This is a preprint of an article published in the *Encyclopedia of Biostatistics*. Please cite as: Smyth, G. K. (1998). Polynomial approximation. In: *Encyclopedia of Biostatistics*, P. Armitage and T. Colton (eds.), Wiley, London, pp. 3425–3429.

Polynomial Approximation

Gordon K. Smyth
Department of Mathematics
University of Queensland

May 1997

Abstract

Polynomials are very simple mathematical functions which have the flexibility to represent very general nonlinear relationships. Approximation of more complicated functions by polynomials is a basic building block for many numerical techniques. This article considers two distinct but related applications. The first is polynomial regression in which polynomials are used to model a nonlinear relationship between a response variable and an explanatory variable. The advantages of using orthogonal polynomials as predictor variables are illustrated using a data set on the height and age of pre-adult girls. The second problem is that of approximating a difficult to evaluate function, such as a density or a distribution function, with the aim of fast evaluation on a computer. The use of Chebychev polynomials is illustrated for the purpose of obtaining a uniformly accurate approximation to a function over a finite interval.

Keywords: polynomial regression, orthogonal polynomials, Legendre polynomials, Chebyshev polynomials, Laguerre polynomials, Hermite polynomials, Chebyshev interpolation.

1 Introduction

A polynomial is a function which can be written in the form

$$p(x) = c_0 + c_1 x + \cdots + c_n x^n$$

for some coefficients $c_0, \ldots, c_n$. If $c_n \neq 0$, then the polynomial is said to be of order n. A first order (linear) polynomial is just the equation of a straight line, while a second order (quadratic) polynomial describes a parabola.

Polynomials are just about the simplest mathematical functions that exist, requiring only multiplications and additions for their evaluation. Yet they also have the flexibility to represent very general nonlinear relationships. Approximation of more complicated functions by polynomials is a basic building block for a great many numerical techniques.

There are two distinct purposes to which polynomial approximation is put in statistics. The first is to model a nonlinear relationship between a response variable and an explanatory variable. The response is usually measured with error, and the interest is on the shape of the fitted curved and perhaps also on the fitted polynomial coefficients. The demands of parsimony and interpretability ensure that one will seldom be interested in polynomial curves of more than 3rd or 4th order in this context.

The second purpose is to approximate a difficult to evaluate function, such as a density or a distribution function, with the aim of fast evaluation on a computer. Here there is no interest in the polynomial curve itself. Rather the interest is on how closely the polynomial can follow the special function, and especially on how small the maximum error can be made. Very high order polynomials may be used here if they provide accurate approximations. Very often a function is not approximated directly, but is first transformed or standardized so to make it more amenable to polynomial approximation.

On either type of problem, substantial benefit can be had from orthogonal polynomials. Orthogonal polynomials can be used to make the polynomial coefficients uncorrelated, to minimize the error of approximation, and to minimize the sensitivity of calculations to round-off error.

Suppose that the function to be approximated, $f(x)$, is observed at a series of values $x_1, \ldots, x_N$. In general we will observe $y_i = f(x_i) + \epsilon_i$ where the ϵ_i are unknown errors. The task is to estimate $f(x)$ for new values of x. If the new x is within the range of the observed absissae then the problem is *interpolation.* If it is outside, then the problem is *extrapolation.* Polynomials are useful for interpolation, but notoriously poor at extrapolation.

Polynomial approximation is relatively straightforward and good enough for many purposes. There are, however, many other ways to approximate functions. Many functions, for example, can be more economically approximated by rational functions, which are quotients of polynomials. A survey of approximation methods is given by Press et al [4, Chapter 4].

Most numerical analysis texts include a treatment of polynomial approximation. Atkinson [2, Chapter 4] gives a nice treatment of minimax approximation using Chebyshev polynomials. Many specific polynomial approximation formulae to functions used by statisticians are given by Abramowitz and Stegun [1]. Many statistical texts mention polynomial regression. Kleinbaum [3, Chapter 13] gives a very accessible treatment, while that of Seber [5, Chapter 8] is more detailed and mathematical.

2 Taylor's Theorem

Use of polynomials can be motivated by Taylor's theorem. A well-behaved function f can be approximated about a point x by

$$f(x+\delta) \approx f(x) + f'(x)\delta + f''(x)\frac{\delta^2}{2!} + \cdots$$

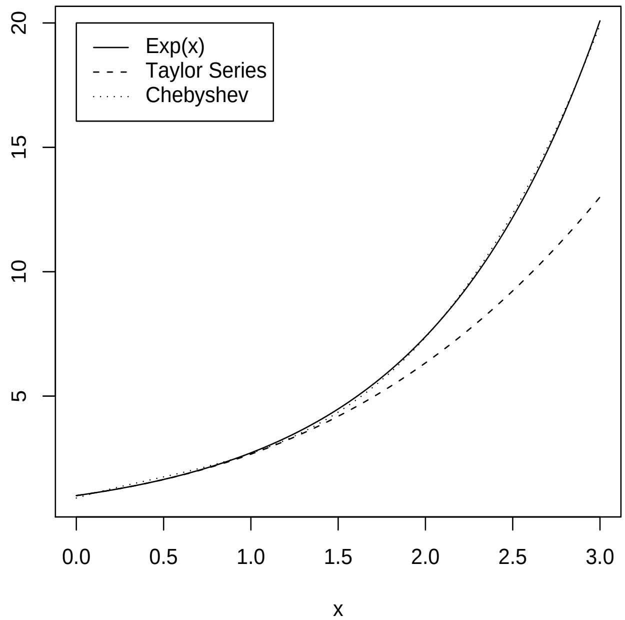

Figure 1: The cubic Taylor series approximation to e^x is accurate only near zero. The cubic Chebyshev polynomial approximation is indistinguishable from the function itself.

The right-hand size, which is a polynomial in δ, is an accurate approximation provided δ is small.

The trouble with Taylor's theorem is that the error of approximation is not evenly distributed. The approximation is accurate for x near zero but becomes poor for larger values of x. Consider the cubic Taylor series expansion for e^x about zero on the interval $[0, 3]$ (Figure 1). The approximation is accurate in a neighborhood of zero, but is very poor at the ends of the interval. Meanwhile there are other cubic polynomials which follow e^x with good accuracy over the entire interval. The holy Grail of polynomial approximation is to the find the polynomial which minimizes the maximum deviation of the polynomial from the function over the entire interval, the so-called *minimax* polynomial.

3 Orthogonal Polynomials

The general polynomial $p(x)$ above was written in terms of the *monomials* x^j. This is known as the *natural form* of the polynomial. The trouble with the natural form is that the monomials all look very similar when plotted on $[0, 1]$, i.e., they are a very highly correlated. This means small changes in $p(x)$ may arise from relatively large changes in the coefficients $c_0, \ldots, c_n$. The coefficients are not well determined when there is measurement or round-off error.

The general polynomial can just as well be written in terms of any sequence of basic polynomials of increasing degree,

$$p(x) = a_0 p_0(x) + a_1 p_1(x) + \ldots + a_n p_n(x)$$

where the degree of $p_j(x)$ is j for $j = 0, \ldots, n$. There is a linear relationship between the original coefficients c_j and the new coefficients a_j to make the resulting polynomial the same in each case.

The idea behind orthogonal polynomials is to select the basic polynomials $p_j(x)$ to be as different from each other as possible. Two polynomials p_i and p_j are said to be *orthogonal* if $p_i(X)$ and $p_j(X)$ are uncorrelated as X varies over some distribution. *Legendre polynomials* are uncorrelated when X is uniform on $(-1, 1)$. *Chebyshev polynomials* are uncorrelated when X is Beta$(\frac{1}{2}, \frac{1}{2})$ on $(-1, 1)$. *Laguerre polynomials* are uncorrelated when X is gamma on $(0, \infty)$. *Hermite polynomials* are uncorrelated when X is standard normal on $(-\infty, \infty)$.

Any sequence of orthogonal polynomials can be calculated recursively using a three-term recurrence formula. For example, the Chebyshev polynomials satisfy

$$\begin{aligned}
p_0(x) &= 1 \\
p_1(x) &= x \\
p_2(x) &= 2x^2 - 1 \\
&\cdots \\
p_{n+1}(x) &= 2xp_n(x) - p_{n-1}(x) \quad n \geq 1
\end{aligned}$$

Another important property of orthogonal polynomials is that $p_n(x)$ changes sign (and has a zero) n times in the interval of interest. The zeros of the nth order Chebyshev polynomial occur at

$$x_k = \cos\left(\pi \frac{k - 0.5}{n}\right), \quad k = 1, \ldots, n$$

The Chebyshev polynomials also have the property of bounded variation. The local maxima and minima of Chebyshev polynomials on $[-1, 1]$ are exactly equal to 1 and -1 respectively, regardless of the order of the polynomial. It is this property which makes them valuable for minimax approximation. In fact, an excellent approximation to the nth order minimax polynomial on an interval can be obtained by finding the polynomial which satisfies $p_n(x) = f(x)$ at the zeros of the $(n+1)$th order Chebyshev polynomial. Figure 1 shows the use of a 3rd order Chebyshev polynomial to approximate the function $\exp(x)$ on the interval $[0, 3]$. The error is less than 0.18 over the whole interval.

As another example, consider the problem of approximating the tail probability of the normal probability distribution function, $1 - \Phi(x)$, for $x > 0$. Since the tail probability decreases rapidly as x increases, we consider the ratio of the tail probability to the normal density function $[1 - \Phi(x)]/\phi(x)$. Finally, we transform

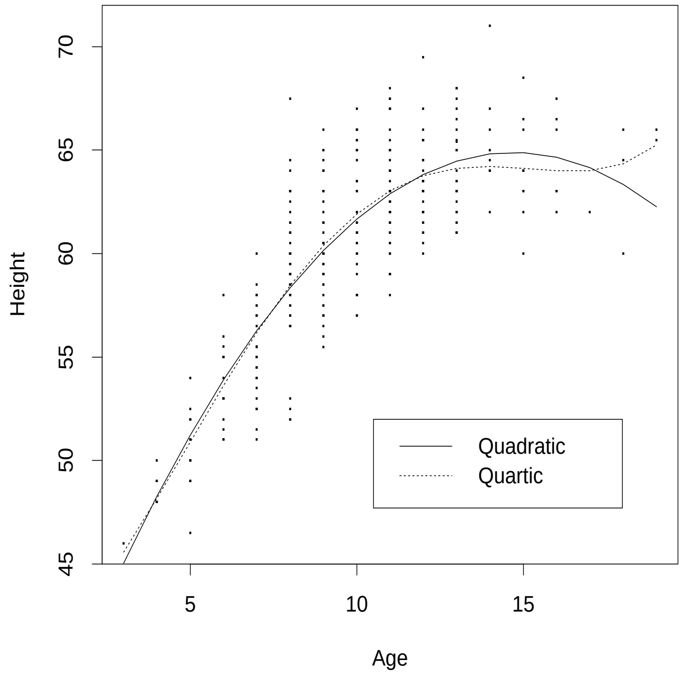

Figure 2: Height (inches) versus age (in years) for 318 girls who were seen in the Childhood Respiratory Disease Study in East Boston, Massachusetts.

x to $y = (x-1)/(x+1)$ which takes values on $[-1, 1]$. The resulting function $f(y) = [1 - \Phi(x(y))]/\phi(x(y))$ looks nearly linear and can be well approximated by a polynomial. The tenth-order polynomial which interpolates $f(y)$ on the Chebyshev points on $[-1, 1]$ approximates $f(y)$ to nine or more significant figures, and hence gives an approximation to $\Phi(x)$ which remains accurate to ten significant figures for very large values of x.

4 Polynomial Regression

Now we turn to the case in which the nonlinear function is observed with error. Suppose that we observe (x_i, y_i), $i = 1, \ldots, N$, where

$$y_i = f(x_i) + \epsilon_i$$

where f is some nonlinear function and the ϵ_i are uncorrelated with mean zero and constant variance.

Consider height as a function of age for 318 girls who were seen in a disease study [6] in East Boston in 1980 (Figure 2). Height might be described roughly by a straight line over a short range of ages, say ages 5 to 10, but over wider age ranges a more

Table 1: Coefficients and standard errors for polynomial regression of Height on Age for the respiratory disease study.

Coefficient	Value	Std. Error	t-value	P-value
β_0	80.2384	32.9342	2.4363	0.0154
β_1	-26.9075	23.0292	-1.1684	0.2435
β_2	7.8563	6.3456	1.2381	0.2166
β_3	-1.0296	0.8856	-1.1627	0.2459
β_4	0.0712	0.0663	1.0737	0.2838
β_5	-0.0025	0.0025	-1.0020	0.3171
β_6	0.0000	0.0000	0.9503	0.3427

Table 2: Correlation matrix for the polynomial regression coefficients.

	β_0	β_1	β_2	β_3	β_4	β_5
β_1	-0.9935					
β_2	0.9774	-0.9950				
β_3	-0.9558	0.9824	-0.9960			
β_4	0.9313	-0.9650	0.9860	-0.9969		
β_5	-0.9058	0.9451	-0.9721	0.9888	-0.9975	
β_6	0.8805	-0.9241	0.9559	-0.9776	0.9910	-0.9980

general function is needed. We fit initially a sixth order polynomial,

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \beta_5 x_i^5 + \beta_6 x_i^6 + \epsilon_i$$

with the intention of decreasing the order later if a simpler polynomial is found to fit just as well. This leads to a multiple linear regression problem for the coefficients $\beta_0, \ldots, \beta_6$ in which the design matrix is

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^6 \\ 1 & x_2 & x_2^2 & \ldots & x_2^6 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{318} & x_{318}^2 & \ldots & x_{318}^6 \end{pmatrix}$$

The columns of this matrix are very nearly collinear, which will make the least squares problem ill-conditioned. Nevertheless we can obtain results from a statistical package: the regression overall is highly significant with an F-statistic of 135.7 on 6 and 311 df. However the table of coefficients and standard errors offers little guidance as to what order of polynomial is required (Table 1). None of the regression coefficients are significantly different from zero, a reflection of the high correlations between the coefficients (Table 2. We could determine the smallest adequate order for the polynomial by fitting, in turn, a 5th-order polynomial, a 4th order, a 3rd-order and

Table 3: Coefficients and standard errors for orthogonal polynomial regression of Height on Age for the respiratory disease study.

Coefficient	Value	Std. Error	t-value	P-value
α_0	60.2119	0.1426	422.1543	0.0000
α_1	65.0285	2.5435	25.5669	0.0000
α_2	-31.3549	2.5435	-12.3276	0.0000
α_3	4.4838	2.5435	1.7629	0.0789
α_4	4.9562	2.5435	1.9486	0.0522
α_5	-2.1465	2.5435	-0.8439	0.3994
α_6	2.4170	2.5435	0.9503	0.3427

so on. At each step we could test for the neglected monomial term using an adjusted F-statistic. A more satisfactory solution however is to use orthogonal polynomials.

Many statistical programs allow one to compute a sequence of polynomials which are orthogonal with respect to the observed values of x, i.e., which satisfy

$$\sum_{k=1}^{N} p_i(x_k)p_j(x_k) = 0, \quad i \neq j.$$

(The function ORPOL is part of PROC MATRIX or PROC IML in SAS. In S-Plus or R the function is `poly`.) It is also convenient to choose the polynomials so that

$$\sum_{k=1}^{N} p_i(x_k)^2 = 1, \quad i = 0, 1, \ldots, N-1$$

In terms of these polynomials, the multiple regression model becomes

$$y_i = \alpha_0 p_0(x_i) + \alpha_1 p_1(x_i) + \cdots + \alpha_6 p_6(x_i) + \epsilon_i$$

where again there is a linear relationship between the coefficients α_j of the orthogonal polynomials and the original β_j. This model has the same fitted values, sums of squares and F-ratio as the original model. However, because of orthogonality, the least squares estimates of the α_j are uncorrelated and have identical standard errors, which greatly simplifies interpretation. In fact each estimated coefficient $\hat{\alpha}_j$ is unchanged by the actual order of the polynomial which has been fit.

Table 3 gives the estimated coefficients and standard errors for the orthogonal polynomial regression. In this case the t-statistics and P-values for the coefficients directly relate to the significance of cubic, quartic, and so on, components of the regression. We can see that the 5th and 6th order terms are not required, but that the 3rd and 4th order terms are approaching significance. A plot of the quadratic and quartic fitted values against Age shows that the quartic fit might be preferred because the quadratic is not monotonic in the observed range (Figure 2).

References

[1] Abramowitz, M., and Stegun, I. A. (1962). *Handbook of mathematical functions.* National Bureau of Standards, Washington. Reprinted by Dover, New York, 1965.

[2] Atkinson, K. E. (1989). *An Introduction to Numerical Analysis. 2nd Edition.* John Wiley, New York.

[3] Kleinbaum, D.G., Kupper, L.L., and Muller, K.E. (1988). Applied regression analysis and other multivariate methods. PWS-Kent, Boston.

[4] Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical recipes in Fortran.* Cambridge University Press. (Also available for C, Basic and Pascal.)

[5] Seber, G. A. F. (1977). *Linear regression analysis.* Wiley, New York.

[6] Tager, I. B., Weiss, S. T., Rosner, B., and Speizer, F. E. (1979). Effect of parental cigarette smoking on pulmonary function in children. *American Journal of Epidemiology*, **110**, 15–26.