MATH 3290 Mathematical Modeling

Chapter 3: Model Fitting

Kuang HUANG

January 26, 2024

Department of Mathematics
The Chinese University of Hong Kong

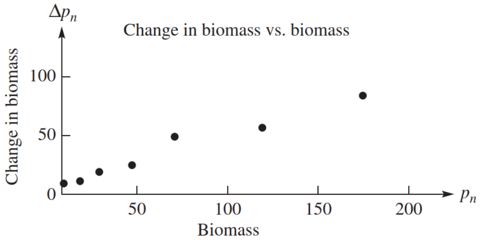
Course webpage

https://www.math.cuhk.edu.hk/course/2324/math3290

SCAN ME

In Chap. 1 (in the example of yeast population), we encounter the problem of finding a function that explains the data.

Time in hours n	Observed yeast biomass p_n	Change in biomass $p_{n+1} - p_n$
0	9.6	8.7
1	18.3	10.7
2	29.0	18.2
3	47.2	23.9
4	71.1	48.0
5	119.1	55.5
6	174.6	82.7
7	257.3	



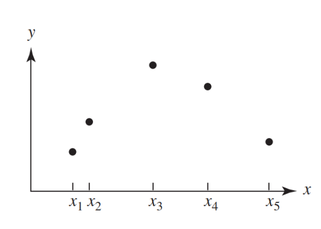
We want to find a constant $k > 0$ such that

$$\Delta p_n = k\, p_n.$$

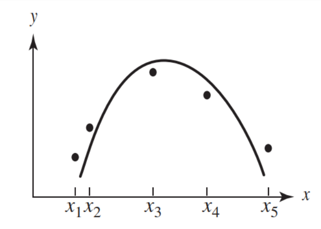
Model fitting

Given a set of data points, we choose a curve (i.e. a function) that best fits the data.

Then we can use the function to make predictions.



Measurement data



The best fitted curve

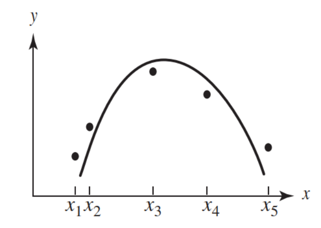
One can predict the value of y for $x_1 \leq x \leq x_5$.

Two main steps

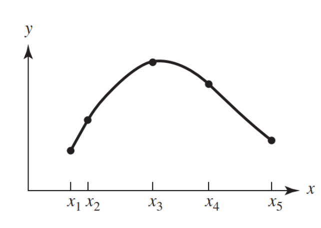
There are two main steps in model fitting.

- When a given model type is chosen, how to find the parameters in the model— e.g., in the yeast population example, we have chosen the model function $\Delta p_n = kp_n$, and our task is to find k.
- When the data set is given, how do you choose the most suitable model function— e.g. in the yeast population example, how do you make the decision of using the model function $\Delta p_n = kp_n$.

Model fitting vs interpolation



- Model fitting (Chap. 3),
- Curve may not meet the points.
- Errors in data expected.
- Theory-driven = a particular form of model function is assumed.



- Interpolation (Chap. 4).
- Curve goes through all points.
- Data are accurate.
- Data-driven = use the data to find the form of the model function.

Methods of model fitting

Given a set of m data points: $(x_i, y_i), \quad i = 1, 2, \dots, m$

Given a type of model function $y = f(x; \theta)$, depending on some parameters θ.

E.g. $f(x; \theta) = ax + b$, $ax^2 + bx + c$, ae^{bx}, $\dots$

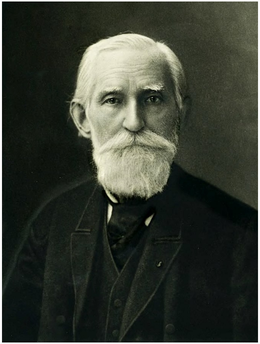
Objective: find the model function that best fits the data

We need to give a precise meaning of best.

There are three commonly used criteria.

Chebyshev criterion

The first one is the Chebyshev criterion.



Pafnuty Chebyshev
(1821-1984)

Pafnuty Chebyshev, is known for Chebyshev's polynomials, Chebyshev's inequality...

Chebyshev criterion

The first one is the Chebyshev criterion.

We will find the parameters in the model function $f(x;\theta)$ such that the largest absolute deviation is minimized.

That is, we will minimize the value

$$\max_{i=1,\dots,m} |y_i - f(x_i;\theta)|.$$

Remarks:

- It is not an easy mathematical problem.
- In some cases, one needs to solve a linear programming problem (see Chap. 7).

Min sum of absolute deviations

The second method is minimizing the sum of the absolute deviations.

We will find the parameters in the model function $f(x;\theta)$ such that the the sum of the absolute deviations is minimized.

That is, we will minimize the value

$$\sum_{i=1}^{m} |y_i - f(x_i;\theta)|.$$

This is again a difficult mathematical problem. E.g. one cannot use calculus techniques to find the minimum.

Least-squares criterion

The third method is the least-squares criterion.

We will find the parameters in the model function $f(x;\theta)$ such that the the sum of the squared deviations is minimized.

That is, we will minimize the value

$$\sum_{i=1}^{m} |y_i - f(x_i;\theta)|^2.$$

Remarks:

- It is a very popular method.
- The solution can be easily obtained by calculus methods.

Connection of two criteria

We give a relation of Chebyshev and least-squares criteria.

- Suppose a function $g_1(x) = f(x; \theta_1)$ is obtained by Chebyshev criterion.
  - We define $c_i = |y_i - g_1(x_i)|, \quad i = 1, 2, \ldots, m.$
  - We define $c_{\max} = \max_i c_i$ the maximum deviation.
  - The Chebyshev criterion implies $g_1(x)$ is chosen so that $c_{\max}$ is smallest among all choices of parameters θ.
- Suppose a function $g_2(x) = f(x; \theta_2)$ is obtained by least-squares criterion.
  - We define $d_i = |y_i - g_2(x_i)|, \quad i = 1, 2, \ldots, m.$
  - We define $d_{\max} = \max d_i$ the maximum deviation.

Then we have the following conclusions.

- By the Chebyshev criterion, $c_{\max} \leq d_{\max}$.
- By the least-squares criterion,

$$
\begin{aligned}
& d_1^2 + d_2^2 + \cdots + d_m^2 \leq c_1^2 + c_2^2 + \cdots + c_m^2 \\
\Rightarrow\ & d_1^2 + d_2^2 + \cdots + d_m^2 \leq c_{\max}^2 + c_{\max}^2 + \cdots + c_{\max}^2 = mc_{\max}^2 \\
\Rightarrow\ & \sqrt{\frac{d_1^2 + d_2^2 + \cdots + d_m^2}{m}} \leq c_{\max}.
\end{aligned}
$$

Combining above

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

We have the following relationship for the above two criteria

$$D := \sqrt{\frac{d_1^2 + d_2^2 + \cdots + d_m^2}{m}} \leq c_{\max} \leq d_{\max}.$$

Suppose you care the maximum deviation, and you know that it is more convenient to use the least-squares criterion.

- If D and $d_{\max}$ are close, then the solution obtained by the least-squares criterion is a good approximation.
- If D and $d_{\max}$ are not close, then one should use the Chebyshev criterion.

Applying least-squares criterion

Given a data set $(x_i, y_i), \quad i = 1, 2, \ldots, m.$

Assume that the model function is $y = f(x; p_1, \ldots, p_k)$, where $p_1, p_2, \ldots, p_k$ are the model parameters.

Consider the least-squares criterion: find $p_1, p_2, \ldots, p_k$ so that

$$S(p_1, p_2, \ldots, p_k) := \sum_{i=1}^{m} |y_i - f(x_i; p_1, p_2, \ldots, p_k)|^2$$

is minimized.

Use standard calculus method, find the solution by

$$\frac{\partial S}{\partial p_j} = -2 \sum_{i=1}^{m} (y_i - f(x_i)) \frac{\partial f}{\partial p_j} = 0, \quad j = 1, 2, \ldots, k.$$

Fitting a line

Assume that the model function is a line

$$y = f(x; a, b) = ax + b.$$

Then we need to minimize

$$S(a, b) = \sum_{i=1}^{m} |y_i - (ax_i + b)|^2.$$

Now we find the partial derivatives

$$\begin{aligned} \frac{\partial S}{\partial a} &= \sum_{i=1}^{m} \{-2x_i(y_i - ax_i - b)\}, \\ \frac{\partial S}{\partial b} &= \sum_{i=1}^{m} \{-2(y_i - ax_i - b)\}. \end{aligned}$$

To find the minimum, we need

$$
\begin{aligned}
0 = \frac{\partial S}{\partial a} &= \sum_{i=1}^{m} \{-2x_i(y_i - ax_i - b)\}, \\
0 = \frac{\partial S}{\partial b} &= \sum_{i=1}^{m} \{-2(y_i - ax_i - b)\}.
\end{aligned}
$$

So, we have

$$
\begin{aligned}
a\left(\sum_{i=1}^{m} x_i^2\right) + b\left(\sum_{i=1}^{m} x_i\right) &= \sum_{i=1}^{m} x_i y_i, \\
a\left(\sum_{i=1}^{m} x_i\right) + b\left(\sum_{i=1}^{m} 1\right) &= \sum_{i=1}^{m} y_i.
\end{aligned}
$$

One can then find a and b by solving the above linear system.

Consider the data set

x	1	2	3	4
y	8.1	22.1	60.1	165

Then we have

$$\sum_{i=1}^{m} x_i^2 = 30, \quad \sum_{i=1}^{m} x_i = 10, \quad \sum_{i=1}^{m} 1 = 4,$$

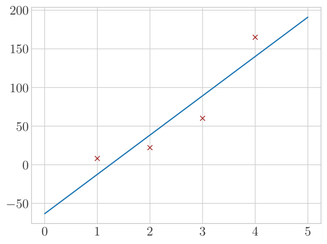
$$\sum_{i=1}^{m} x_i y_i = 892.6, \quad \sum_{i=1}^{m} y_i = 255.3.$$

The linear system is

$$30a + 10b = 892.6, \quad 10a + 4b = 255.3.$$

Solving, we have $a = 50.87$ and $b = -63.35$.

Hence, our model function is $f^*(x) = 50.87x - 63.35$.



One can use the model to predict the value at $x = 2.5$

$$y = f^*(2.5) = 50.87(2.5) - 63.35 = 63.825.$$

Fitting a more general model

Assume that the model function is

$$f(x; a, b) = ag(x) + bh(x).$$

Then we need to minimize

$$S(a, b) = \sum_{i=1}^{m} \left| y_i - \Big(ag(x_i) + bh(x_i) \Big) \right|^2.$$

Taking partial derivatives,

$$\begin{aligned} \frac{\partial S}{\partial a} &= \sum_{i=1}^{m} \Big\{ -2g(x_i) \Big(y_i - ag(x_i) - bh(x_i) \Big) \Big\}, \\ \frac{\partial S}{\partial b} &= \sum_{i=1}^{m} \Big\{ -2h(x_i) \Big(y_i - ag(x_i) - bh(x_i) \Big) \Big\}. \end{aligned}$$

To find the minimum, we solve

$$
\begin{aligned}
0 = \frac{\partial S}{\partial a} &= \sum_{i=1}^{m} \left\{ -2g(x_i)\Big(y_i - ag(x_i) - bh(x_i)\Big) \right\}, \\
0 = \frac{\partial S}{\partial b} &= \sum_{i=1}^{m} \left\{ -2h(x_i)\Big(y_i - ag(x_i) - bh(x_i)\Big) \right\}.
\end{aligned}
$$

We obtain the linear system

$$
\begin{aligned}
a\Big(\sum_{i=1}^{m} g(x_i)^2\Big) + b\Big(\sum_{i=1}^{m} g(x_i)h(x_i)\Big) &= \sum_{i=1}^{m} g(x_i)\, y_i, \\
a\Big(\sum_{i=1}^{m} g(x_i)h(x_i)\Big) + b\Big(\sum_{i=1}^{m} h(x_i)^2\Big) &= \sum_{i=1}^{m} h(x_i)\, y_i.
\end{aligned}
$$

Consider the data set

x	-1	$-\frac{1}{2}$	0	$\frac{1}{2}$	1
y	-1	0	1	2	1

and the model function

$$f(x; a, b) = a\cos(\pi x) + b\sin(\pi x).$$

Then we define

$$g(x) = \cos(\pi x), \quad h(x) = \sin(\pi x).$$

To find the linear system, we need

$$\sum_{i=1}^{m} g(x_i)^2 = 3, \quad \sum_{i=1}^{m} g(x_i)h(x_i) = 0, \quad \sum_{i=1}^{m} h(x_i)^2 = 2,$$

$$\sum_{i=1}^{m} g(x_i)\, y_i = 1, \quad \sum_{i=1}^{m} h(x_i)\, y_i = 2.$$

Hence, the linear system

$$\begin{aligned} a\Big(\sum_{i=1}^{m} g(x_i)^2\Big) + b\Big(\sum_{i=1}^{m} g(x_i)h(x_i)\Big) &= \sum_{i=1}^{m} g(x_i)\, y_i, \\ a\Big(\sum_{i=1}^{m} g(x_i)h(x_i)\Big) + b\Big(\sum_{i=1}^{m} h(x_i)^2\Big) &= \sum_{i=1}^{m} h(x_i)\, y_i. \end{aligned}$$

To find the linear system, we need

$$\sum_{i=1}^{m} g(x_i)^2 = 3, \quad \sum_{i=1}^{m} g(x_i)h(x_i) = 0, \quad \sum_{i=1}^{m} h(x_i)^2 = 2,$$

$$\sum_{i=1}^{m} g(x_i)\, y_i = 1, \quad \sum_{i=1}^{m} h(x_i)\, y_i = 2.$$

becomes

$$3a + 0b = 1, \quad 0a + 2b = 2.$$

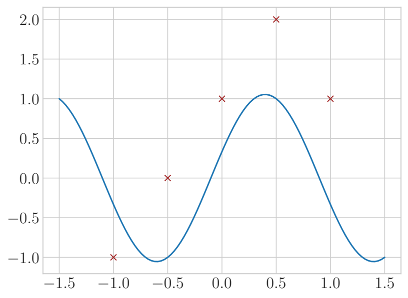
To find the linear system, we need

$$\sum_{i=1}^{m} g(x_i)^2 = 3, \quad \sum_{i=1}^{m} g(x_i)h(x_i) = 0, \quad \sum_{i=1}^{m} h(x_i)^2 = 2,$$

$$\sum_{i=1}^{m} g(x_i)\, y_i = 1, \quad \sum_{i=1}^{m} h(x_i)\, y_i = 2.$$

We have $a = 1/3$ and $b = 1$.

Hence, the model function is $f^*(x) = \frac{1}{3}\cos(\pi x) + \sin(\pi x)$.



Note that the trigonometric polynomial approximation

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{N} a_n \cos(n\pi x) + b_n \sin(n\pi x)$$

is widely used in many engineering areas.

Transformed least-squares fitting

Consider the model function $y = f(x; a, b) = be^{ax}$.

To fit this function to the data, we minimize

$$S(a, b) = \sum_{i=1}^{m} |y_i - be^{ax_i}|^2.$$

Taking partial derivatives,

$$\begin{aligned} \frac{\partial S}{\partial a} &= \sum_{i=1}^{m} \Big(-2bx_i e^{ax_i}(y_i - be^{ax_i}) \Big), \\ \frac{\partial S}{\partial b} &= \sum_{i=1}^{m} \Big(-2e^{ax_i}(y_i - be^{ax_i}) \Big). \end{aligned}$$

Note: the model function depends nonlinearly on a and b.

Given the model function $y = be^{ax}$, we have

$$\ln y = \ln b + ax.$$

Introduce the new variables $\tilde{y} = \ln y$ and $\tilde{b} = \ln b$.

Now consider the data set $(x_i, \tilde{y}_i)$, and the model function $\tilde{y} = \tilde{b} + ax$.

Note: the new model depends linearly on a and $\tilde{b}$.

Now,

$$\begin{aligned} a\Big(\sum_{i=1}^{m} x_i^2\Big) + \tilde{b}\Big(\sum_{i=1}^{m} x_i\Big) &= \sum_{i=1}^{m} x_i \tilde{y}_i, \\ a\Big(\sum_{i=1}^{m} x_i\Big) + \tilde{b}\Big(\sum_{i=1}^{m} 1\Big) &= \sum_{i=1}^{m} \tilde{y}_i. \end{aligned}$$

We can get a and $\tilde{b}$. Then $b = e^{\tilde{b}}$.

Consider the data set

x	1	2	3	4
y	8.1	22.1	60.1	165

We will fit the model function $y = be^{ax}$ to the data by the transformed least-squares criterion.

The transformed data set is

x	1	2	3	4
$\ln y$	2.1	3.1	4.1	5.1

x	1	2	3	4
$\ln y$	2.1	3.1	4.1	5.1

Then we have

$$\sum_{i=1}^{m} x_i \tilde{y}_i = 41, \quad \sum_{i=1}^{m} \tilde{y}_i = 14.4.$$

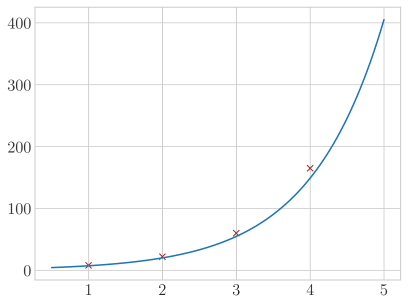
The linear system is

$$30a + 10\tilde{b} = 41, \quad 10a + 4\tilde{b} = 14.4.$$

Solving, we obtain $a = 1$ and $\tilde{b} = 1.1$.

Hence, the model function is $y = e^{\tilde{b}} e^{x} = 3.0042 e^{x}$.

x	1	2	3	4
$\ln y$	2.1	3.1	4.1	5.1



Polynomial approximation

If we consider in a continuous level, i.e., we have a data set $(x, f(x))$ where $x \in [-1, 1]$ $(f(x) \in C[-1, 1])$. We want to find a polynomial $p(x)$ whose degree is no greater than n $(p(x) \in \mathcal{P}_n)$ that best fits the data.

Why polynomials?

- Evaluations of polynomials only require additions and multiplications, which computers are very good at.
- We have a lot of fast algorithms available. For example, Horner's method—

$$
\begin{aligned}
&p(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + a_n x^n \Longrightarrow \\
&b_n = a_n, \\
&b_{n-1} = a_{n-1} + b_n x, \\
&\ldots, \\
&b_0 = a_0 + b_1 x \longrightarrow p(x).
\end{aligned}
$$

Three criteria again

- The Chebyshev criterion

$$\min_{p\in\mathcal{P}_n} \|p-f\|_{L^\infty} = \min_{p\in\mathcal{P}_n} \max_{x\in[-1,1]} |p(x)-f(x)|.$$

- Minimize the sum of the absolute deviations

$$\min_{p\in\mathcal{P}_n} \|p-f\|_{L^1} = \min_{p\in\mathcal{P}_n} \int_{-1}^{1} |p(x)-f(x)|\,\mathrm{d}x.$$

- The least-squares criterion

$$\min_{p\in\mathcal{P}_n} \|p-f\|_{L^2}^2 = \min_{p\in\mathcal{P}_n} \int_{-1}^{1} |p(x)-f(x)|^2\,\mathrm{d}x.$$

least-squares criterion

Note that $p(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + a_n x^n$, and we need to determine $a_0, a_1, \ldots, a_n$.

According to the least-squares criterion, we have

$$\int_{-1}^{1} |p(x) - f(x)|^2 \,\mathrm{d}x = \int_{-1}^{1} |p(x)|^2 - 2p(x)f(x) + |f(x)|^2 \,\mathrm{d}x$$

$$= \sum_{i,j=0,\ldots,n} a_i a_j \int_{-1}^{1} x^{i+j} \,\mathrm{d}x - \sum_{i=0,\ldots,n} 2a_i \int_{-1}^{1} f(x)x^i \,\mathrm{d}x + \int_{-1}^{1} |f(x)|^2 \,\mathrm{d}x.$$

Take partial derivatives, we obtain a linear system for $a = (a_0, a_1, \ldots, a_n)^\mathsf{T}$

$$Ma = b,$$

where $M_{i,j} = \int_{-1}^{1} x^{i+j} \,\mathrm{d}x$ and $b_i = \int_{-1}^{1} f(x)x^i \,\mathrm{d}x$.

Disclaimer

All figures, tables, and data appearing in the slides are only used for teaching under guidelines of Fair Use.