). \tag{9а}$$

Здесь $\omega(f'')=\max\limits_i \omega_i(f'')$, $\omega_i(f'')=\max\limits_{\xi,\,\eta\in[x_i;x_{i+1}]} |f''(\xi)-f''(\eta)|$,

$$\bar{h}=\max_i h_i;\ i=0,\,1,\,\ldots,\,n-1;$$

$$\left\| \tilde{S}(x)-S(x) \right\| \le \frac{19}{96}\bar{h}^2 \omega(S''). \tag{9б}$$

Для $R(x)$, используя неравенства (9а) и (9б), построим оценку

$$\left\| f(x)-S(x) \right\| \le \frac{19}{96}\bar{h}^2 \Omega,$$

где $\Omega=\max\{\omega(f''),\omega(S'')\}$.

Если сетка является равномерной, то на промежутке $[x_i;x_{i+1}]$ отличные от нуля B-сплайны имеют вид

$$B_{i-1}(x)=(1-t)^3/6,$$

$$B_i(x)=[1+3\cdot(1-t)+3\cdot t\cdot(1-t)^2]/6,$$

$$B_{i+1}(x)=[1+3\cdot t+3\cdot t^2\cdot(1-t)]/6,$$

$$B_{i+2}(x)=t^3/6. \tag{10}$$

Здесь $t=(x-x_i)/(x_{i+1}-x_i)$ и, с учетом выражений для B-сплайнов (10), получим

$$\int\limits_{x_i}^{x_{i+1}} f(x)\,dx=\frac{h}{24}(z_{i-1}+11z_i+11z_{i+1}+z_{i+2}),\ i=\overline{0,n-1}, \tag{11}$$

где $h=x_{i+1}-x_i$ – шаг равномерной сетки. Дополним систему (10) уравнениями:

$$\frac{1}{6}z_{-1}+\frac{2}{3}z_0+\frac{1}{6}z_1=f(x_0),$$

$$\frac{1}{6}z_{n-1}+\frac{2}{3}z_n+\frac{1}{6}z_{n+1}=f(x_n),$$

$$\frac{1}{6}z_{n-2}+\frac{2}{3}z_{n-1}+\frac{1}{6}z_{n+1}=f(x_{n-1}). \tag{12}$$

Выразим из первого уравнения системы (12) $z_{-1}=-4z_0-z_1+f(x_0)$, подставим в уравнение (11) при $i=0$ и получим следующие выражения для $W_0=-10/7$, $Q_0=-1/7$, $R_0=\dfrac{24}{7h}\int\limits_{x_0}^{x_1} f(x)\,dx-f(x_0)/7$. Для получения решения на правой границе введем обозначения $\alpha=1/6,\beta=2/3,\gamma=1/6$ и запишем систему уравнений на правой границе в виде

$$\alpha \cdot z_{n-1} + \beta \cdot z_n + \gamma \cdot z_{n+1} = f(x_n),$$
$$\alpha \cdot z_{n-2} + \beta \cdot z_{n-1} + \gamma \cdot z_{n+1} = f(x_{n-1}),$$
$$z_{n-2} = W_{n-2} z_{n-1} + Q_{n-2} z_n + R_{n-2},$$
$$z_{n-1} = W_{n-1} z_n + Q_{n-1} z_{n+1} + R_{n-1}. \qquad (13)$$

Первое уравнение системы (13) добавим к (5), зададим значения $a_n = \alpha$, $b_n = \beta$, $c_n = \gamma$, $d_n = 0$ и получим (n+1)-е уравнение системы (5) в виде

$$a_n z_{n-1} + b_n z_n + c_n z_{n+1} + d_n z_{n+2} = f(x_n).$$

Вычислим коэффициенты W_i, Q_i, R_i для $i = 1,2,\ldots,n$. Из системы (13) выразим $z_{n+1} = U/V$:

$$U = f(x_{n-1}) - \alpha(W_{n-1}W_{n-2}R_n + W_{n-2}R_{n-1} + Q_{n-2}R_n) - \beta(W_{n-1}R_n + R_{n-1}) - \gamma R_n,$$

$$V = \alpha(W_{n-2}W_{n-1}W_n + W_{n-1}Q_{n-1} + W_n Q_{n-2}) + \beta(W_{n-1}W_n + Q_{n-1}) + \gamma W_k.$$

Будем рассматривать сплайн как плотность некоторого распределения, тогда для первого начального момента $\mu_1 = \int\limits_D x \cdot S(x)dx$, интегрируя (1) для случая равноотстоящих узлов, получаем выражение

$$\mu_1 = \sum_{i=0}^{n-1}\left(z_{i-1}\left(\frac{h^2}{120} + \frac{x_i h}{24}\right) + z_i\left(\frac{11h^2}{60} + \frac{11x_i h}{24}\right)\right) + \sum_{i=0}^{n-1}\left(z_{i+1}\left(\frac{11h^2}{40} + \frac{11x_i h}{24}\right) + z_i\left(\frac{h^2}{30} + \frac{x_i h}{24}\right)\right).$$

Применим построенный сплайн для аппроксимации функции плотности вероятностей, например, стандартного нормального распределения

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \qquad (14)$$

Здесь $x \in (-\infty; \infty)$. Однако, $\int\limits_{-6}^{6} f(x)dx$ = 0,999999998026825, если $x_0 = -6$; $x_n = 6$, поэтому полагаем область D = [– 6; 6].

В табл. 1 даны максимальные отклонения функции $f(x)$ от сплайна $S(x)$ для различных значений числа разбиений n.

Таблица 1

Максимальное отклонение значений функции (13) от значений сплайна для различных *n*

Число разбиений	10	20	30	50
Максимальное отклонение	0,010041	0,000353	0,000045	0,000002

Проведенные расчеты показали, что значения интеграла от $S(x)$ по области D = [– 6; 6] отклоняются от единицы менее чем на 10^{-7}, при этом интегралы $\int\limits_{x_i}^{x_{i+1}} f(x)dx$ вычислялись с точностью порядка 10^{-8}. Значение μ_1, вычисляемое по формуле (12), равно нулю также с точностью порядка 10^{-7}. Заметим, что функция

(14) при $|x| \to \infty$ асимптотически приближается к нулю, поэтому для верификации применяемого способа задания дополнительных условий приведем результаты, построения сплайна для функции

$$g(x) = 0{,}75x^{0{,}25} - 2{,}5x^3 - 0{,}5x^2 + 1$$

на промежутке D = [0; 4]. Значения функции $g(x)$ на границах не равны нулю: $g(0) = 1$ и $g(4) = 25$, интеграл по области D = [0; 4]: $\int\limits_D g(x)\,dx = -13{,}06666(6)$, $\int\limits_D x \cdot g(x)\,dx = -24$. В табл. 2 приведены величины максимального отклонения сплайна от функции $\max\limits_{x \in D}|S(x) - g(x)|$, интеграла от сплайна $\int\limits_D S(x)\,dx$ и $\hat{\mu}_1 = \int\limits_D x \cdot S(x)\,dx$, вычисленные при нескольких n.

Таблица 2

Результаты расчетов для функции $g(x)$ при различных n

Число разбиений	10	20	30	50
Максимальное отклонение	0,005066	0,001278	0,000589	0,000220
Интеграл от сплайна по промежутку [0;4]	– 13,064400	– 13,066100	– 13,66415	– 13,066576
$\hat{\mu}_1$	– 23,991763	– 23,997935	– 23,999082	– 23,999669

Общее решение системы разностных уравнений четвертого порядка с постоянными коэффициентами (11)

$$z_i = c_1(-1)^i + c_2\left(-5 + 2\sqrt{6}\right)^i + c_3\left(-5 - 2\sqrt{6}\right)^i + z_i^*.$$

Здесь z_i^* – частное решение системы (11).

Рассмотрим применение предлагаемого сплайна к приближению эмпирической плотности вероятности, построенной по модельным данным. В работе [2] предложено семейство линейных конгруэнтных методов генерирования базовой случайной величины.

В общем случае линейный конгруэнтный метод определяется рекуррентной последовательностью вида

$$a_n = a_n^* / m, \text{ где } a_n^* = (\beta a_{n-1}^* + c) \bmod (m),\ n = 1, 2, \ldots \qquad (15)$$

Эта последовательность является циклической. Согласно (15), $a_n^* \in \{0,1,\ldots,m\}$, поэтому период $T \le m$. Максимальная длина периода $T = m$ достижима, об этом говорит следующая теорема [2].

Теорема 1. Максимально возможная величина периода последовательности (15) равна m и достигается тогда и только тогда, когда:

1) c и m взаимно просты,

2) $\beta-1$ кратно p для всех простых p, являющихся делителем m,

3) $\beta-1$ кратно четырем, если m кратно четырем.

Для проведения численного эксперимента будем использовать генераторы базовой случайной величины с параметрами $c = 51$, $x_0 = 5$, $\beta = 101$ и $c = 7$, $x_0 = 2$, $\beta = 81$ и длиной периода $m = 1\,000\,000$. Методом полярных координат построим последовательность случайных чисел, имеющих стандартное нормальное распределение. Принадлежность используемых для численного эксперимента выборок к нормальному распределению проверялась при помощи критерия согласия χ^2.

Анализ результатов численного эксперимента показал, что сплайн, приближающий эмпирическое распределение, демонстрирует сходимость по норме пространства $C[-6;6]$ к плотности вероятности стандартного нормального распределения (рис. 1).

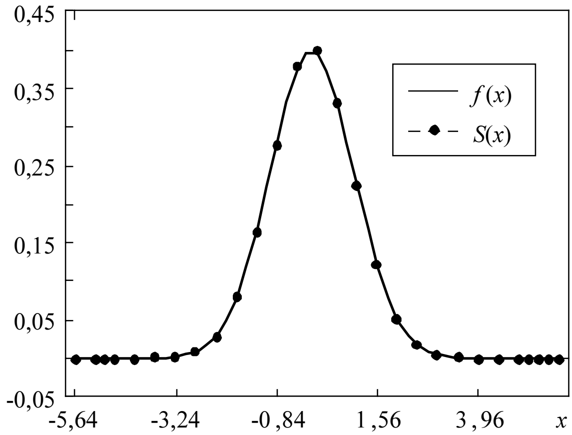

Рис. 1. Сплайн-аппроксимация $S(x)$ эмпирической плотности вероятности и график плотности стандартного нормального распределения

В табл. 3 показаны значения $\max\limits_{x\in[-6;6]} |S(x) - f(x)|$, вычисляемые при различных n.

Таблица 3

Объем выборки	$n = 10$	$n = 30$	$n = 100$
5000	0,008653	0,033278	0,078358
50000	0,004839	0,009868	0,022497
500000	0,000846	0,003912	0,008984
1000000	0,000862	0,003321	0,004401

Здесь $f(x)$ – функция плотности вероятностей стандартного нормального распределения. По построению, $S(x)$ для каждого разбиения есть плотность некоторого распределения. Задавая увеличивающееся число разбиений k, получаем последовательность функций $S_k(x)$, которая отклоняется от эмпирической плотности на величину порядка $O(h^2)$, т.е. при увеличении объема выборки и увеличении числа разбиений так, что $\overline{h} \to 0$, сплайн-функция сходится к плотности стандартного нормального распределения со вторым порядком малости относительно $\overline{h}$.

ЛИТЕРАТУРА

1. *Завьялов Ю.С., Квасов Б.И., Мирошниченко В.Л.* Методы сплайн-функций. М.: Наука, 1980. 352 с.
2. *Кнут Д.* Искусство программирования для ЭВМ: В 3 т. М.: Мир, 1977. Т. 2: Получисленные алгоритмы. 728 с.

Статья принята в печать 05.07.2008 г.