ISSN: P-2455-0515
E- 2394-8450

Educreator Research Journal

A Peer Reviewed Referred Journal

OPEN ACCESS

Volume–IX, Special Issues–I
March – April 2022

Original Research Article

EXTRAPOLATING DATA BY USING POLYNOMIALS: A CASE STUDY ON PANVEL TAHSIL POPULATION

Sagar Lahanu Khairnar

**Assistant Professor, Department of Mathematics, Changu Kana Thakur Arts, Commerce and Science College, New Panvel, Raigad, Maharashtra.*

Introduction:

In the numerical analysis, spline interpolation is a form of interpolation where the interpolant is a special type of piecewise polynomial called a spline. That is, instead of fitting a single, high-degree polynomial to all of the values at once, spline interpolation fits low-degree polynomials to small subsets of the values. Spline interpolation is often preferred over polynomial interpolation because the interpolation error can be made small even when using low-degree polynomials for the spline. Spline interpolation also avoids the problem of Runge's phenomenon, in which oscillation can occur between points when interpolating using high-degree polynomials.

Literature Review:

A comprehensive review is made on the application of cubic spline interpolation techniques in the field of power systems. Domains like available transfer capability (ATC), electric arc furnace modeling, static var. compensation, voltage stability margin, and market power determination in deregulated electricity market are taken as samples to illustrate the significance of cubic spline interpolation.(Prasad et al., 2018)

Edmund Whittaker, a professor of Numerical Mathematics at the University of Edinburgh from 1913 to 1923, observed “the most common form of interpolation occurs when we seek data from a table which does not have the exact values we want.” Some of the first surviving evidence of the use of interpolation came from ancient Babylon and Greece. Around 300 BC, they were using not only linear, but also more complex forms of interpolation to predict the positions of the sun, moon, and the planets they knew of. Farmers, timing the planting of their crops, were the primary users of these predictions. Also in Greece sometime around 150 BC, Hipparchus of Rhodes used linear interpolation to construct a “chord function”, which is similar to a sinusoidal function, to compute positions of celestial bodies. Farther east, Chinese evidence of interpolation dates back to around 600 AD. Liu Zhuo used the equivalent of second order Gregory-Newton interpolation to construct an “Imperial Standard Calendar”. In 625 AD, Indian astronomer and mathematician Brahmagupta introduced a method for second order interpolation of the sine function and, later on, a method for interpolation of unequal-interval data. (Lavanya & Achireddy, 2016)

(Bahadori, 2011) used spline interpolation to extract natural gas from underground reservoirs is saturated with water. (Bahadori & Nouri, 2012) developed a simple model to estimate the critical oil rate for bottom water coning in anisotropic and homogeneous formations with the well completed from the top of the formation. In a variety of contexts, physicists study complex, nonlinear models with many unknown or tuneable parameters to explain experimental data.

Ghazali et. al. presented an application of ridge polynomial neural network to forecast the future trends of financial time series data (Ghazali et al., 2006). Harju used an FIR polynomial predictor for data in which some samples are missing. The method is compared with a computationally lighter algorithm that is based on decision-driven recursion. Both schemes are found to perform almost identically well on predicting a sinusoidal signal corrupted by both impulsive and Gaussian noise (Harju, 1997). Recently a scheme named CryptoNets is proposed to perform prediction on encrypted data using neural networks. Wu et al. conducted extensive experiments to show the expressiveness of PPoly activations and discussed the tradeoff between accuracy and efficiency for the prediction on encrypted data(Wu et al., 2018). Zjavka developed Accurate short-term wind speed forecasting model for the planning of a renewable energy power generation and utilization, especially in grid systems(Zjavka, 2015). Zaw and Naing forecasted rainfall in Myanmar using MPR and MLR model and found that MPR model gave the more accurate results (Zaw & Naing, 2009). Ostertagová used the polynomial regression model to find the relationship between the strains and drilling The data were analyzed using computer program(Ostertagová, 2012). Hussain proposed a novel type of higher-order pipelined neural network: the polynomial pipelined neural network (Hussain et al., 2008).

Research Methodology:

A polynomial with degree n means a linear combination of the term $x^i, i = 0,1,2, \dots, n$. For every set of data point there is a unique interpolation polynomial with degree= no. of points -1. From the complicated mathematical model or a function f(x) we can get the data points. If we get the interpolation polynomial the we can easily replaced the original model that is the function by the interpolation polynomial for the purpose of analysis and design.

We can write the straight line by using two points i.e. in the form of slope and y-intercept. The slope= α_1, and y-intercept= α_0, the polynomial of degree 1 is $p(x) = \alpha_1 x + \alpha_0$. For the two points $(x_1, f(x_1))$ and $(x_2, f(x_2))$ we get the linear system

$f(x_1) = \alpha_1 x_1 + \alpha_0, f(x_2) = \alpha_1 x_2 + \alpha_0$ or, in matrix form, $\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_0 \end{bmatrix} = \begin{bmatrix} f(x_1) \\ f(x_2) \end{bmatrix}$

Similarly, for three points are compulsory to find the coefficients of a parabola, $p(x) = \alpha_2 x^2 + \alpha_1 x + \alpha_0$.

For the three coefficients of the parabola that passes through the different (non-collinear) points $(x_1, f(x_1)), (x_2, f(x_2))$ and $(x_3, f(x_3))$ are unique and are the solution of the system of three linear equations: $f(x_1) = \alpha_2 x_1^2 + \alpha_1 x_1 + \alpha_0$ $f(x_2) = \alpha_2 x_2^2 + \alpha_1 x_2 + \alpha_0$ $f(x_3) = \alpha_2 x_3^2 + \alpha_1 x_3 + \alpha_0$ or, in matrix form,

$$\begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ x_3^2 & x_3 & 1 \end{bmatrix} \begin{bmatrix} \alpha_2 \\ \alpha_1 \\ \alpha_0 \end{bmatrix} = \begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \end{bmatrix}$$

The above formulation can be extended to interpolating the n points $\{(x_1, f(x_1)), (x_2, f(x_2)), \dots, (x_n, f(x_n))\}$ Using the an $(n-1)$-order polynomial

$$p(x) = \alpha_{n-1} x^{n-1} + \cdots + \alpha_2 x^2 + \alpha_1 x + \alpha_0$$

which results in the following nxn system of linear equation:

$$\begin{bmatrix} x_1^{n-1} & \cdots & x_1 & 1 \\ x_2^{n-1} & \dots & x_2 & 1 \\ \vdots & \ddots & \vdots & \vdots \\ x_n^{n-1} & \cdots & x_n & 1 \end{bmatrix} \begin{bmatrix} \alpha_{n-1} \\ \alpha_{n-2} \\ \vdots \\ \alpha_1 \\ \alpha_0 \end{bmatrix} = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix}$$

The above Coefficient matrices are known as Vandermonde matrices.

The solution for the above system is very subtle to roundoff errors. For the large data set we required a high-order interpolating polynomial. High order interpolating polynomial gives overfitting. If a set of n data points is given and a single $(n-1)$-degree polynomial is used for interpolation, then the polynomial gives the exact values at the points (passes through the points) and gives approximations (interpolated) values between the points. If the number of points is small, the order of the polynomial is low and, typically that leads to practically accurate interpolation. We can obtained better interpolation by using multiple low order polynomials. In this case every polynomial is valid in one interval between two or several points. Generally every polynomial has the same degree with different coefficients. This is known as a spline interpolation or piecewise interpolation.

The point at which the two adjacent splines meet and the slopes changes abruptly is called knot. At this point derivative is not continuous. By using higher order polynomial we can make the underlying function being interpolated smooth. We can improved smoothness by making second derivative continuous at each knot point.

The following equations are the central part of the spline method. We will try to find the lowest-order polynomial spline s_i , which passing through the two interpolating points $(x_i, f(x_i))$ and $(x_{i+1}, f(x_{i+1}))$, with a slope (first derivative) equal to the slope of spline s_{i-1} at x_i and have a concavity (second derivative) that is equal to that of spline s_{i-1} at x_i. We get the following four constraints,

$$s_i(x_i) = f(x_i)$$
$$s_i(x_{i+1}) = f(x_{i+1})$$
$$\frac{ds_i}{dx}(x_i) = \frac{ds_{i-1}}{dx}(x_i)$$
$$\frac{d^2s_i}{dx^2}(x_i) = \frac{d^2s_{i-1}}{dx^2}(x_i)$$

The polynomial that satisfies these four constraints must have at least four degrees of freedom (coefficients). That would be a 3rd -order polynomial, or a cubic. Therefore, s_i has the form,

$$s_i(x) = \alpha_{i,3}x^3 + \alpha_{i,2}x^2 + \alpha_{i,1}x + \alpha_{i,0} \quad \forall i = 0,1,2, \ldots n-1$$

(To fit n data points, we require $n - 1$ splines.)

To compute all $4(n-1)$ spline coefficients, we need to solve $4(n-1)$ equations. We have computed for the cubic splines with three data points x_1, x_2, x_3. For three points, we need 2 cubic splines:

$s_1(x) = \alpha_{1,3}x^3 + \alpha_{1,2}x^2 + \alpha_{1,1}x + \alpha_{1,0}$ (spline between x_1 and x_2)

$s_2(x) = \alpha_{2,3}x^3 + \alpha_{2,2}x^2 + \alpha_{2,1}x + \alpha_{2,0}$ (spline between x_2 and x_3)

$$\frac{ds_1}{dx}(x) = 3\alpha_{1,3}x^2 + 2\alpha_{1,2}x + \alpha_{1,1}$$
$$\frac{d^2s_1}{dx^2}(x) = 6\alpha_{1,3}x + 2\alpha_{1,2}$$
$$\frac{ds_2}{dx}(x) = 3\alpha_{2,3}x^2 + 2\alpha_{2,2}x + \alpha_{2,1}$$
$$\frac{d^2s_1}{dx^2}(x) = 6\alpha_{2,3}x + 2\alpha_{2,2}$$

To satisfy the four constraints, we need to solve eight equations. Each spline pass through two end data points, therefore we get the following four equations

$$s_1(x_1) = \alpha_{1,3}x_1^3 + \alpha_{1,2}x_1^2 + \alpha_{1,1}x_1 + \alpha_{1,0} = f(x_1) \quad (1)$$
$$s_1(x_2) = \alpha_{1,3}x_2^3 + \alpha_{1,2}x_2^2 + \alpha_{1,1}x_2 + \alpha_{1,0} = f(x_2) \quad (2)$$
$$s_2(x_2) = \alpha_{2,3}x_2^3 + \alpha_{2,2}x_2^2 + \alpha_{2,1}x_2 + \alpha_{2,0} = f(x_2) \quad (3)$$
$$s_2(x_3) = \alpha_{2,3}x_3^3 + \alpha_{2,2}x_3^2 + \alpha_{2,1}x_3 + \alpha_{2,0} = f(x_3) \quad (4)$$

The 1st derivatives(as well as 2nd derivative) of splines $s_1(x)$ and $s_2(x)$, at the interior knot (x_2) must be equal, therefore we get two additional equations:

$$3\alpha_{1,3}x^2 + 2\alpha_{1,2}x + \alpha_{1,1} = 3\alpha_{2,3}x^2 + 2\alpha_{2,2}x + \alpha_{2,1} \quad (5)$$
$$6\alpha_{1,3}x + 2\alpha_{1,2} = 6\alpha_{2,3}x + 2\alpha_{2,2} \quad (6)$$

We got a total six equations. We need two additional equations to solve the system. We make the 1st derivative at point x_1 and x_n equal to zero. So we get two additional equations.

$$3\alpha_{1,3}x^2 + 2\alpha_{1,2}x + \alpha_{1,1} = 0 \tag{7}$$

$$3\alpha_{2,3}x^2 + 2\alpha_{2,2}x + \alpha_{2,1} = 0 \tag{8}$$

The above eight equations can be written as in the form of matrix.

$$\begin{bmatrix} x_1^3 & x_1^2 & x_1 & 1 & 0 & 0 & 0 & 0 \\ x_2^3 & x_2^2 & x_2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & x_2^3 & x_2^2 & x_2 & 1 \\ 0 & 0 & 0 & 0 & x_3^3 & x_3^2 & x_3 & 1 \\ 3x_2^2 & 2x_2 & 1 & 0 & -3x_2^2 & -2x_2 & -1 & 0 \\ 3x_2 & 1 & 0 & 0 & -3x_2 & -1 & 0 & 0 \\ 3x_1^2 & 2x_1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3x_3^2 & 2x_3 & 1 & 0 \end{bmatrix} \begin{bmatrix} \alpha_{1,3} \\ \alpha_{1,2} \\ \alpha_{1,1} \\ \alpha_{1,0} \\ \alpha_{2,3} \\ \alpha_{2,2} \\ \alpha_{2,1} \\ \alpha_{2,0} \end{bmatrix} = \begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_2) \\ f(x_3) \\ 0 \\ 0 \\ f'(x_1) \\ f'(x_3) \end{bmatrix}$$

For the spline polynomial interpolation, we have used MATLAB.

Results and Conclusion:

We have considered the population of the Panvel Tahsil from the year 1971 to 2011. Fig. 1 represents the actual data of the year 1971,1981,1991, 2001 and 2011.

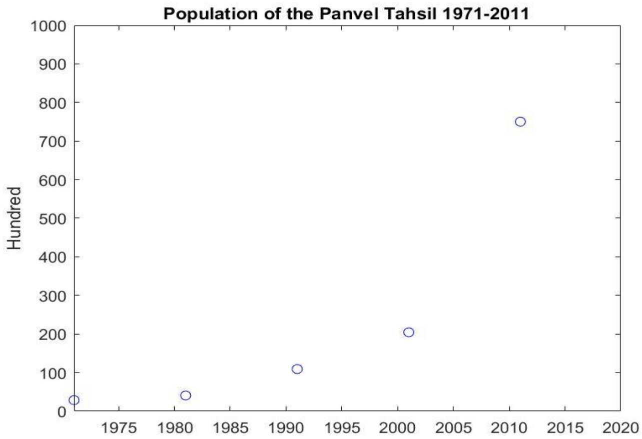

Figure 1

From Figure 2 it is clear that the data for the year 1971, 1981 and 2001 gave the first spline curve and 1971, 1991 and 2011 gave the second spline curve.

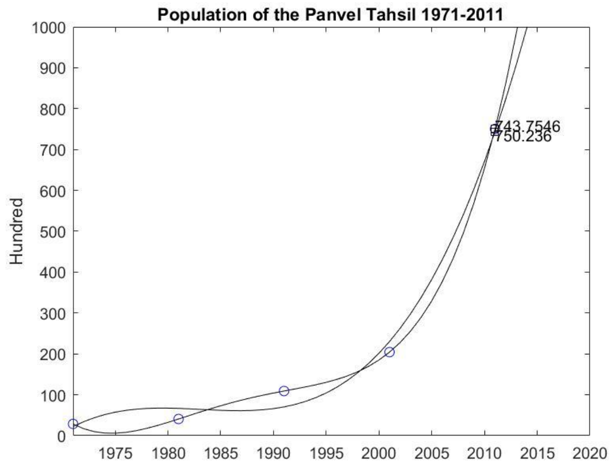

Figure 2

In Figure 3, we can observe that, there are no point on the degree 1 polynomial (linear polynomial) denoted by red colour. Similarly, there is no point on the second-degree polynomial (quadratic polynomial) shown by yellow colour. If we observe the light green colour line i.e. of degree 3 polynomial (cubic polynomial) then there are three points on the line. Most interesting fact that the all the points on the green line. i.e. polynomial of degree 4(quartic polynomial).

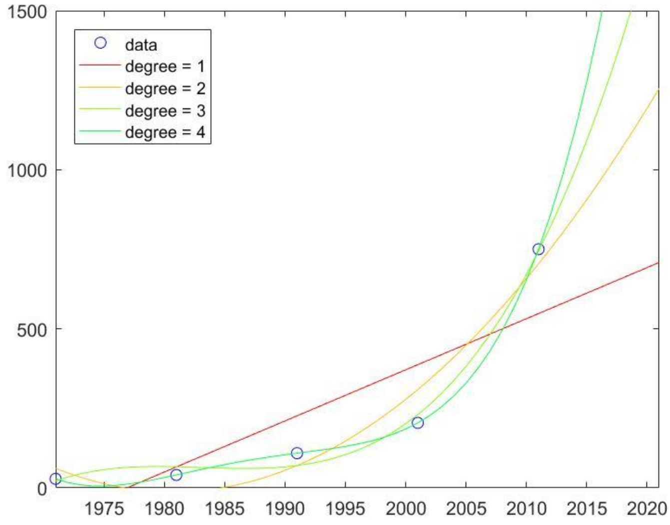

Figure 3

As compared to the degree 1,2,3 and 4 polynomial the degree 4 polynomial i.e. quartic polynomial gave the best approximation for the data. Linear and quadratic polynomial not gave the perfect approximation of the data. It is clear that by using quartic polynomial we can get more reliable predictions as compared to linear, quadratic polynomial. This example shows that extrapolating data using polynomials of even modest degree is dicey and defective.

References:

Bahadori, A. (2011). Prediction of moisture content of natural gases using simple Arrhenius-type function. *Central European Journal of Engineering*, *1*(1), 81–88.

Bahadori, A., & Nouri, A. (2012). Prediction of critical oil rate for bottom water coning in anisotropic and homogeneous formations. *Journal of Petroleum Science and Engineering*, *82*, 125–129.

Ghazali, R., Hussain, A., & El-Deredy, W. (2006). Application of Ridge Polynomial Neural Networks to Financial Time Series Prediction. *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, 913–920.

Harju, P. T. (1997). Polynomial prediction using incomplete data. *IEEE Transactions on Signal Processing*, *45*(3), 768–770.

Hussain, A. J., Knowles, A., Lisboa, P. J. G., & El-Deredy, W. (2008). Financial time series prediction using polynomial pipelined neural networks. *Expert Systems with Applications*, *35*(3), 1186–1199.

Lavanya, Y., & Achireddy, C.(2016). *APPLICATIONS OF INTERPOLATION*. *3*(5), 5.

Ostertagová, E. (2012). Modelling using Polynomial Regression. *Procedia Engineering*, *48*, 500–506.

Prasad, A., Manmohan, A., & D.p, P. K. S. and K. (2018). Application of Cubic Spline Interpolation Technique in Power Systems: A Review. In *Topics in Splines and Applications*. IntechOpen.

Wu, W., Liu, J., Wang, H., Tang, F., & Xian, M. (2018). PPolyNets: Achieving High Prediction Accuracy and Efficiency With Parametric Polynomial Activations. *IEEE Access*, *6*, 72814–72823.

Zaw, W. T., & Naing, T. T. (2009). Modeling of Rainfall Prediction over Myanmar Using Polynomial Regression. *2009 International Conference on Computer Engineering and Technology*, *1*, 316–320.

Zjavka, L. (2015). Wind speed forecast correction models using polynomial neural networks. *Renewable Energy*, *83*, 998–1006.

District Census Handbook Ser 11 Maharashtra Kolaba Eboo. (n.d.).

District Census Handbook Kolaba Ebook and Read Online. (n.d.).

Cite This Article:

***Sagar Lahanu Khairnar,** (2022).* **Extrapolating Data by Using Polynomials: A Case Study on Panvel Tahsil Population**, *Educreator Research Journal IX (Special Issues - I), March –April, 188-193.*