

An Approximation Method of Spline Functions

Abunawas Khaled Abdallah

Department of Mathematics, Qassim University, College of Science and Arts at ArRass, ArRass, Saudi Arabia

Abstract In this paper we construct the spline which approximates the function one variable. Spline coefficients are chosen so that the integrals over the domain of the spline function and the same. estimate the error spline. The results of numerical experiments. The use of spline functions in the analysis of empirical two-dimensional data (x_i, y_i) is described. The definition of spline functions as piecewise polynomials with continuity conditions give them unique properties as empirical functions [2]. They can represent any variation of y with x arbitrarily well over wide intervals of x . Furthermore, due to the local properties of the spline functions, they are excellent tools for differentiation and integration of empirical data. Hence, spline functions are excellent empirical functions which can be used with advantage instead of other empirical functions, such as polynomials or exponentials. [1]

Keywords Approximation spline, Function approximation, Estimation error of approximation

1. Introduction

Spline functions are denned as piecewise polynomials of degree n . The pieces join in the so called knots and fulfill continuity conditions for the function itself and the first $w - 1$ derivatives. Thus a spline function of degree n is a continuous function with $n - 1$ continuous derivatives.

Spline functions have lately found great use as approximating functions in mathematics and numerical analysis; to quote [2] Rice (1969, page 123): "spline functions are the most successful approximating functions for practical applications so far discovered. The reader may be unaware of the fact that ordinary polynomials are inadequate in many situations. This is particularly the case when one approximates functions which arise from the physical world rather than from the mathematical world. Functions which express physical relationships are frequently of a disjointed or disassociated nature. That is to say that their behavior in one region may be totally unrelated to their behavior in another region. Polynomials along with most other

Moreover, since the splines are everywhere represented by simple polynomials, they are computationally easy to handle; their integrals and derivatives are also spline functions, of one degree higher and lower, respectively. Also, the definition of splines in terms of polynomials has the statistically important consequence that a spline function, when fitted to data by least squares, conserves the first two moments of the data [3] (Whittaker 1926, page 307).

These nice properties make spline functions excellent also for use in curve fitting in data analysis [3] of observations with random components. However, few applications of splines in this area are found in the literature [4], except for the trivial case of moving averages (which can be seen as splines of degree zero). In our laboratory, spline functions have been used in the analysis of chemical data for some years. The present article reviews my experience with spline functions, specifically as empirical functions used in least squares curve fitting

2. Explanation of Methods

Let a function $f(x)$, $x \in D$ with the property $\int_D f(x)dx = 1$.

We approximate spline function so that the value of the definite integral on the domain of the function being approximated spline and the same. We use the approximation of a spline defect.

On the grid $\Delta: -\infty < x_{-m} < \dots < x_0 < x_1 < \dots < x_n < x_{n+1} < \dots < x_{n+m} < \infty$ spline $S(x)$ degrees of m on the segment $[x_0, x_n]$ It is expressed by a linear combination of normalized B -splines:

$$S(x) = \sum_i z_i(f) B_i(x), \quad (1)$$

where $z_i(f)$ – a sequence of linear functional, whose form determines the type of approach.

Let us choose $z_i(f)$, so that the value the definite integral at the domain of the function to be approximated,

* Corresponding author:

xsx6666@yahoo.com (Abunawas Khaled Abdallah)

Published online at <http://journal.sapub.org/ajms>

Copyright © 2015 Scientific & Academic Publishing. All Rights Reserved

and the spline are equal:

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

Suppose the domain D coincides with the segment $[x_0, x_n]$, that is $D = [x_0, x_n]$,

Then

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

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

Proceeding from the purposes of the spline, we require that

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

In the interval $[x_i, x_{i+1}]$ the cubic spline $S(x)$ defect (1) written in the form

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

Let us substitute $S(x)$ in the form (3) into the formula (2) and we get to find the coefficients z_i system of linear equations

$$a_i z_{i-1} + b_i z_i + c_i z_{i+1} + d_i z_{i+2} = \int_{x_i}^{x_{i+1}} f(x)dx, i = \overline{0, n-1}. \quad (4)$$

Here

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

To determine the $n+3$ spline coefficients have n of conditions of the form (2). Additional conditions can be set in a different way, based on knowledge of the properties of the approximated function $f(x)$.

For example, we require that the value of spline or r -th derivative values, respectively, and function or its derivative coincide in three of the four points: x_0, x_1, x_{n-1}, x_n .

$$\sum_{p=-1}^1 z_p B_p^{(r)}(x_k) = f^{(r)}(x_k), k = 0, 1, n-1, n \quad (5)$$

The system (4), (5) has four diagonal matrix of coefficients whose solution will be sought by the sweep method in the form

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

Here

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

Consider a function $R(x) = S(x) - f(x)$ - approximation error. If the function $f(x) \in C^2[D]$ and spline - cubic defect (1), the error $R(x)$ is a twice continuously differentiable function.

By condition (2)

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

Or

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

From this equation, it follows that $\exists \tilde{x}_i \in [x_i, x_{i+1}]$ such that $R(\tilde{x}) = 0$, that is for each interval there exists at least one point at which the value of spline $S(x)$ and the function $f(x)$ match. We write for $\forall x \in [x_i, x_{i+1}]$ Taylor formula for the error $R(x) = R(\tilde{x}_i + \delta_i)$:

$$R(x) = R(\tilde{x}_i) + R'(\theta_i) \delta_i, \quad (6)$$

Where

$\theta_i \in [x_i, x_{i+1}]$. Formula (6), taking into account that $R(\tilde{x}) = 0$, each interval can be written as $R(x) = R'(\theta_i) \delta_i$. Thus, for the remainder of approximation in the estimate of D .

$$|R(x)| \leq M \delta, \quad (7)$$

Here $M = \max_i |\theta_i|$, $\delta = \max_i |\delta_i|$. To construct error estimates of approximation, you can use the following approach. We construct cubic spline $\tilde{S}(x)$ on points \tilde{x}_i , the spline interpolating function for $f(x)$, and for the spline $S(x)$. At the selected boundary conditions, the following inequality [4]:

$$\|f(x) - \tilde{S}(x)\| \leq \frac{19}{96} \bar{h}^2 \omega(f''). \quad (8a)$$

Here

$$\begin{aligned} \omega(f'') &= \max_i \omega_i(f''), \\ \omega_i(f'') &= \omega_i(f'') = \max_{\xi, \eta \in [x_i; x_{i+1}]} |f''(\xi) - f''(\eta)|, \\ \bar{h} &= \max_i h_i; i = 0, 1, \dots, n-1; \\ \|\tilde{S}(x) - S(x)\| &\leq \frac{19}{96} \bar{h}^2 \omega(S'') \end{aligned} \quad (8b)$$

For $R(x)$, using the inequalities (8a) and (8b), we construct the evaluation

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

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

If the mesh is uniform, then the interval $[x_i; x_{i+1}]$ nonzero B -splines have the form

$$\begin{aligned} B_{i-1}(x) &= \frac{(1-t)^3}{6}, \\ B_i(x) &= \frac{[1 + 3 \cdot (1-t) + 3 \cdot t \cdot (1-t)^2]}{6}, \\ B_{i+1}(x) &= \frac{[1 + 3 \cdot t + 3 \cdot t^2 \cdot (1-t)]}{6}, \\ B_{i+2}(x) &= \frac{t^3}{6}. \end{aligned} \quad (9)$$

Here $t = \frac{(x - x_i)}{(x_{i+1} - x_i)}$ and, taking into account the expressions for B -splines (9), we obtain

$$\int_{x_i}^{x_{i+1}} f(x) dx = \frac{h}{24} (z_{i-1} + 11z_i + 11z_{i+1} + z_{i+2}), i = \overline{0, n-1}, \quad (10)$$

Where $h = x_{i+1} - x_i$ - step uniform grid.

We supplement the system (9) by the equations:

$$\begin{aligned} \frac{1}{6} z_{-1} + \frac{2}{3} z_0 + \frac{1}{6} z_1 &= f(x_0), \\ \frac{1}{6} z_{n-1} + \frac{2}{3} z_n + \frac{1}{6} z_{n+1} &= f(x_n), \\ \frac{1}{6} z_{n-2} + \frac{2}{3} z_{n-1} + \frac{1}{6} z_{n+1} &= f(x_{n-1}). \end{aligned} \quad (11)$$

We express from the first equation of the system (11) $z_{-1} = 4z_0 - z_1 + f(x_0)$, substitute in equation (10) for $i = 0$ and we get the following expression for the

$$W_0 = -\frac{10}{7}, \quad Q_0 = -\frac{1}{7}, \quad R_0 = \frac{24}{7h} \int_{x_0}^{x_1} f(x) dx - \frac{f(x_0)}{7}.$$

To get a solution at the right boundary introduce the notation $\alpha = \frac{1}{6}, \beta = \frac{2}{3}, \gamma = \frac{1}{6}$ and write a system of equations on the right border of a

$$\begin{aligned} \alpha \cdot z_{n-1} + \beta \cdot z_n + \gamma \cdot z_{n+1} &= f(x_n), \\ \alpha \cdot z_{n-2} + \beta \cdot z_{n-1} + \gamma \cdot z_{n+1} &= f(x_{n-1}), \\ z_{n-2} &= W_{n-2} z_{n-1} + Q_{n-2} z_n + R_{n-2}, \\ z_{n-1} &= W_{n-1} z_n + Q_{n-1} z_{n+1} + R_{n-1}. \end{aligned} \quad (12)$$

The first equation of (12) add to (4), enter values $a_n = \alpha, b_n = \beta, c_n = \gamma, d_n = 0$ and we get $(n+1)^{th}$ equation system (4) as $a_n z_{n-1} + b_n z_n + c_n z_{n+1} + d_n z_{n+2} = f(x_n)$.

We calculate coefficients W_i, Q_i, R_i for $i = 1, 2, \dots, n$.

From the system (12) express $z_{i+1} = \frac{U}{V}$

$$\begin{aligned} U &= f(x_{n-1}) - \alpha(W_{n-1}W_{n-2}R_n + W_{n-2}R_{n-1} + Q_{n-2}R_n) \\ &\quad - \beta(W_{n-1}R_n + R_{n-1}) - \gamma R_n, \\ V &= \alpha(W_{n-2}W_{n-1}W_n + W_{n-1}Q_{n-1} + W_nQ_{n-2}) \\ &\quad + \beta(W_{n-1}W_n + Q_{n-1}) - \gamma W_n. \end{aligned}$$

We consider a spline density distribution, then for a first start time $\mu_1 = \int_D x \cdot S(x) dx$, integrating (1) in the case of equidistant points, we get the expression

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

We apply the spline to approximate the probability density function, for example, the standard normal distribution

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \quad (13)$$

Here $x \in (-\infty, \infty)$. But, $\int_{-6}^6 f(x) dx = 0.999999998026825$, if $x_0 = -6; x_n = 6$, therefore considers the region $D = [-6, 6]$.

Table. 1 shows the maximum deviation function $f(x)$

from the spline $S(x)$ for the different values of the number of partitions of n .

Table 1. Maximum deviation of the function (12) on the values of the spline for any n

number of partitions	10	20	30	50
Maximum deviation	0,010041	0,000353	0,000045	0,000002

Carried out calculations shown that value of integral of $S(x)$ the region $D = [-6, 6]$ deviate from unity by at least 10^{-7} , While integrals $\int_{x_i}^{x_{i+1}} f(x)dx$ calculated to an

accuracy of 10^{-8} . Significance μ_1 , calculated by the formula (11) is equal to zero and with an accuracy of 10^{-7} . Note that the function (13) $|x| \rightarrow \infty$ asymptotically approaching zero, Therefore the method used for verification tasks other conditions present the results, construct a spline for the function

$g(x) = 0.75x^{0.25} - 2.5x^3 - 0.5x^2 + 1$ on the interval $D = [0; 4]$. The values of the function $g(x)$ on the borders not equal to zero: $g(0) = 1$ and $g(4) = 25$, the integral over the area of $D = [0; 4]$: $\int_D g(x)dx = -13.06666(6)$, $\int_D x.g(x)dx = -24$.

Table. 2 shows the value of the maximum deviation of spline functions $\max_{x \in D} |S(x) - g(x)|$, integral of spline $\int_D S(x)dx$ and $\tilde{\mu}_1 = \int_D x.S(x)dx$, calculated under several n .

The general solution of differential equations of the fourth order with constant starts coefficients (10)

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

Here z_i^* – a particular solution of (10).

Consider the application of the proposed spline approximation to the empirical probability density, built at the model data. In [5] It invited the family of linear congruent method for generating base random variable.

In general, the linear congruential generator defined by recurrent sequence type $a_n = \frac{a_n^*}{m}$, where

$$a_n^* = (\beta a_{n-1}^* + c) \bmod(m), n = 1, 2, \dots \quad (14)$$

This sequence is cyclic. According to (14), $a_n^* \in \{0, 1, 2, \dots, m\}$, so that period of $T \leq m$. The maximum length of time $T = m$ be achieved, it says the following theorem [5].

Theorem 1.

The maximum possible value of the period sequence (14) and is equal to m if and only if:

- 1) c and m are coprime
- 2) $\beta - 1$ multiple of p for all simple p is a divisor of m ,
- 3) $\beta - 1$ a multiple of four if m is a multiple of four.

To conduct numerical experiment will use the base generator of the random variable with parameters of $c = 51$, $x_0 = 5$, $\beta = 101$, and $c = 7$, $x_0 = 2$, $\beta = 81$, and lengths the period $m = 1000000$. The method polar coordinates construct a sequence of random numbers having standard normal distribution. Membership used for numerical experiment samples to normal distribution was checked using the agreement criterion χ^2 . [6]

Analysis of results numerical experiment showed that spline approximating the empirical distribution shows convergence is the norm of $C[-6; 6]$ – to the the probability density the standard normal distribution (Fig. 1)

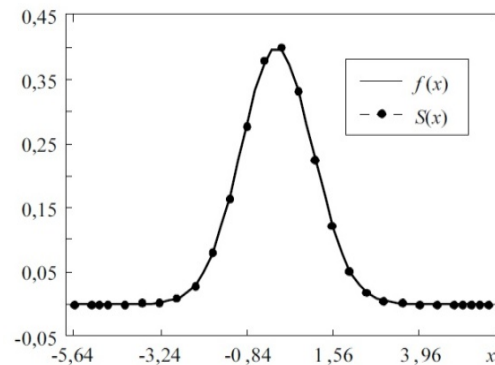


Figure 1. Spline approximation of $S(x)$ the empirical the probability density and timetable density standard normal distribution

Table 2. The results of calculations for the function $g(x)$ for different n

number of partitions	10	20	30	50
Maximum deviation	0,005066	0,001278	0,000589	0,000220
Integral of spline at interval $[0, 4]$	- 13,064400	- 13,066100	- 13,66415	- 13,066576
$\tilde{\mu}_1$	- 23,991763	- 23,997935	- 23,999082	- 23,999669

Table. 3 shows the values $\max_{x \in [-6;6]} |S(x) - f(x)|$, calculated for different n .

Table 3. Shows the values $\max_{x \in [-6;6]} |S(x) - f(x)|$, calculated for different n

size of the sample	$n = 10$	$n = 30$	$n = 100$
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

Here $f(x)$ - the probability density function standard normal distribution. By construction of, $S(x)$ for every partition has a density distribution. By asking increasing number of partitions k , we get the sequence of functions $S_k(x)$, which deviates from the empirical density on the order of $O(h^2)$, that is with increasing size of the sample and increase the number partitions so that $\bar{h} \rightarrow 0$, spline function is convergent to the density a standard normal distribution with the second order smallness with respect to \bar{h} .

3. Conclusions

Spline functions can be used instead of other empirical functions such as polynomials or exponentials for curve fitting of any kind of "continuous" data. Their extreme flexibility and pronounced local properties make them little

inclined to introduce bias in the results due to "filtering" effects.* This makes the number of possible applications enormous; the rather simple examples shown in this paper are only a first limited indication of these possibilities [1].

It is interesting to note the advantages with the use of empirical spline function descriptions of data even when these are thought to conform well to a theoretical "high information" mathematical model. This description corresponds to a direct investigation of the behavior of the data instead of looking on the data through a sophisticated model. From the applied view, this leads to a higher certainty that the information obtained in the analysis is indeed relevant for the scope of the investigation and does not contain artifacts due to the assumed model. For the theoretician, the spline analysis is an excellent tool for study of the behavior of an accustomed model in relation to the data in question.

REFERENCES

- [1] Svante. W. Spline Functions in Data Analysis Technometrics, Vol. 16, No. 1. (Feb., 1974), pp. 1-11.
- [2] Rice, J. R. (1969). The approximation of functions, Vol 2. Addison – Wesly, Reading Massachusetts.
- [3] Whittaker, E .T. and Robinson G. (1926). The calculus of observations. Blackie and Son Ltd., London.
- [4] Zavyalov U. S, Kvassov B. I., Miroshnichenko V. L Methods of spline functions. M.: science, 1980. 352 p.
- [5] Knuth D, The Art of Computer Programming: In 3 t. M.: Mir, 1977. T. 2: Seminumerical algorithms. 728.
- [6] Larry. Schumaker. Spline Functions: Basic theory; 3rd Edition, September 2007.