Newton’s Divided Difference Interpolation formula: Representation of Numerical Data by a Polynomial curve

Biswajit Das¹, Dhritikesh Chakrabarty ²

¹Department of Mathematics, Chhaygaon College, Chhaygaon, Assam, India
²Department of Statistics, Handique Girls’ College, Guwahati, Assam, India

Abstract: Due to the necessity of a formula for representing a given set of numerical data on a pair of variables by a suitable polynomial, in interpolation by the approach which consists of the representation of numerical data by a suitable polynomial and then to compute the value of the dependent variable from the polynomial corresponding to any given value of the independent variable, one such formula has been derived from Newton’s divided difference interpolation formula. This paper describes the derivation of the formula with numerical example as its application.

Keywords: Interpolation, divided difference formula, polynomial curve, representation of numerical data.

1. Introduction:

Interpolation, which is a tool for estimating the value of the dependent variable corresponding to a value of the independent variable lying between its two extreme values on the basis of the given values of the independent and the dependent variables [Hummel (1947), Erdos & Turan (1938) et al]. A number of interpolation formulas such as Newton’s Forward Interpolation formula, Newton’s Backward Interpolation formula, Lagrange’s Interpolation formula, Newton’s Divided Difference Interpolation formula, Newton’s Central Difference Interpolation formula, Stirlings formula, Bessel’s formula and some others are available in the literature of numerical analysis [Bathe & Wilson (1976), Jan (1930), Hummel (1947) et al].

In case of the interpolation by the existing formulae, the value of the dependent variable corresponding to each value of the independent variable is to be computed afresh from the used formula putting the value of the independent variable in it. That is if it is wanted to interpolate the values of the dependent variable corresponding to a number of values of the independent variable by a suitable existing interpolation formula, it is required to apply the formula for each value separately and thus the numerical computation of the value of the dependent variable based on the given data are to be performed in each of the cases. In order to get rid of these repeated numerical computations from the given data, one can think of an approach which consists of the representation of the given numerical data by a suitable polynomial and then to compute the value of the dependent variable from the polynomial corresponding to any given value of the independent variable. However, a formula is necessary for representing a given set of numerical data on a pair of variables by a suitable polynomial. Due to this necessity, one such formula has been developed by Das & Chakrabarty (2016). They have been derived this formula from Lagrange’s interpolation formula. In this study, another formula has been derived for the same purpose. The formula has here been derived from Newton’s divided difference interpolation formula. This paper describes the derivation of the formula with numerical example as its application.

2. Newton’s Divided Difference Interpolation Formula:

Newton's Divided Difference is a way of finding an interpolation polynomial (a polynomial that fits a particular set of points or data). Similar to Lagrange's method for finding an interpolation polynomial, it finds the same interpolation polynomial due to the uniqueness of
interpolation polynomials. Newton’s Divided Difference uses the following equation called the divided difference to accomplish this task:

\[ f(x_0, x_1, x_2, \ldots, x_n) - f(x_0, x_1, x) = \frac{x_n - x_0}{x - x_0} \]

and the following equation called Newton’s divided difference formula for the interpolation polynomial is where the polynomial is derived from:

\[ f(x) = f(x_0) + (x - x_0)f'(x_0, x_1) + (x - x_0)(x - x_1)f(x_0, x_1, x_2) + \ldots + (x - x_0)(x - x_1)(x - x_2) \ldots (x - x_{n-1})f(x_0, x_1, x_2, \ldots, x_n) \]

(2.1)

3. Representation of Numerical Data by Polynomial Curve:

By algebraic expansion, one can obtain that

\[ (x - x_0)(x - x_1) = x^2 - (x_0 + x_1)x + x_0x_1 \]

Also,

\[ (x - x_0)(x - x_1)(x - x_2) = x^3 - (x_0 + x_1 + x_2)x^2 + (x_0x_1 + x_0x_2 + x_1x_2)x - x_0x_1x_2 \]

Again,

\[ (x - x_0)(x - x_1)(x - x_2)(x - x_3) = x^4 - (x_0 + x_1 + x_2 + x_3)x^3 + (x_0x_1 + x_0x_2 + x_1x_2 + x_0x_3 + x_1x_3 + x_2x_3)x^2 - (x_0x_1x_2 + x_1x_2x_3 + x_0x_2x_3 + x_1x_3x_4)x + x_0x_1x_2x_3 \]

In general, one can obtain that

\[ (x - x_0)(x - x_1)(x - x_2)(x - x_3)\ldots(x - x_{n-1}) = x^n - \left( \sum_{i=0}^{n-1} x_i \right)x^{n-1} + \left( \sum_{i=0}^{n-2} \sum_{j=1}^{n-1} x_i x_j \right)x^{n-2} - \left( \sum_{i=0}^{n-3} \sum_{j=1}^{n-2} \sum_{k=2}^{n-1} x_i x_j x_k \right)x^{n-3} + \left( \sum_{i=0}^{n-4} \sum_{j=1}^{n-3} \sum_{k=2}^{n-2} \sum_{l=3}^{n-1} x_i x_j x_k x_l \right)x^{n-4} + \ldots \]

+ \ldots + \ldots + \ldots + \ldots + \ldots + \ldots + \ldots

Now, divided difference interpolation formula, described by equation (2.1), can be expressed as

\[ f(x) = C_0 + C_1(x - x_0) + C_2(x^2 - \left( \sum_{i=0}^{n-1} x_i \right)x + x_0x_1) + C_3(x^3 - \left( \sum_{i=0}^{n-2} \sum_{j=1}^{n-1} x_i x_j \right)x - \ldots + \ldots + \ldots + \ldots + \ldots + \ldots + \ldots + \ldots \]

where

\[ C_0 = f(x_0) \]

\[ C_1 = f(x_0, x_1) \]

\[ C_2 = f(x_0, x_1, x_2) \]

\[ C_3 = f(x_0, x_1, x_2, x_3) \]

\[ \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \]

(3.1)
\[ C_n = f(x_0, x_1, x_2, x_3 \ldots x_n) \]

Now, we have

- **Constant term**: \( C_0 = -C_0 x_0 + C_2 x_0 x_1 - C_3 x_0 x_1 x_2 + \ldots \)

\[ C_n (-1)^n (x_0 x_1 x_2 x_3 \ldots \ldots x_{n-1}) \]

- **Coefficient of** \( x \): \( C_1 = C_1 - C_2 (\sum_{i=0}^1 x_i) + C_3 (\sum_{i=0}^1 \sum_{j=1}^2 x_i x_j) \)

\[ -C_4 (\sum_{i=0}^1 \sum_{j=1}^2 \sum_{k=2}^3 x_i x_j x_k) + \ldots \]

\[ (-1)^n C_n x_0 x_1 x_2 x_3 \ldots \ldots x_{n-2} + x_0 x_1 x_2 x_3 \ldots \ldots x_{n-1} \]

- **Coefficient of** \( x^2 \): \( C_2 = C_2 - C_3 (\sum_{i=0}^2 x_i) + \ldots \)

\[ + C_4 (\sum_{i=0}^2 \sum_{j=1}^3 x_i x_j) - \ldots \]

\[ (-1)^n C_n (x_0 x_1 x_2 x_3 \ldots \ldots x_{n-3} + x_0 x_1 x_2 x_3 \ldots \ldots x_{n-2} + x_0 x_1 x_2 x_3 \ldots \ldots x_{n-1}) \]

- **Coefficient of** \( x^3 \): \( C_3 = C_3 - C_4 (\sum_{i=0}^3 x_i) + \ldots \)

\[ + (-1)^n C_n x_0 x_1 x_2 x_3 \ldots \ldots x_{n-4} + x_0 x_1 x_2 x_3 \ldots \ldots x_{n-3} + x_0 x_1 x_2 x_3 \ldots \ldots x_{n-2} + x_0 x_1 x_2 x_3 \ldots \ldots x_{n-1} \]

-coefficient of \( x^n = C_n \)

The equation (3.1), can be expressed as

\[ f(x) = A_0 + A_1 x + A_2 x^2 + A_3 x^3 + \ldots + A_n x^n \]  

(3.2)

which is the required formula for representation of numerical data by a polynomial curve.

- \( A_0 = C_0 - C_1 x_0 + C_2 x_0 x_1 - C_3 x_0 x_1 x_2 + \ldots \)

\[ C_4 x_0 x_1 x_2 x_3 - \ldots \ldots + C_n (-1)^n (x_0 x_1 x_2 x_3 \ldots \ldots x_{n-1}) \]

- \( A_1 = C_1 - C_2 (\sum_{i=0}^1 x_i) + C_3 (\sum_{i=0}^1 \sum_{j=1}^2 x_i x_j) - C_4 \)

\[ (\sum_{i=0}^1 \sum_{j=1}^2 \sum_{k=2}^3 x_i x_j x_k) + \ldots \]

\[ (-1)^n C_n x_0 x_1 x_2 x_3 \ldots \ldots x_{n-2} + x_0 x_1 x_2 x_3 \ldots \ldots x_{n-1} \]

- \( A_2 = C_2 - C_3 (\sum_{i=0}^2 x_i) + C_4 (\sum_{i=0}^2 \sum_{j=1}^3 x_i x_j) - \ldots \)

\[ (-1)^n C_n x_0 x_1 x_2 x_3 \ldots \ldots x_{n-3} + x_0 x_1 x_2 x_3 \ldots \ldots x_{n-2} + x_0 x_1 x_2 x_3 \ldots \ldots x_{n-1} \]

- \( A_3 = C_3 - C_4 (\sum_{i=0}^3 x_i) + \ldots \ldots + (-1)^n C_n \)

\[ (x_0 x_1 x_2 x_3 \ldots \ldots x_{n-4} + x_0 x_1 x_2 x_3 \ldots \ldots x_{n-3} + x_0 x_1 x_2 x_3 \ldots \ldots x_{n-2} + x_0 x_1 x_2 x_3 \ldots \ldots x_{n-1}) \]

\[ \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \]

- \( A_i = C_i - C_{i+1} (\sum_{j=0}^i x_j) + \ldots \ldots + (-1)^{i+1} C_n \)

\[ (x_0 x_1 x_2 x_3 \ldots \ldots x_{n-i+1} + x_0 x_1 x_2 x_3 \ldots \ldots x_{n-1}) \]

\[ \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \]

- \( A_n = C_n \)

Equation (3.2), with the coefficients \( A_0, A_1, A_2, A_3, \ldots, A_n \), as defined above, is the required formula for representing a given set of numerical data on a pair of variables by a suitable polynomial we have aimed at.

**4. Example of Application of the Formula:**

The following table shows the data on total population of Assam corresponding to the years:

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>1971</td>
<td>14625152</td>
</tr>
<tr>
<td>1981</td>
<td>18041248</td>
</tr>
<tr>
<td>1991</td>
<td>22414322</td>
</tr>
</tbody>
</table>

**ISSN:** 2231-5373  [http://www.ijmttjournal.org](http://www.ijmttjournal.org) Page 199
Taking 1971 as origin and changing scale by 1/10, one can obtain the following table for independent variable \( x \) (representing time) and \( f(x) \) (representing total population of Assam):

<table>
<thead>
<tr>
<th>Year</th>
<th>( x_i )</th>
<th>Total Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>1971</td>
<td>0</td>
<td>14625152</td>
</tr>
<tr>
<td>1981</td>
<td>1</td>
<td>18041248</td>
</tr>
<tr>
<td>1991</td>
<td>2</td>
<td>22414322</td>
</tr>
<tr>
<td>2001</td>
<td>3</td>
<td>26638407</td>
</tr>
<tr>
<td>2011</td>
<td>4</td>
<td>31205576</td>
</tr>
</tbody>
</table>

Now here \( x_0 = 0, x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4 \)
\[
f(x_0) = 14625152, \quad f(x_1) = 18041248,
f(x_2) = 22414322, \quad f(x_3) = 26638407,
f(x_4) = 31205576
\]

### Difference Table

<table>
<thead>
<tr>
<th>( x )</th>
<th>( f(x) )</th>
<th>( f(x_0, x_1) )</th>
<th>( f(x_0, x_1, x_2) )</th>
<th>( f(x_0, x_1, x_2, x_3) )</th>
<th>( f(x_0, x_1, x_2, x_3, x_4) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>14625152</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>18041248</td>
<td>3416096</td>
<td>478489</td>
<td>-184327.83</td>
<td>66584.99</td>
</tr>
<tr>
<td>2</td>
<td>22414322</td>
<td>4373074</td>
<td>-74494.5</td>
<td>82012.16</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>26638407</td>
<td>4224085</td>
<td>171542</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>31205576</td>
<td>4567169</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Now, \( C_0 = f(x_0) = 14625152 \)
\[
C_1 = f(x_0, x_1) = 3416096
\]
\[
C_2 = f(x_0, x_1, x_2) = 478489
\]
\[
C_3 = f(x_0, x_1, x_2, x_3) = -184327.83
\]
\[
C_4 = f(x_0, x_1, x_2, x_3, x_4) = 66584.99
\]

The polynomial is
\[
f(x) = A_0 + A_1 x + A_2 x^2 + A_3 x^3 + A_4 x^4
\]
(4.1)

where
\[
A_0 = C_0 = 14625152
\]
\[
A_1 = C_1 - C_2 x_0 - C_3 x_0 x_1 - C_4 x_0 x_1 x_2 + C_4 x_0 x_1 x_2 x_3
\]
\[
= 14625152 - 3416096 \times 0 + 478489 \times 0 \times 1
+ 184327.83 \times 0 \times 1 \times 2 + 66584.99 \times 0 \times 1 \times 2
\]
\[
\times 3
= 14625152
\]
\[
A_2 = C_2 - C_1 x_0 - C_3 x_0 x_1 - C_4 x_0 x_1 x_2 + x_0 x_3 + x_1 x_3 + x_2 x_3
\]
\[
= 478489 + 184327.83 \times (0 + 1 + 2) + 66584.99 \times (0 + 0 + 0 + 6)
\]
\[
= 478489 + 184327.83 \times 3 + 66584.99 \times 11
\]
\[
= 478489 + 552983.49 + 732434.89
\]
\[
= 1763907.38
\]
\[
A_3 = C_3 - C_4 x_0 - x_1 + x_2 + x_3
\]
Thus, the polynomial that can represent the given numerical data is

\[(4.1)\quad f(x) = 14625152 + 2169441.4x + 1763907.38x^2 - 583837.77x^3 + 66584.99x^4\]

This polynomial yields the values of the function \(f(x)\) corresponding to the respective observed values as follows:

\[f(0) = 14625152 + 2169441.4 \times 0 + 1763907.38 \times 0 - 583837.77 \times 0 + 66584.99 \times 0 = 14625152\]
\[f(1) = 14625152 + 2169441.4 \times 1 + 1763907.38 \times 1 - 583837.77 \times 1 + 66584.99 \times 1 = 14625152 + 2169441.4 + 1763907.38 - 583837.77 + 66584.99 = 18625085.77\]
\[f(2) = 14625152 + 2169441.4 \times 2 + 1763907.38 \times 4 - 583837.77 \times 8 + 66584.99 \times 16 = 14625152 + 4338882.8 + 7055629.52 - 4670702.16 + 1065359.84 = 22414322\]
\[f(3) = 14625152 + 2169441.4 \times 3 + 1763907.38 \times 9 - 583837.77 \times 27 + 66584.99 \times 81 = 14625152 + 6508324.2 + 15875166.42 - 15763619.79 + 5393384.19 = 31205576\]
\[f(4) = 14625152 + 2169441.4 \times 4 + 1763907.38 \times 16 - 583837.77 \times 64 + 66584.99 \times 256 = 14625152 + 8677765.6 + 28222518.08\]

Example (ii):

**PROBLEM :** (Total population of India)

The following table shows the data on total population of India corresponding to the years:

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>1971</td>
<td>548159652</td>
</tr>
<tr>
<td>1981</td>
<td>683329097</td>
</tr>
<tr>
<td>1991</td>
<td>846302688</td>
</tr>
<tr>
<td>2001</td>
<td>1027015247</td>
</tr>
<tr>
<td>2011</td>
<td>1210193422</td>
</tr>
</tbody>
</table>

Taking 1971 as origin and changing scale by 1/10, one can obtain the following table for independent variable \(x\) (representing time) and \(f(x)\) (representing total population of India):

<table>
<thead>
<tr>
<th>Year</th>
<th>(x_i)</th>
<th>Total Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>1971</td>
<td>0</td>
<td>548159652</td>
</tr>
<tr>
<td>1981</td>
<td>1</td>
<td>683329097</td>
</tr>
<tr>
<td>1991</td>
<td>2</td>
<td>846302688</td>
</tr>
<tr>
<td>2001</td>
<td>3</td>
<td>1027015247</td>
</tr>
<tr>
<td>2011</td>
<td>4</td>
<td>1210193422</td>
</tr>
</tbody>
</table>

Now here \(x_0 = 0, x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4\)

\[f(x_0) = 548159652, \quad f(x_1) = 683329097, \quad f(x_2) = 846302688, \quad f(x_3) = 1027015247,\]
\[ f(x_4) = 1210193422 \]

### Difference Table

<table>
<thead>
<tr>
<th>( x )</th>
<th>( f(x) )</th>
<th>( f(x_0, x_1) )</th>
<th>( f(x_0, x_1, x_2) )</th>
<th>( f(x_0, x_1, x_2, x_3) )</th>
<th>( f(x_0, x_1, x_2, x_3, x_4) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>548159652</td>
<td>135169445</td>
<td>13902073</td>
<td>-1677529.66</td>
<td>-217007.25</td>
</tr>
<tr>
<td>1</td>
<td>683329097</td>
<td>135169445</td>
<td>13902073</td>
<td>-1677529.66</td>
<td>-217007.25</td>
</tr>
<tr>
<td>2</td>
<td>846302688</td>
<td>135169445</td>
<td>13902073</td>
<td>-1677529.66</td>
<td>-217007.25</td>
</tr>
<tr>
<td>3</td>
<td>1027015247</td>
<td>135169445</td>
<td>13902073</td>
<td>-1677529.66</td>
<td>-217007.25</td>
</tr>
</tbody>
</table>

Now, \( C_0 = f(x_0) = 548159652 \)
\[
C_1 = f(x_0, x_1) = 135169445 \\
C_2 = f(x_0, x_1, x_2) = 13902073 \\
C_3 = f(x_0, x_1, x_2, x_3) = -1677529.66 \\
C_4 = f(x_0, x_1, x_2, x_3, x_4) = -217007.25
\]

The polynomial is
\[
f(x) = A_0 + A_1 x + A_2 x^2 + A_3 x^3 + A_4 x^4 \ldots \quad (4.2)
\]

Where
\[
A_0 = C_0 - C_1 x_0 - C_2 x_0 x_1 - C_3 x_0 x_1 x_2 + C_4 x_0 x_1 x_2 x_3 = 548159652 - 135169445 \times 0 + 13902073 \times 0 \times 1 + 1677529.66 \times 0 \times 1 \times 0 \times 1 \times 2 \times 0.25 \times 1 = 548159652 \\
A_1 = C_1 - C_2 (x_0 + x_1) + C_3 (x_0 x_1 + x_1 x_2) = 135169445 - 13902073 \times (0 + 1) - 1677529.66 \times (0 + 0 + 0 + 6) = 135169445 - 13902073 - 1677529.66 \times 2 + 217007.25 \times 6 = 135169445 - 13902073 - 3355059.32 + 1302043.5 = 136471488.5 - 17257132.32 = 119214356.18
\]

\[
A_2 = C_2 - C_3 (x_0 + x_1 + x_2) + C_4 (x_0 x_1 + x_0 x_2 + x_1 x_2 + x_0 x_3 + x_1 x_3 + x_2 x_3) = 13902073 + 1677529.66 \times (0 + 1 + 2) - 217007.25 \times (0 + 0 + 2 + 0 + 3 + 6) = 13902073 + 1677529.66 \times 3 - 217007.25 \times 11 = 13902073 + 503288.98 - 2387079.75 = 18934661.98 - 2387079.75 = 16547582.23
\]

\[
A_3 = C_3 - C_4 (x_0 + x_1 + x_2 + x_3) = -1677529.66 + 217007.25 \times (0 + 1 + 2) - 217007.25 \times 6 - 217007.25 \times 11 = 1677529.66 + 1302043.5 = -375486.16
\]

\[
A_4 = C_4 = -217007.25
\]

Thus, the polynomial that can represent the given numerical data is
\[
(4.2) \Rightarrow f(x) = 548159652 + 119214356.18 x + 16547582.23 x^2 - 375486.16 x^3 - 217007.25 x^4
\]

which is the required polynomial.

This polynomial yields the values of the function \( f(x) \) corresponding to the respective observed values as follows
\[
\begin{align*}
f(0) & = 548159652 + 119214356.18 \times 0 + 16547582.23 \times 0 - 375486.16 \times 0 - 217007.25 \times 0 \\
& = 548159652 \\
f(1) & = 548159652 + 119214356.18 \times 1 + 16547582.23 \times 1 - 375486.16 \times 1 - 217007.25 \times 1 \\
& = 548159652 + 119214356.18 + 16547582.23
\end{align*}
\]
The approach of interpolation, described here, can be suitably applied in inverse interpolation also.

Newton’s forward interpolation formula is valid for estimating the value of the dependent variable under the following two conditions:

(i) The given values of the independent variable are at equal interval.

(ii) The value of the independent variable corresponding to which the value of the dependent variable is to be estimated lies in the first half of the series of the given values of the independent variable.

However, Newton’s divided difference interpolation formula is valid for estimating the value of the dependent variable beyond these two conditions. Therefore, the formula derived here is valid for representing a set of numerical data on a pair of variables by a polynomial beyond these two conditions.

References:


5. Conclusion:

The formula described by equation (3.2) can be used to represent a given set of numerical data on a pair of variables, by a polynomial.

The degree of the polynomial is one less than the number of pairs of observations.

The polynomial that represents the given set of numerical data can be used for interpolation at any position of the independent variable lying within its two extreme values.