Math/Stat 2300 Choosing Models (3.4)
from text *A First Course in Mathematical Modeling*, Giordano, Fox, Horton, Weir, 2009.

Qualitative description of the criteria

Minimizing the sum of the absolute deviations tends to treat each data point with equal weight and to average the deviations.

The Chebyshev criterion gives more weight to a single point potentially having a large deviation.

The Least-squares criterion is somewhere in between as far as weighing individual points with significant deviations

How do we relate the curve fitting criteria?

For *Chebyshev criterion*:
Suppose Chebyshev criterion is applied to some problem and the resulting optimization problem is solved to give the function $F(x)$. The absolute deviations resulting from this fit are defined as

$$|y_i - F(x_i)| = c_i, \quad i = 1, 2, \ldots, m$$

Let

c_{max} be defined as the largest of the absolute deviations c_i

This means

$$c_i \leq c_{max}, i = 1, 2, \ldots, m$$

What is c_{max}?

Since the parameters of $F(x)$ are determined so as to minimize the value of the largest deviation, c_{max} is the minimal largest absolute deviation

For *Least-Squares criterion*:
Suppose the Least-squares criterion is applied and the resulting optimization problem is solved to give the function $G(x)$. The absolute deviations resulting from this fit are defined as

$$|y_i - G(x_i)| = d_i, i = 1, 2, \ldots, m$$

Let

d_{max} be defined as the largest of the absolute deviations d_i

Note that the sum of the d_i squares is the smallest such sum.
$d_1^2 + d_2^2 + \ldots + d_m^2 \leq c_1^2 + c_2^2 + \ldots + c_m^2$
$c_{max} \leq d_{max}$

Because $c_i \leq c_{max}$ for every i,

$$d_1^2 + d_2^2 + \ldots + d_m^2 \leq c_{max}^2 + \ldots c_{max}^2 + \ldots + c_{max}^2$$
$$d_1^2 + d_2^2 + \ldots + d_m^2 \leq mc_{max}^2$$
$$\sqrt{\frac{d_1^2 + d_2^2 + \ldots + d_m^2}{m}} \leq c_{max} \leq d_{max}$$

Choosing a Best Model

How can we evaluate how well the model fits the data?

Example. Let's compare the different ways of fitting data:
Previously, we fit the data

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

to the curve $y = Ax^2$.

Using Least-squares, we determined

$$y = 3.1869x^2 \approx 3.19x^2$$

Using transformed Least-squares, we determined

$$\ln a = 1.14322$$

giving

$$y = 3.1368x^2 \approx 3.14x^2$$

For the sake of comparing fitting methods, if the optimization problem resulting from applying the Chebyshev criterion was solved, we would obtain

$$y = 3.17073x^2 \approx 3.17x^2$$

To compare the methods, we consider the deviations

x_i	y_i	$y_i - 3.19x_i^2$	$y_i - 3.14x_i^2$	$y_i - 3.17x_i^2$
0.5	0.7	-0.0975	-0.085	-0.0925
1.0	3.4	0.21	0.26	0.23
1.5	7.2	0.0225	0.135	0.0675
2.0	12.4	-0.36	-0.16	-0.28
2.5	20.1	0.1625	0.475	0.2875

Criterion	Model	$\sum[y_i - y(x_i)]^2$	$\text{Max}\lvert y_i - y(x_i)\rvert$
Least-Squares	$y = 3.19x^2$	0.21	0.36
Transformed Least-Squares	$y = 3.14x^2$	0.34	0.475
Chebyshev	$y = 3.17x^2$	0.23	0.2875

Is one of the models better than the others? Why?

maybe or maybe not
We could choose the model with the smallest absolute deviation (Chebyshev) or choose model with smallest sum of squares (Least Squares)

the model with the smallest absolute deviation or smallest sum of squares may fit very poorly over the range you intend to use it most

Some considerations:

interpolation - construct models that pass through each data point $\Longrightarrow$ will give zero sum of squares and zero maximum deviation
Which model is best must be decided on a case-by-case basis.

Math/Stat 2300 Interpolation

Higher Order Polynomials Models (4.2)
Recall *polynomials* are functions of the form

$$P_n(x) = a_n x^n + a_{n-1}x^{n-1} + \dots a_1 x + a_0$$

where n is the *degree* of the polynomial and $a_0, \dots, a_n$ are real constants.

A unique polynomial of degree of at most degree 2 can be passed through three data points BUT an infinite number of polynomials of any degree greater than 2 can be passed through three data points.

We call a polynomial that interpolates a set of data points an *interpolant*.

Consider the following data

x	1	2	3	4	5	6
y	205	430	677	945	1233	1542

Because we have 6 data points, a unique polynomial of at most degree 5 is expected to interpolate the data

$$P_5(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5$$

We can use the 6 data points to determine the constants $a_0, a_1, \dots, a_5$:

$$\begin{aligned}
205 &= a_0 + a_1(1) + a_2(1)^2 + a_3(1)^3 + a_4(1)^4 + a_5(1)^5 \\
430 &= a_0 + a_1(2) + a_2(2)^2 + a_3(2)^3 + a_4(2)^4 + a_5(2)^5 \\
&\vdots \\
1542 &= a_0 + a_1(6) + a_2(6)^2 + a_3(6)^3 + a_4(6)^4 + a_5(6)^5
\end{aligned}$$

We would solve this system of linear equations to determine the constants $a_0, a_1, \dots, a_5$.

For more data points, we would have larger systems of equations to solve. Large systems of linear equations can be difficult to solve with great numerical precision.

Lagrange Polynomials

There is an alternative form of the interpolant: **Lagrangian form of the polynomial**

Suppose the following data has been collected

x	x_1	x_2	x_3	x_4
y	y_1	y_2	y_3	y_4

Consider the following cubic polynomial

$$\begin{aligned} P_3(x) &= \frac{(x-x_2)(x-x_3)(x-x_4)}{(x_1-x_2)(x_1-x_3)(x_1-x_4)}y_1 + \frac{(x-x_1)(x-x_3)(x-x_4)}{(x_2-x_1)(x_2-x_3)(x_2-x_4)}y_2 \\ &\quad + \frac{(x-x_1)(x-x_2)(x-x_4)}{(x_3-x_1)(x_3-x_2)(x_3-x_4)}y_3 + \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_4-x_1)(x_4-x_2)(x_4-x_3)}y_4 \end{aligned}$$

Does this interpolate the data?

YES

$$\begin{aligned} P_3(x_1) &= \frac{(x_1-x_2)(x_1-x_3)(x_1-x_4)}{(x_1-x_2)(x_1-x_3)(x_1-x_4)}y_1 + \frac{(x_1-x_1)(x_1-x_3)(x_1-x_4)}{(x_2-x_1)(x_2-x_3)(x_2-x_4)}y_2 \\ &\quad + \frac{(x_1-x_1)(x_1-x_2)(x_1-x_4)}{(x_3-x_1)(x_3-x_2)(x_3-x_4)}y_3 + \frac{(x_1-x_1)(x_1-x_2)(x_1-x_3)}{(x_4-x_1)(x_4-x_2)(x_4-x_3)}y_4 = y_1 \end{aligned}$$

Similarly,

$$P_3(x_2) = y_2, \;\; P_3(x_3) = y_3, \;\; P_3(x_4) = y_4$$

When does the polynomial not exist?

When any of the x_i are equal.

The x_i values must all be different (to prevent division by zero).

Theorem

If $x_0, x_1, \ldots, x_n$ are $(n+1)$ distinct points and $y_0, y_1, \ldots, y_n$ are corresponding observations at these points, then there exists a unique polynomial $P(x)$, of at most degree n, with the property that

$$y_k = P(x_k) \quad \text{for each } k = 0, 1, \ldots, n$$

This polynomial is given by

$$P(x) = y_0 L_0(x) + \ldots + y_n L_n(x)$$

where

$$L_k(x) = \frac{(x-x_0)(x-x_1)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)}{(x_k-x_0)(x_k-x_1)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)}$$

What are some advantages of using higher order polynomials?

Because the polynomial passes through each of the data points, the resultant sum of absolute deviations is zero.

easy to integrate and differentiate

What are some disadvantages?

High order polynomials oscillate severely near the end points

Coefficients of high order polynomials are sensitive to small changes in the data

Example. Using the following data, determine the second order interpolating polynomial

x	2	2.5	4
y	0.5	0.4	0.25

Let $x_0 = 2$**,** $x_1 = 2.5$**,** $x_2 = 4$**.**

$$\begin{aligned} L_0(x) &= \frac{(x-2.5)(x-4)}{(2-2.5)(2-4)} = (x-2.5)(x-4) \\ L_1(x) &= \frac{(x-2)(x-4)}{(2.5-2)(2.5-4)} = \frac{-4}{3}(x-2)(x-4) \\ L_2(x) &= \frac{(x-2)(x-2.5)}{(4-2)(4-2.5)} = \frac{1}{3}(x-2)(x-2.5) \end{aligned}$$

$$P(x) = L_0(x)y_0 + L_1(x)y_1 + L_2(x)y_2$$

$$P(x) = (x-2.5)(x-4)(0.5) + (-4/3)(x-2)(x-4)(0.4) + (1/3)(x-2)(x-2.5)(0.25)$$