# Biostatistics Department Technical Report

# BST2023-001

Calculation of Orthogonal Polynomials and Their First Two Derivatives with Examples

**Charles R. Katholi, Phd**

**September 2023**

**Department of Biostatistics**
**School of Public Health**
**University of Alabama at Birmingham**
**ckatholi@uab.edu**

# Calculation of Orthogonal Polynomials and Their Derivatives, Revision 2

Charles R. Katholi, PhD

September 20, 2023

## 1 Introduction

The purpose of this report is to explore the calculation of orthogonal polynomials and their derivatives. The basic method follows the approach given by Emerson (1968)[2]. Given a set of points $x_1, \cdots, x_n$, polynomials $p_j(x_i), j = 0, \cdots, m$ where $p_j(x_i)$ is of degree $j$, are found such that the matrix of values

$$P = \begin{bmatrix} p_0(x_1) & p_1(x_1) & p2(x_1) & \cdots & p_m(x_1) \\ p_0(x_2) & p_1(x_2) & p2(x_2) & \cdots & p_m(x_2) \\ p_0(x_3) & p_1(x_3) & p2(x_3) & \cdots & p_m(x_3) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_0(x_n) & p_1(x_n) & p2(x_n) & \cdots & p_m(x_n) \end{bmatrix} \tag{1.1}$$

is orthonormal; that is such that $P^T P = I_{m+1}$ where $I_{m+1}$ is the $(m+1) \times (m+1)$ identity matrix. From this matrix of values recursion coefficients $A_j, B_j$ and $C_j$, $j = 1, \cdots, m$ are found and used to calculate values of the polynomials at any point $x$. It will be shown that the derivatives of the polynomials can also be found recursively utilizing constants $A_j, B_j$ and $C_j$ defined in the next section.

# 2 Derivations of the methods

## 2.1 Calculation of the matrix $P$ of equation (1.1)

We shall utilize the notation of Emerson's paper [2]. Thus, let $x_i, i = 1, \cdots, n$ be given values of $x$ and let $w_i$ be a weight associated with each $x_i$. We shall find the $A_j, B_j$ and $C_j$ such that at any $x$ the values of the orthogonal polynomials are given by the simple recursion (Equation (6) in Emerson),

$$p_j(x) = (A_j x + B_j) p_{j-1}(x) - C_j p_{j-2}(x)\,,\; j = 1, 2, 3, \cdots, m < n \tag{2.1}$$

where $p_{-1}(x) = 0, \forall x$ and $p_0(x) = \left(\sqrt{\sum_{i=1}^n w_i}\right)^{-1}, \forall x$. In the derivation of these recursion constants we make use of the following conditions

$$\sum_{i=1}^{n} w_i p_j(x_i) p_k(x_i) = \left\{ \begin{array}{ll} 1 & ,\ \text{if } j = k \\ 0 & ,\ \text{if } j \neq k \end{array} \right. \tag{2.2}$$

The values of $A_j, B_j$ and $C_j$ are then found recursively from the following equations

$$\begin{aligned} A_j &= \left\{ \sum_{i=1}^{n} w_i x_i^2 p_{j-1}^2(x_i) - \left[ \sum_{i=1}^{n} w_i x_i p_{j-1}^2(x_i) \right]^2 - \left[ \sum_{i=1}^{n} w_i x_i p_{j-1}(x_i) p_{j-2}(x_i) \right]^2 \right\}^{-1/2} \\ B_j &= -A_j \sum_{i=1}^{n} w_i x_i p_{j-1}^2(x_i) \\ C_j &= A_j \sum_{i=1}^{n} w_i x_i p_{j-1}(x_i) p_{j-2}(x_i) \end{aligned} \tag{2.3}$$

The steps in the calculation of the $A_j, B_j$ and $C_j$ given $w_i$ and $x_i, i = 1, \cdots, n$ can be summarized in the following steps

1. For the weights, $w_i$ calculate $s_w = \left[\sum_{i=1}^n w(i)\right]^{-1/2}$. For $i = 1, \cdots, n$ set $p_{-1}(x_i) = 0$ and $p_0(x_i) = s_w$.

2. For $j = 1, 2, \cdots, m$

**(i)** Calculate

$$s_1 = \sum_{i=1}^{n} w_i x_i^2 p_{j-1}^2(x_i)$$
$$s_2 = \sum_{i=1}^{n} w_i x_i p_{j-1}(x_i)$$
$$s_3 = \sum_{i=1}^{n} w_i x_i p_{j-1}(x_i) p_{j-2}(x_i)$$

**(ii)** Calculate

$$A_j = \left\{ s_1 - s_2^2 - s_3^2 \right\}^{-1/2}$$
$$B_j = -A_j s_2$$
$$C_j = A_j s_3$$

(iii) For $i = 1, 2, \cdots, n$ calculate

$$p_j(x_i) = (A_j x_i + B_j) p_{j-1}(x_i) - C_j p_{j-2}(x_i)$$

3. End of loop started at step 2.

## 2.2 Approximating a function using the orthogonal polynomials

At this point the $n \times (m+1)$ orthogonal matrix $P$ has been calculated. Next given observations of a function $y = f(x)$ for $x = x_1, x_2, \cdots, x_n$ we determine coefficients $\alpha_j$, $j = 0, \cdots, m$ such that $f(x)$ is approximated on the range of the $x_i$ by $f(x) \approx \sum_{j=0}^{m} \alpha_j p_j(x)$. This is accomplished by finding the least squares solution of the $n \times (m+1)$ system of linear equations

$$P\boldsymbol{\alpha} = \boldsymbol{y}\,,\ \boldsymbol{y} = (y_1, y_2, \cdots, y_n)^T$$

Because of the orthogonality of $P$ the solution is trivially found to be $\boldsymbol{\alpha} = P^T\boldsymbol{y}$. Noting that $C_1$ is arbitrary and so can be set to zero, the approximation of the function $f(x)$ can be found for any $x$ by setting $p_0(x) = s_w$ and using the recursion of equation(2.1) to calculate $p_1(x), \cdots, p_m(x)$ and then form the linear combination $\widehat{f(x)} = \sum_{j=0}^{m} \alpha_j p_j(x)$.

## 2.3 The first and second derivatives of the model function

If we wish to use the model to estimate the derivative of the approximating polynomial function we require the derivatives $p_j^{'}(x)$ , $j = 0, 1, \cdots, m$. Noting that $p_0(x)$ is a constant independent of $x$ we see that $p_0^{'}(x) = 0$ for all $x$ and that $C_1 = 0$ we can find $p_j^{'}(x)$ by differentiating equation (2.1) to obtain the recursion

$$p_j^{'}(x) = A_j p_{j-1}(x) + (A_j x + B_j) p_{j-1}^{'}(x) - C_j p_{j-2}^{'}(x)\,,\; j = 1, 2, \cdots, m \quad (2.4)$$

Since the recurrence of equation(2.4) requires $p_{j-1}(x)$ it is necessary for any $j$ to first utilize the recursion of equation (2.1) and then the recursion of equation (2.4). The justification for equation (2.4) is given in Appendix A. Differentiation of the recursion a second time leads to a recursion for the second derivative:

$$p_j^{''}(x) = 2A_j p_{j-1}^{'}(x) + (A_j x + B - j) p_{j-1}^{''}(x) - C_j p_{j-2}^{''}(x) \quad (2.5)$$

$$p_{-1}^{''}(x) = 0\,;\; p_0^{''}(x) = 0,\; \forall x \quad (2.6)$$

It should be noted that although these yield the correct second derivatives of the orthogonal polynomials there is no guarantee that the value calculated are good representations of the second derivatives of the function being approximated by the polynomial model. A few examples follow which should help to clarify this point.

# 3 Some examples

We shall consider three examples of approximating functions with orthogonal polynomials of different degrees and the use of these models to estimate the first and second derivative of the function. The first two of these can be classified as “difficult” problems. They are designed to illustrate certain difficulties that can occur when approximating a general function by an expansion in orthogonal polynomials. The third is estimating a cubic polynomial based on “noisy” data. In this case, we expect the approximations to be good and the derivatives to be correct. We shall see, however, if one fits a polynomial of degree larger than the underlying function, then the estimates of the derivatives (especially the second derivative) give incorrect values.

## 3.1 Eample 1:

*Approximating the Runge Function:*

A well known example from elementary Numerical Analysis is the interpolation of the Runge Function

$$f(x) = \frac{1}{1+25x^2}\ ,\ x \in [-1,1] \tag{3.1}$$

Standard divided difference interpolation based on equally spaced points fails completely due to rapid oscillation of the polynomial approximation near $-1$ and 1.[1],[3] The problem is due to the equal spacing of the interpolation data points and becomes worse as the degree of the polynomial increases. The problem can be remedied by making the interpolation table based on the Chebyshev points on the interval $[a,b]$.

$$x_j = \frac{1}{2}\left((a+b)-(a-b)\cos\left[\frac{(2j-1)}{2n}\pi\right]\right) \tag{3.2}$$

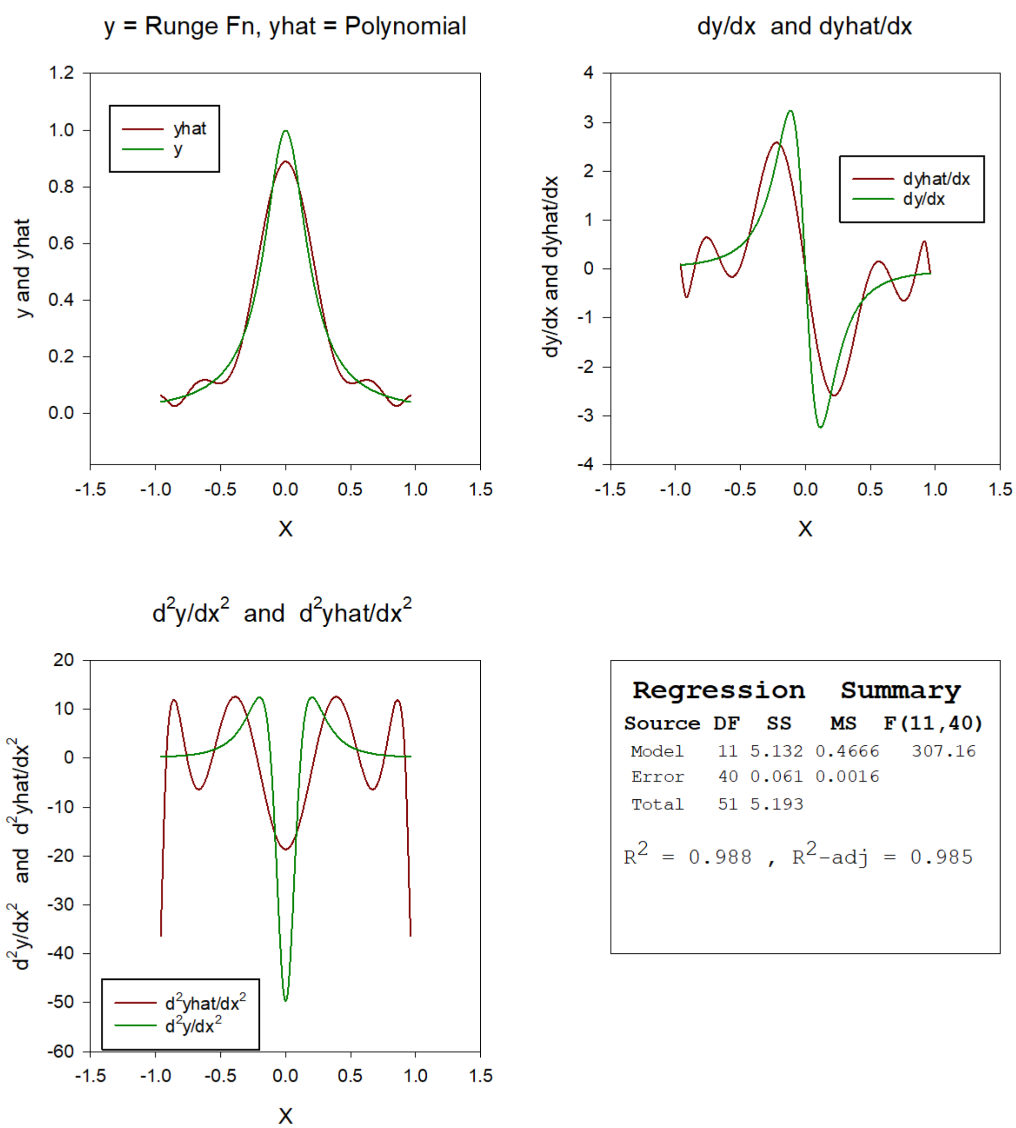
In this case $a=-1$ and $b=1$. Rather than approach this problem by means of approximation by an interpolating polynomial we shall use the methods of least squares for approximation by a set of orthogonal polynomials based on function evaluations at the Chebyshev points. We consider three approximations based on $n=51$ Chebyshev points on $[-1,1]$ and polynomials of degree 10 , 20 and 30 respectively. Noting that the Runge function is symmetric about $x=0$ we expect that only the polynomials of even degree will contribute to the approximation and that proves to be true. Figures 1 and 2 below illustrates the fit based on polynomials up to degree 10, 20 and 30 respectively. The ANOVA table associated with the approximation based on $n=51$ points and polynomials up to degree 10 is

| Source | df | Sum of Squares | Mean Square | F(11,40) |
|---|---|---|---|---|
| Regression | 11 | 5.13254 | 0.46659 | 307.15556 |
| Error | 40 | 0.06076 | 0.00152 | |
| Total | 51 | 5.19331 | | |

| $R^2$ | $R^2-\text{adj}$ |
|---|---|
| 0.98829968 | 0.98537461 |

| $j$ | $\hat{\alpha}_j$ | Sum of Squares | $F(1,40)$ | p-Value |
|---|---|---|---|---|
| 0 | 1.40055 | 1.96154 | 1291.26473 | 0.00010 |
| 1 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |
| 2 | -1.33117 | 1.77201 | 1166.50126 | 0.00010 |
| 3 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |
| 4 | 0.89465 | 0.80040 | 526.89629 | 0.00010 |
| 5 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |
| 6 | -0.60128 | 0.36153 | 237.99349 | 0.00010 |
| 7 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |
| 8 | 0.40410 | 0.16330 | 107.49915 | 0.00010 |
| 9 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |
| 10 | -0.27159 | 0.07376 | 48.55623 | 0.00010 |

The $95^{th}$ percentile point of a F distribution with $\nu_1 = 11$ and $\nu_2 = 40$ degrees of freedom is 2.03 so it is clear that the model is fitting the Runge Function very well. Since the Runge Function is an even function, as expected every coefficient associated with an odd power of $x$ is zero. The p-values in the table denoted by 0.00010 stand for probabilities less than or equal to 0.0001. These statistics paint a far more rosy picture of the quality of the estimate than we see when we observe a plot of the function and its approximation over the interval $[-1, 1]$ as given in Figure 1.

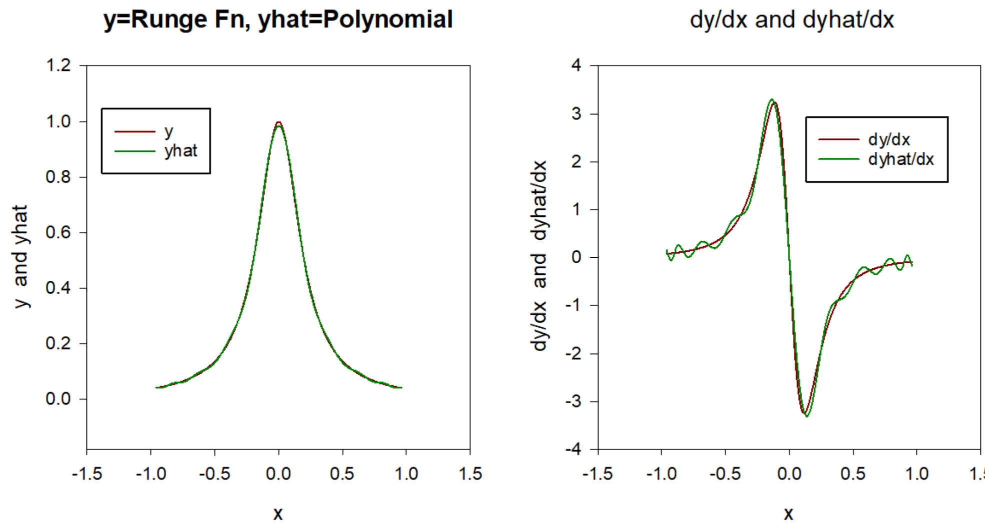


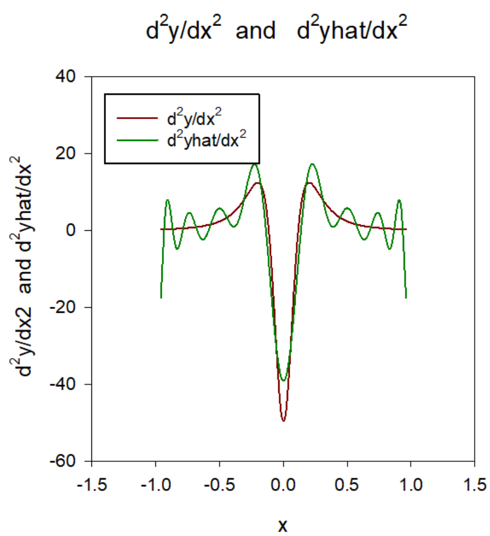

**Regression Summary**

| Source | DF | SS | MS | F(11,40) |
|---|---|---|---|---|
| Model | 11 | 5.132 | 0.4666 | 307.16 |
| Error | 40 | 0.061 | 0.0016 | |
| Total | 51 | 5.193 | | |

$R^2 = 0.988$ , $R^2\text{-adj} = 0.985$

Figure 1: Approximation of the Runge function, $f(x) = (1 + 25x^2)^{-1}$ by orthogonal polynomials of maximum degree 10 by the method of Least Squares.




**Regression Summary**

| Source | DF | SS | MS | F(21.30) |
|---|---|---|---|---|
| Model | 21 | 5.1922 | 0.24725 | 6191.4 |
| Error | 30 | 0.0011 | 0.00004 | |
| Total | 51 | 5.1933 | | |

$R^2$=0.99965, $R^2$-adj=0.99941

Figure 2: Approximation of the Runge function, $f(x) = (1 + 25x^2)^{-1}$ by orthogonal polynomials of maximum degree 20 by the method of Least Squares.

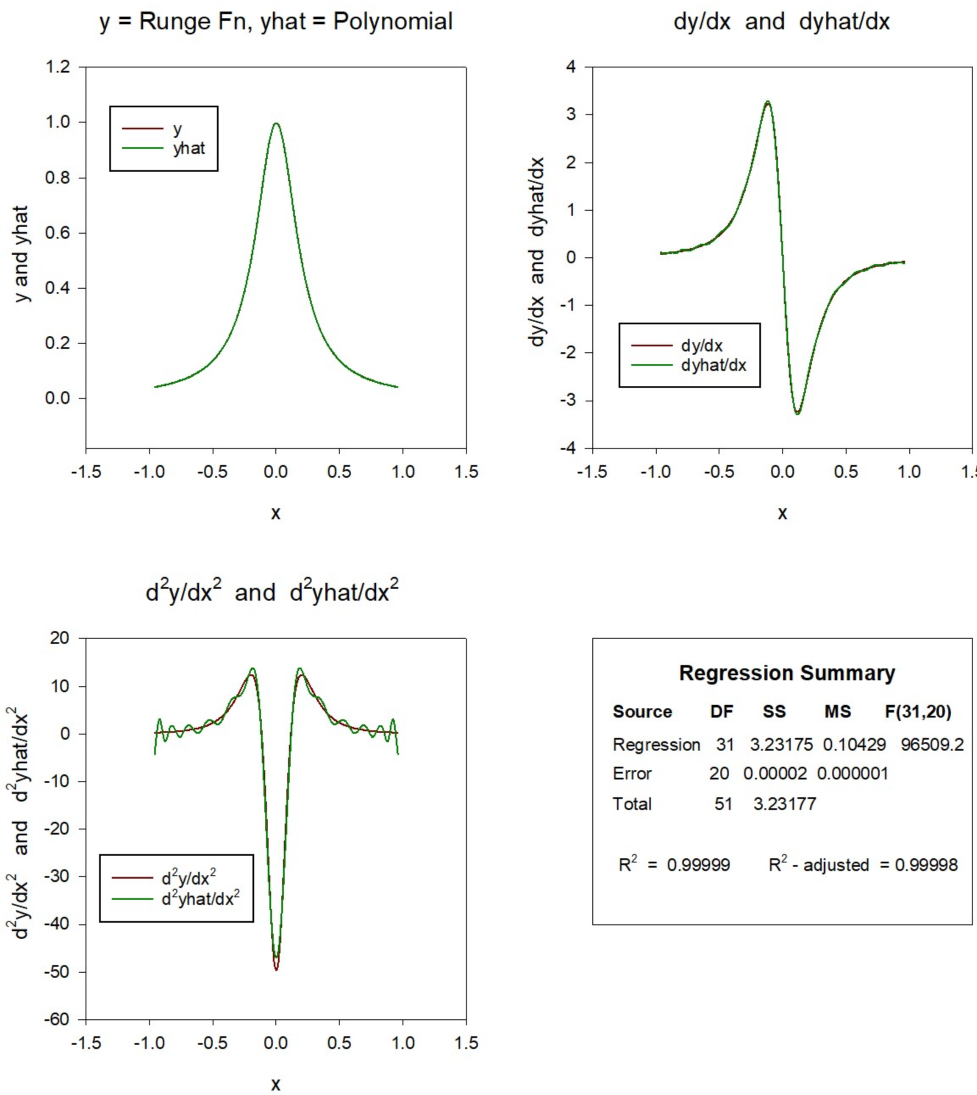


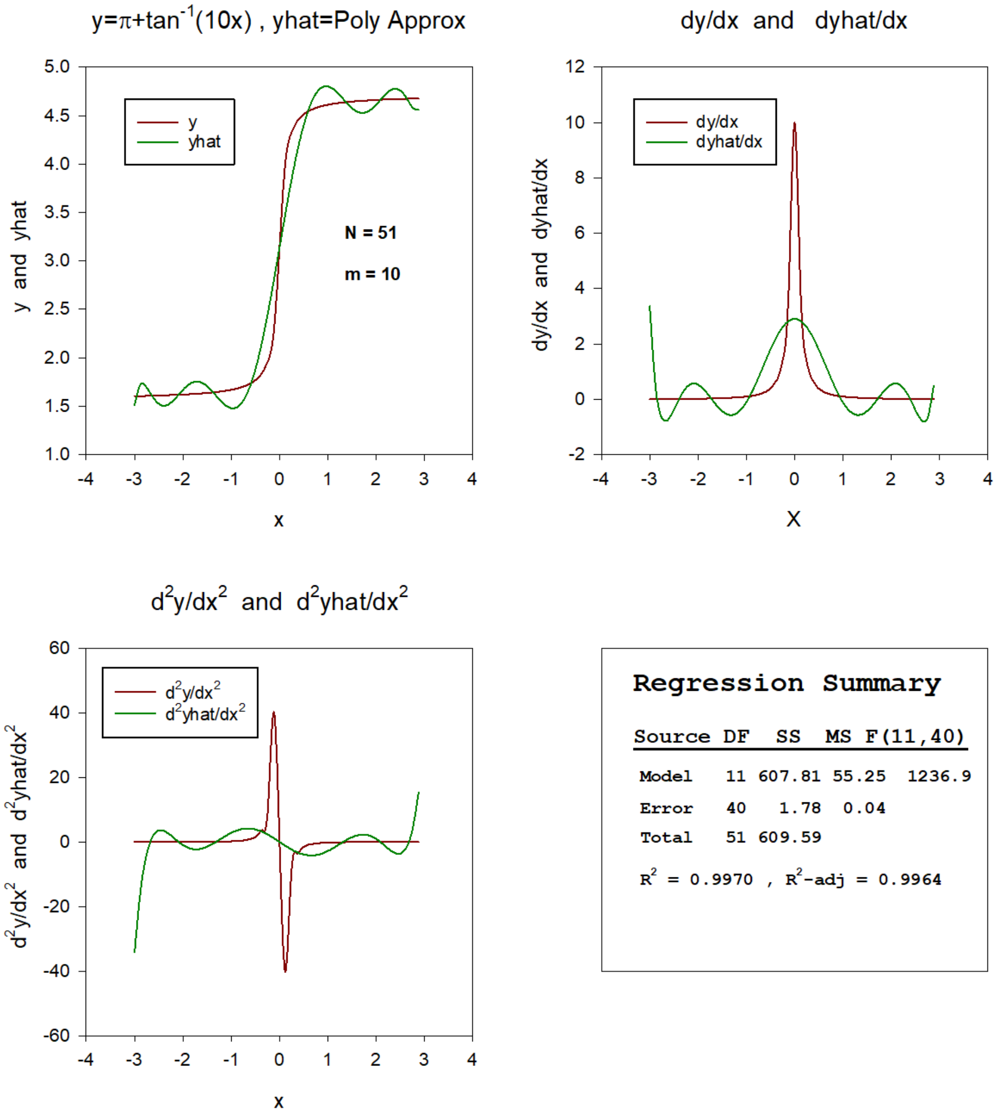

Figure 3: Approximation of the Runge function, $f(x) = (1 + 25x^2)^{-1}$ by orthogonal polynomials of maximum degree 30 by the method of Least Squares.

Increasing the maximum degree of the polynomials to 20 and then to 30 improves the fit to the function. This is illustrated in Figures 2 and 3. We see that the approximation to the function improves steadily as do the approximations to the first and second derivative. However, even when the maximum degree is 30 the derivatives continue to show oscillations in the tail regions. Thus we would judge that approximation of the derivatives, especially the second derivative, by differentiating the polynomial model is a questionable practice.

### 3.2 Example 2:

*Approximation of the function* $f(x) = \pi + tan^{-1}(10x)$ , $-3 \leq x \leq 3$

This function has been chosen to approximate a step function at $x = 0$. It is well known that approximation of a step function by a Fourier series leads to oscillations in the regions away from the step function. These oscillations are called Gibbs Phenomenon. It is apparent from this example that similar behavior occurs for this function when approximating it with orthogonal polynomials on a evenly spaced set of points for $x \in [-3, 3]$. Figures 4 through 6 illustrate respectively the approximation of the function by polynomials of degrees $10, 20$ and $30$ respectively. Even though the $R^2$ values in each case would suggest a reasonable fit to the function, we see that the oscillations persist even when the degree of the polynomial is $m = 30$. More importantly, we note that the first derivative of the function resembles a Dirac delta function with a single narrow spike at $x = 0$ and a nearly flat contour elsewhere. The derivatives, both first and second, of the approximating polynomial are wildly different from the true function and exhibit extreme fluctuations near the ends of the interval $[-3, 3]$. In figure 6 the first derivative is plotted twice; once on the interval $[-3, 3]$ and then again on the subinterval $[-2.64, 2.64]$. This is done because the wild behavior at the ends masks the real behavior near $x = 0$ because of the size of the fluctuations. In short, the graphs tell the whole story. If the function being approximated has regions where there is little change and a narrow area where there is an abrupt change, then the approximation of such a function can adequately model the function but completely miss-represent the derivatives.

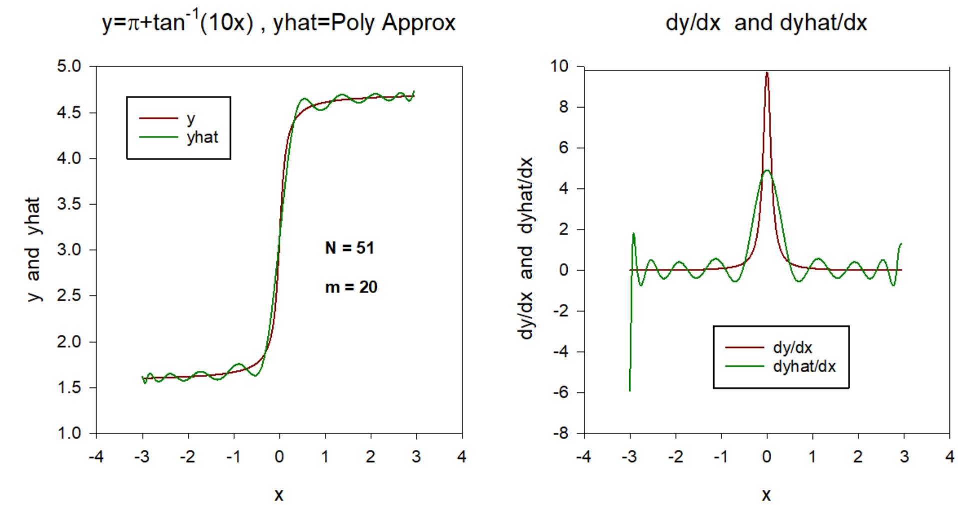


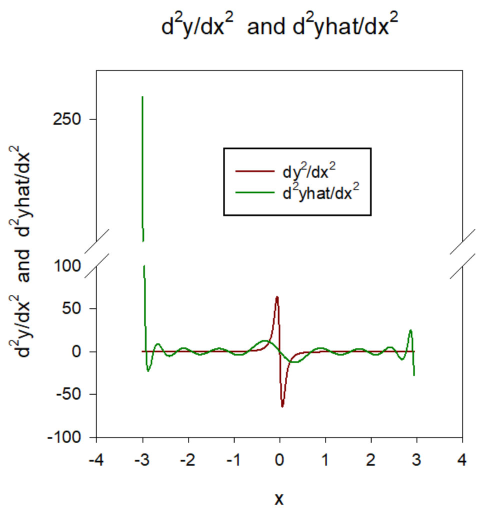

**Regression Summary**

| Source | DF | SS | MS | F(11,40) |
|---|---|---|---|---|
| Model | 11 | 607.81 | 55.25 | 1236.9 |
| Error | 40 | 1.78 | 0.04 | |
| Total | 51 | 609.59 | | |

$R^2 = 0.9970$ , $R^2$-adj $= 0.9964$

Figure 4: Approximation of the function, $f(x) = \pi + tan^{-1}(10x)$ , $x \in [-3, 3]$ by orthogonal polynomials of maximum degree 10 by the method of Least Squares.




**Regression Summary**

| Source | DF | SS | MS | F(21,30) |
|---|---|---|---|---|
| Model | 21 | 609.28 | 29.01 | 2736.7 |
| Error | 30 | 0.32 | 0.01 | |
| Total | 51 | 609.59 | | |

$R^2 = 0.99948$ , $R^2$-adj $= 0.99913$

Figure 5: Approximation of the function, $f(x) = \pi + tan^{-1}(10x)$ , $x \in [-3, 3]$ by orthogonal polynomials of maximum degree 20 by the method of Least Squares.

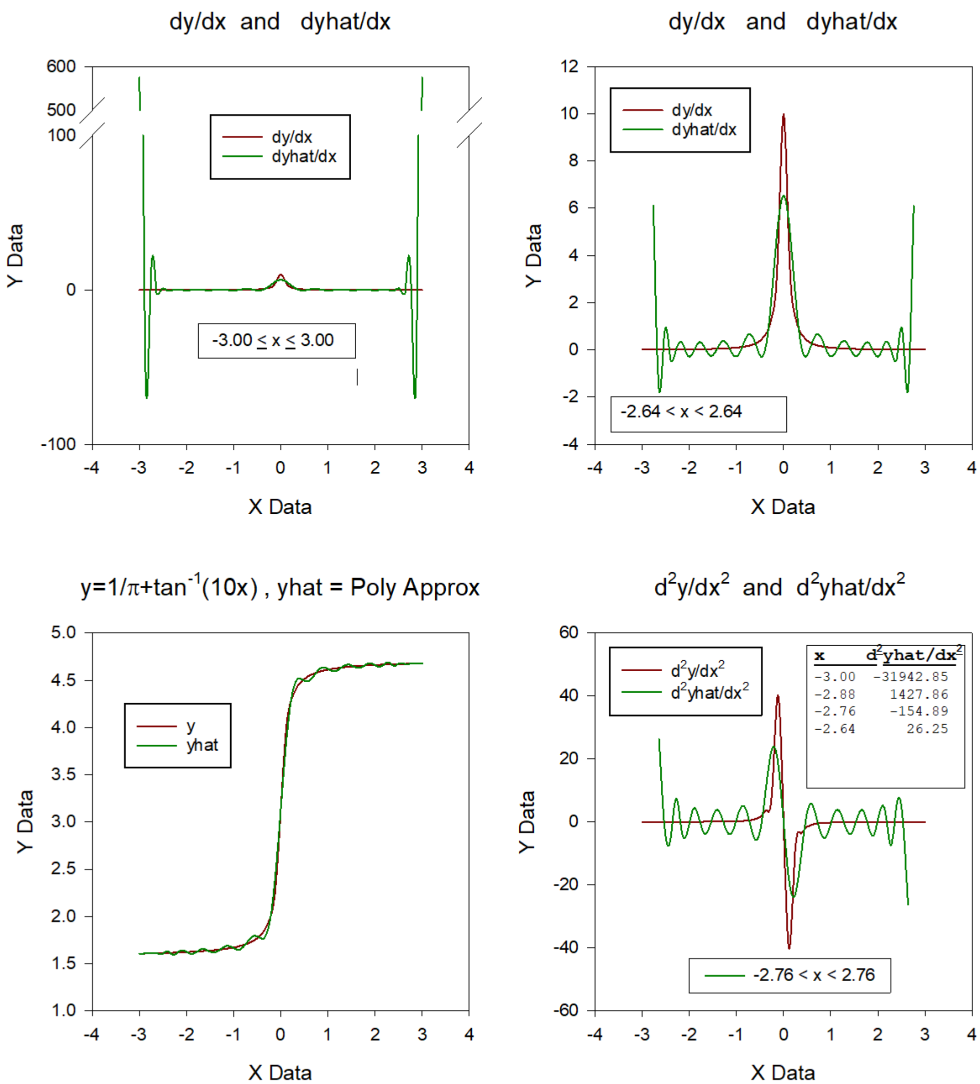



Figure 6: Approximation of the function, $f(x) = \pi + tan^{-1}(10x)$ , $x \in [-3, 3]$ by orthogonal polynomials of maximum degree 30 by the method of Least Squares.Note that although the model captures the rapid rise of the function at $x = 0$ it still exhibits the oscillations at $x^-$ and $x^+$. The oscillations are even more pronounced for the first and second derivative of the modeling polynomial

## 3.3 Example 3:

*Modeling a function subject to errors in the observations*

In this example we consider the common problem of modeling noisy data by a polynomial model. That is, we assume that we have $n$ observations $y(x_i) = y_t(x_i) + \epsilon_i, i = 1, \cdots, n$ where $\epsilon_i \sim N(0, \sigma^2)$ and we estimate $y_t$ by means of a polynomial model $y(x_i) = \sum_{k=0}^{m} \alpha_k p_k(x_i)$ , $i = 1, \cdots, n$ where the $p_k, k = 0, 1, \cdots, m$ are a set of orthonormal polynomials. Thus suppose that $\boldsymbol{y} = (y(x_1), \cdots, y(x_n))^T \in \mathbb{R}^n$ and that the $\boldsymbol{p_k} = (p_k(x_1), \cdots, p_k(x_n))^T \in \mathbb{R}^n$ are such that

$$\boldsymbol{p_j}^T \boldsymbol{p_k} = \begin{cases} 0 & , if j \neq k \\ 1 & , if j = k \end{cases}$$

then the Least Squares estimates of the $\alpha_k$ are given by $\hat{\alpha}_k = \boldsymbol{p_k}^T \boldsymbol{y}$.

For this example we consider a pseudo-random sample, $\boldsymbol{\epsilon} \in \mathbb{R}^n$, from the Normal, $N(0, 6.25)$ and the cubic polynomial model

$$y(x_i) = ((0.035x_i + 1.3)x_i + 13.1)x_i + 60.9 + \epsilon_i \, , \, i = 1, \cdots, n$$

We begin by examining a scatter plot of the raw data to get a feel for the possible degree of the polynomial model. To this end we have figure 7,

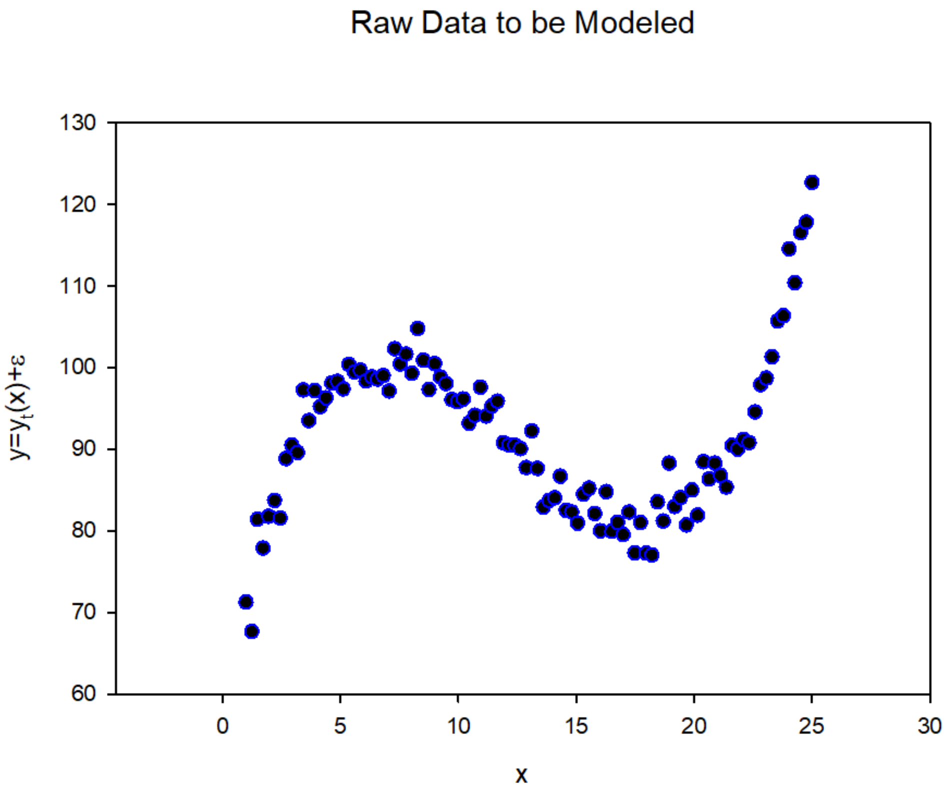



Figure 7: Scatter plot of the observed pseudo-data $y(x_i) = y_t(x_i) + \epsilon_i$ , $i = 1, \cdots, n$.

From this plot it is clear that the polynomial model should be at least of degree m=3 and possibly of a higher degree, say m=5. taking $m = 5$ leads to the following Analysis of Variance table

| Source | df | Sum of Squares | Mean Square | F( 6,94) |
|---|---|---|---|---|
| Regression | 6 | 849254.59514 | 141542.43252 | 32170.94316 |
| Error | 94 | 413.57161 | 4.39970 | |
| Total | 100 | 849668.16674 | | |

| $R^2$ | $R^2 - \text{adj}$ |
|---|---|
| 0.99951326 | 0.99948736 |

| $j$ | $\hat{\alpha}_j$ | Sum of Squares | $F(1,94)$ | p-Value |
|---|---|---|---|---|
| 0 | 916.56773 | 840096.40408 | 190944.10902 | 0.00010 |
| 1 | 13.08182 | 171.13399 | 38.89676 | 0.00010 |
| 2 | 23.75683 | 564.38701 | 128.27858 | 0.00010 |
| 3 | 91.73522 | 8415.34994 | 1912.71084 | 0.00010 |
| 4 | 1.41578 | 2.00443 | 0.45558 | 0.50135 |
| 5 | 2.30558 | 5.31569 | 1.20819 | 0.27450 |

Looking at the ANOVA table for the we see that the $4^{th}$ and $5^{th}$ coefficients corresponding to the polynomial model of degree 5 are not significant $p > 0.275$. In figure 8 we see that the presence of the nonsignificant parameters has an undesirable effect on the first and second derivatives.

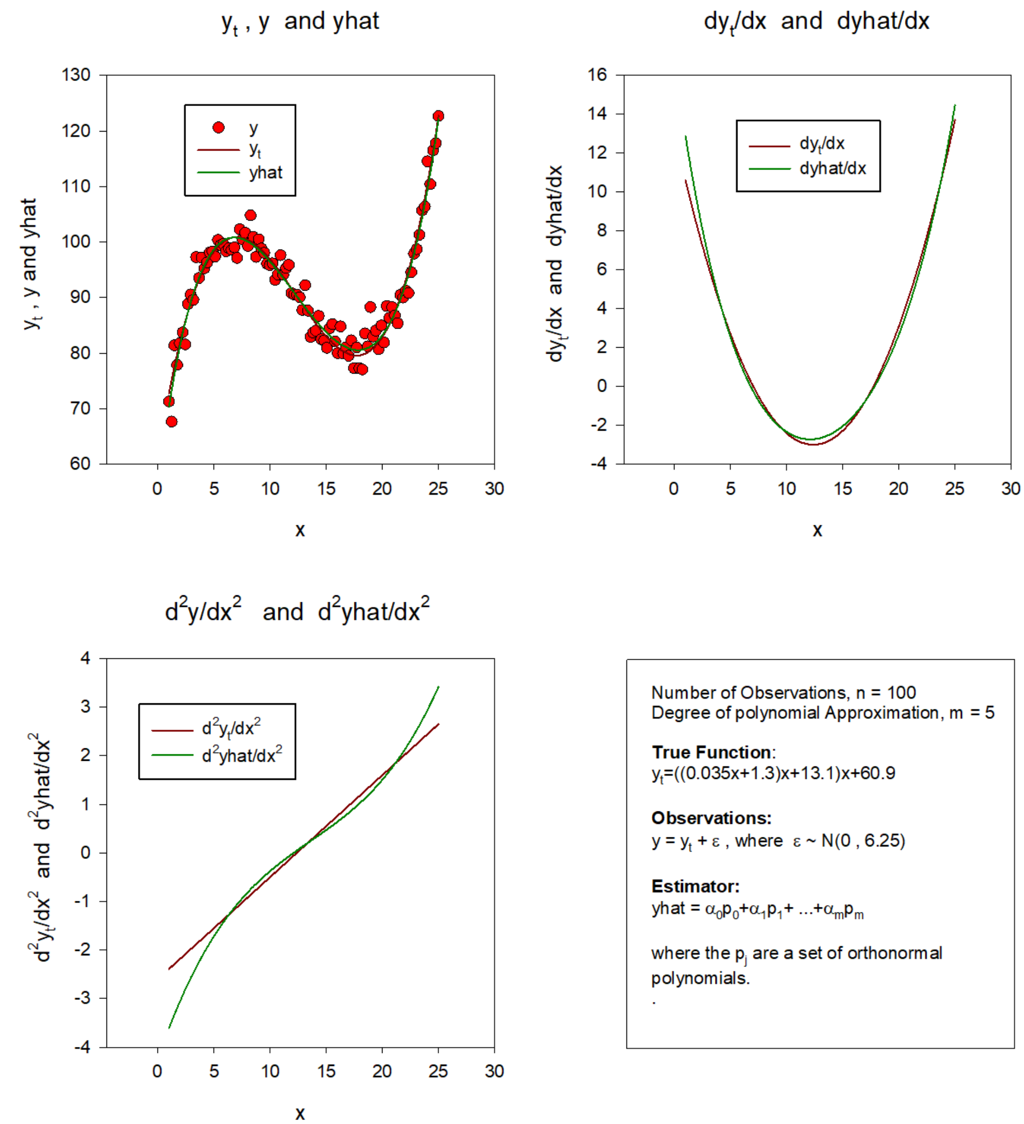



Figure 8: Plot of the true value of the function and it's first and second derivatives and the estimated value with it's derivatives. We see that there is some lack of fit for the first derivative but a more extensive lack of fit for the second derivative.

Thus, in Figure9 we have a plot of the model and derivatives when the nonsignificant parameters are not included in the model.

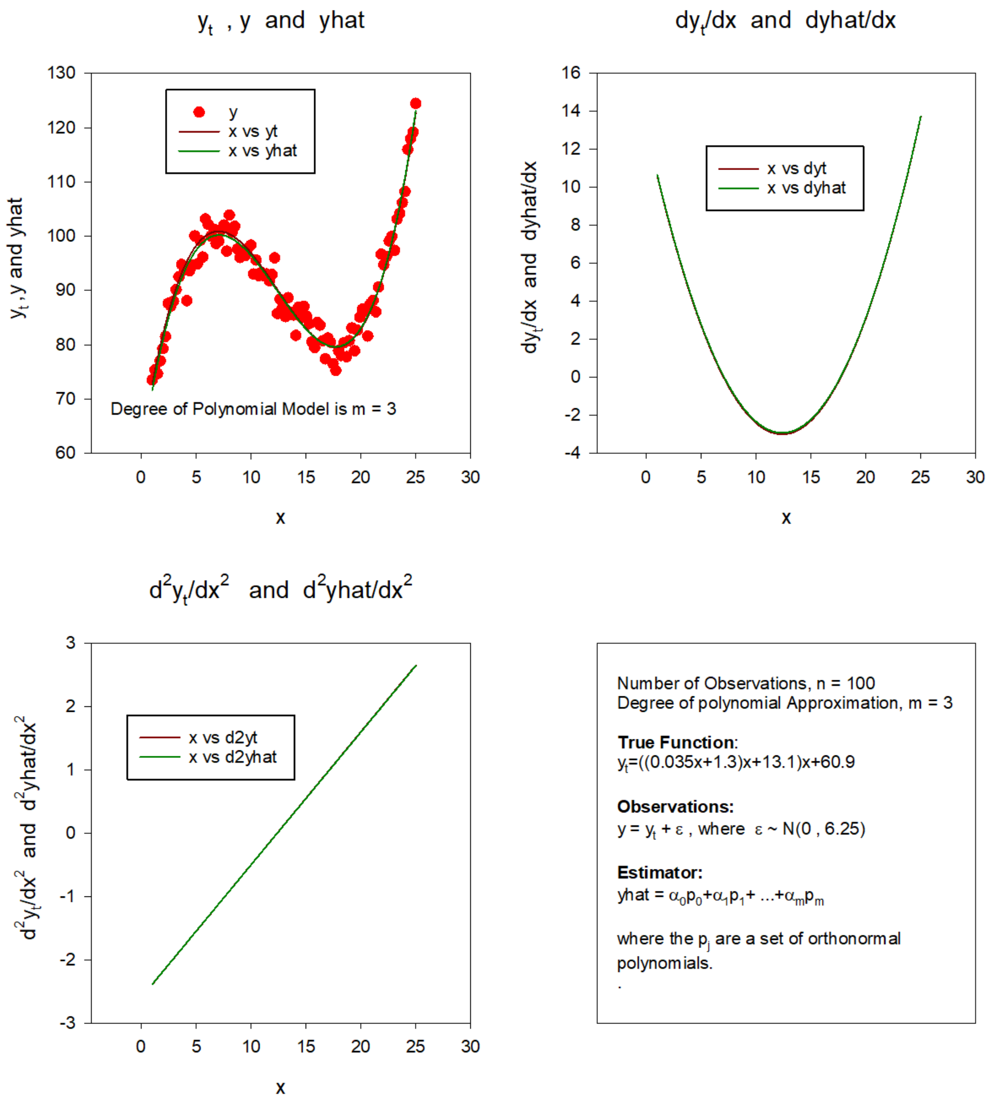



Figure 9: Plot of the true value of the function and it's first and second derivatives and the estimated value with it's derivatives. We see that there is no lack of fit for the first derivative and second derivative.

For completeness we look at the polynomial model when the degree of the polynomial is too small. In this case we choose to take $m = 2$ so that the polynomial approximation is of degree 2. The analysis of variance table in this case is

| Source | df | Sum of Squares | Mean Square | F( 3,97) |
|---|---|---|---|---|
| Regression | 3 | 840831.92508 | 280277.30836 | 3076.74914 |
| Error | 97 | 8836.24166 | 91.09527 | |
| Total | 100 | 849668.16674 | | |

| $R^2$ | $R^2 - \text{adj}$ |
|---|---|
| 0.98960036 | 0.98938594 |

| $j$ | $\hat{\alpha}_j$ | Sum of Squares | $F(1,97)$ | p-Value |
|---|---|---|---|---|
| 0 | 916.56773 | 840096.40408 | 9222.17322 | 0.00010 |
| 1 | 13.08182 | 171.13399 | 1.87863 | 0.17365 |
| 2 | 23.75683 | 564.38701 | 6.19557 | 0.01451 |

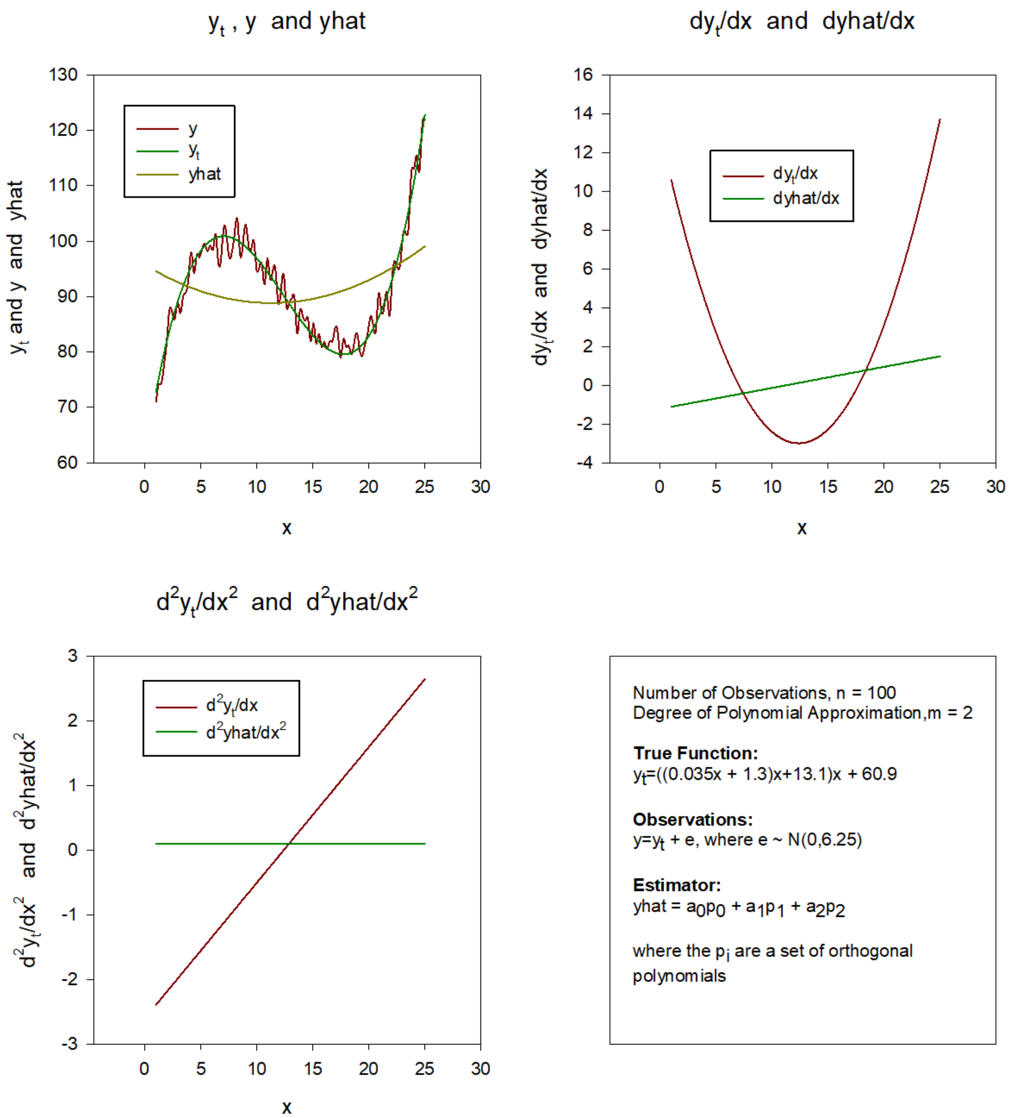



Figure 10: Plot of the true value of the function and it's first and second derivatives and the estimated value with it's derivatives in the case where we have underestimated the degree of the polynomial.

Not surprisingly we see in Figure 10 that if the polynomial model has degree too small to allow capture of the shape of the data, not only are the derivative estimates bad but so is the fit to the data. That is, we observe a simple quadratic curve versus the plot of the raw data and of the true function, denoted as previously by $y_t$. Thus we see that if we choose the degree of the polynomial too small, the resulting polynomial model simply fails to capture the shape of the data.

## 4 Summary

We have shown how to calculate the derivatives of a set of orthogonal polynomials given the constants for the three term recursion which generates the polynomial values. We have seen that these same constants can be used in a recursion to calculate the derivatives of the polynomials and hence of any model function based on a linear combination of the polynomials. A down side of this approach for SAS users is that the procedure ORPOL does not return the recurrence coefficients that are calculated by the methods described in this report [2]. The computations for this report were done using a program written in FORTRAN2003. A subroutine called ORPOLY.F95 is available from the author.

# Appendices

## A Derivation of equation(2.4)

Note that in the construction of the recursion given in equation(2.1), the constants $A_j, B_j$ and $C_j$, $j = 1, 2, \cdots, m$ are found recursively and once they are found, the value of the polynomials at any point $x$ can be found using equation(2.1),

$$p_j(x) = (A_j x + B_j)p_{j-1}(x) - C_j p_{j-2}(x)\ ,\ j = 1, 2, 3, \cdots,\ ,\ m < n$$

In particular, given any $x$ we can use the recursion to find values of the polynomials at the points $x+h$ where $h$ is a small number, the by elementary calculus we know that

$$p_j^{'}(x) = \lim_{h \to 0} \frac{p_j(x+h) - p_j(x)}{h}$$

and from application of the Taylor expansion we can write,

$$p_j(x+h) = p_j(x) + hp_j^{'}(x) + \circ(h) \tag{A.1}$$

From equation(2.1) we have that

$$p_j(x+h) = (A_j(x+h) + B_j)p_{j-1}(x+h) - C_j p_{j-2}(x+h)$$

if we now replace $p_k(x+h)$ for $k = j, j-1, j-2$ with the appropriate expressions from equation(A.1) in equation(2.1) , subtracting equation(2.1) leads (after considerable algebra) to

$$\begin{aligned} \frac{p_j(x+h) - p_j(x)}{h} &= (A_j x + b_j)p_{j-1}^{'}(x) - C_j p_{j-2}^{'}(x) + A_j p_{j-1} \\ &+ hA_j p_{j-1}^{'} + \circ(h) \end{aligned} \tag{A.2}$$

Taking the limit of both sides of this equation leads to the result given in equation(2.4).

# References


[1] de Boor, Carl (2001),*A Proactial Guide to Splines: revised edition*, Springer, New York.

[2] Emerson, Phillip L.,(1968) Numerical Construction of Orthogonal Polynomials From a General Recurrence Formula, *Biometrics*, Vol 24, pp 695-701.

[3] Forsythe, G. ,Malcolm, M. and Moler, C.,*Computer Methods for Mathematical Computations*, Prentice-Hall series in automaticd computation, Prentice Hall, 1977.