Lecture 4

Model Fitting

PRELIMINARIES

Data and modeling: three situations

Model fitting

Model fitting

Inter-polation

extrapolation

Parameter identification / calibration!

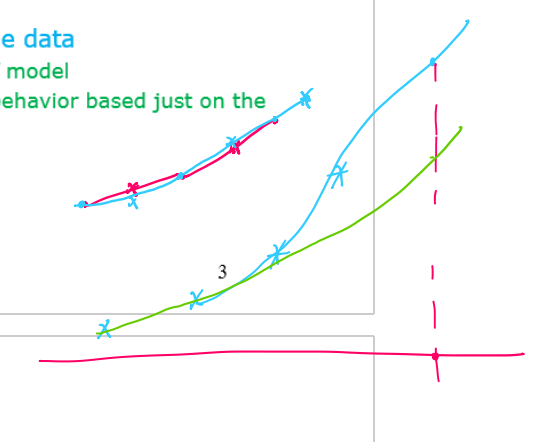
- Given a data set we may take 3 approaches
  1. Fit an already selected model type to the data
     - The model type is already fixed
     - For example: a linear, or a quadratic model; mass-action model
  2. Choose the most appropriate model from several alternative models that have been fitted to the data
     - Decide whether the best-fitting exponential model is better than the best-fitting polynomial model
  3. Make predictions based solely on the data
     - No hypothesis regarding the type of model
     - Predict intermediate and/or future behavior based just on the data set

(x_i, y_i)

$f(\alpha, \beta, x) = y$

f Given. / to decide $\alpha, \beta, \dots$

blood pressure	blood oxygen
60	1
72.5	2
85	3
100	4
105	5
	6



Model fitting vs. interpolation

- **Model fitting** model is known. but parameters are unknown. $\alpha, \beta, \dots$
  - The modeler has a hypothesis regarding the mathematical form of the model (s)he is building
  - It is just a matter of finding the numerical parameters that make the chosen model explain (fit) the experimental data best
  - Some deviations are going to be willingly accepted
  - Emphasis on the model exact, data inexact
- **Interpolation**
  - No a-priori hypothesis regarding the mathematical form of the model
  - The modeler is strongly guided by the data
  - Aims to capture the trend of the data to predict in-between the data points (and sometimes outside the data interval)
  - Emphasis on the data exact, model inexact!
- **Fitting and interpolating at the same time**
  - A fitted model may need to be replaced with an interpolating curve
  - For example because the interpolating curve may have better mathematical properties, needed for model analysis
  - Called sometimes model approximation

Example: model fitting vs. interpolation

- We are given a set of data
- Two approaches
  - Model fitting: we look for a quadratic polynomial to explain the data
  - Interpolation: look for a curve going exactly through the data points

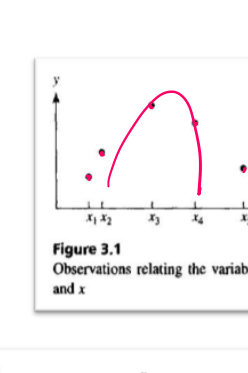


Figure 3.1 Observations relating the variables y and x

$y = ax^2 + bx + c$

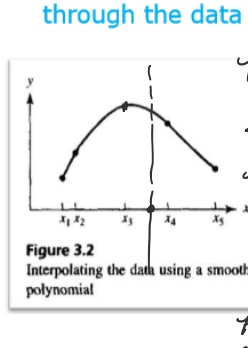


Figure 3.2 Interpolating the data using a smooth polynomial

piecewise quadratic interpolation

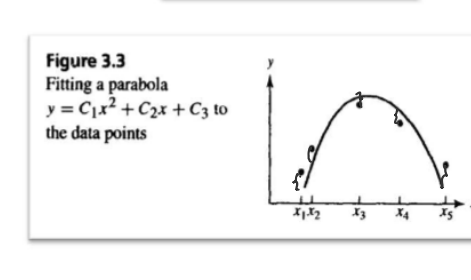
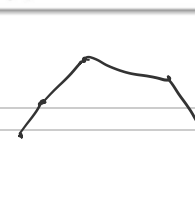


Figure 3.3 Fitting a parabola $y = C_1x^2 + C_2x + C_3$ to the data points



piecewise linear interpolation

Sources of error in modeling

- Formulation errors
  - Result from errors in the model formulation
  - Significant variables were ignored
  - Interrelationships between variables were ignored or simplified
  - Relating the data to the model in the wrong way: see for example reporter systems
- Truncation errors $e^x =$ $\pi = \dots$
  - Come from the math techniques used in building the model
  - For example, an infinite series expansion may be truncated to a polynomial
- Round-off errors
  - Numerical errors coming from representing real numbers with finite precision
- Measurement errors
  - Imprecision in the collection of data
  - Physical limitations of the instruments
  - Human errors

UQ: Uncertainty quantification!

Content of this lecture

- Model fitting
  - Visual model fitting
  - (Confidence intervals)
  - Analytic model fitting
  - Choosing the best model

Visual model fitting

- The modeler has a hypothesis regarding the type of model to be fitted
- The numerical value of parameters needs to be fixed so that the model explains the available data
  - The range of the parameters is known
- Data set
  - The size of the data set is a trade-off between the cost of obtaining the data and the accuracy to be obtained for the model
  - Minimum: at least as many data points as the number of parameters to fix
  - Spacing of the data points important
    - More data points on those intervals where the model should be fitted particularly well,
    - or where the maximum use of the model is expected,
    - or where abrupt changes in the model behavior are expected
  - Use the data together with its confidence intervals

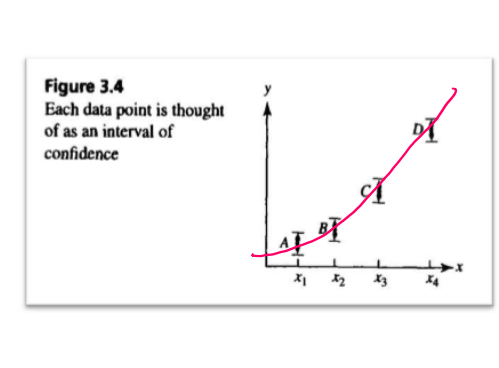


Figure 3.4 Each data point is thought of as an interval of confidence

errorbar!

Giordano et al. A first course in mathematical modeling. (3rd edition) Figure 3.4, page 101

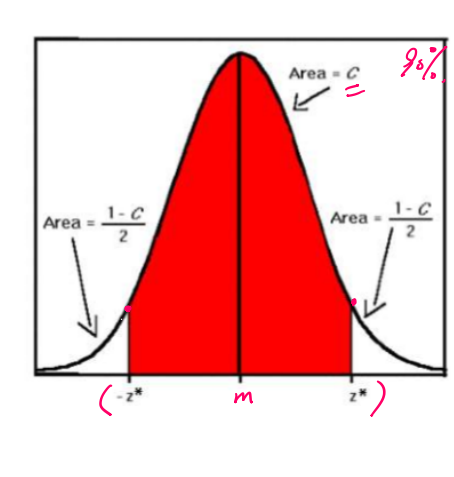
About confidence intervals

Confidence intervals

- Setup: a population and a parameter taking different values for different members of the population.
  - Goal: We need to estimate the population mean value of that parameter only through a population sample
  - Outcome: an interval where (the true value of) the population mean lies with large probability
  - The interval is called confidence interval 置信区间
  - The probability is called confidence level 置信水平
- Example
  - Measure the voting intentions for a certain party in an election
    - The answer may be presented as 40%
    - It should be interpreted as an interval centered at 40% that is smaller or larger depending on the confidence level
    - The 90% confidence interval that could be calculated from the data might be for example 38% to 42%, while the 95% interval might be for example 36% to 44%

Confidence intervals (cont.): mathematical setup

- For a given sample, its mean is a stochastic variable
- If the measurements follow a normal distribution, the sample mean will have the distribution N(μ,σ/n½)
- From the calculated sample mean we need to report an interval where the real population mean μ lies with large probability
- The interval is related to the percentages of the area of the normal density curve
  - For example a 95% confidence interval should cover 95% of the area under the normal curve
- The value z* of the point on the standard normal density curve N(0,1) such that the probability of observing a value greater than z* is equal to p is known as the upper p-critical value of the standard normal distribution.



90% 95%

(−z* m z*)

Confidence intervals: unknown mean, known standard deviation

- Confidence intervals for unknown mean μ and known standard deviation σ
  - Given a sample of size n with mean m, a C-confidence interval for the population mean is (m-z*σ/n½, m+z*σ/n½), where z* is the upper (1-C)/2 critical value for the standard normal dsitribution 2.5%
  - The error margin is z*σ/n½ ×2
- Note: the interval above is exact only for populations that are normally distributed. For other populations, the interval is approximately correct for large samples by the central limit theorem

$\{X_i\}_{i=1\dots n}$

$\bar{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$ μ unknown

σ^2

Confidence intervals: unknown mean, unknown standard deviation

- Confidence intervals for unknown mean μ and unknown standard deviation σ
  - The unknown standard deviation is replaced by the estimated standard deviation s: s²=1/(n-1)∑(xᵢ-m)²
  - Given a sample of size n, its mean follows the t distribution t(n-1) with mean μ and standard deviation s/n½
  - As the sample size n increases, the t distribution approaches the normal distribution (because the estimated standard deviation approaches the standard deviation)
  - A C-confidence interval for the population mean is (m-t*s/n½, m+t*s/n½), where t* is the upper (1-C)/2 critical value for the t(n-1) distribution 2.5%
  - The error margin is t*s/n½

μ unknown

σ unknown

$\bar{\mu} = \frac{1}{n}\sum X_i$

$\hat{\sigma}^2 = \frac{1}{n-1}\sum (X_i - \bar{\mu})^2$

HW3: 1. Generate k groups of random numbers

rand ...

randn ...

rand(...) ...

type

N: 10, 100, 1000 ...

errorbar

Back to visual model fitting

Visual model fitting

- Using the original data
  - Look at the deviations between the model prediction and the available data set
  - Aim to minimize the deviations
    - The largest deviation $\min \max_i |e_i|$
    - Sum of all deviations $\min \sum_i |e_i|$
    - The sum of squares of deviations $\min \sum_i |e_i|^2$
    - …
  - Example: fit a linear model y=ax+b to the data on the previous slide
- Note: Although these methods of visually fitting the data may seem imprecise, they might be quite compatible with the accuracy of the modeling process
  - Grossness of the assumptions and the imprecision in the data collection may not warrantee a more sophisticated analysis

? $\min \sum_i |e_i|$

p = 3, 4, ...

p = 1.5

0.5

0

−1

nonzero

$\{e_i\}$

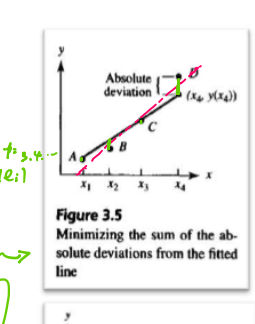


Figure 3.5 Minimizing the sum of the absolute deviations from the fitted line

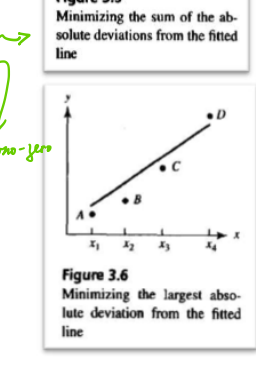


Figure 3.6 Minimizing the largest absolute deviation from the fitted line

Giordano et al. A first course in mathematical modeling. (3rd edition) Figs 3.5,3.6, page 102

Visual model fitting

- Transforming the data
  - Much easier to fit visually to linear data
  - Problem: How about if the data is not linear?
  - Solution: transform the data!
- Examples: $y = \ln x \to Y = e^y = cx$
  - Fit a model y=Ceˣ is the same as fitting a model ln(y)=ln(C)+x exp.
    - Replace the (x,y) data with the (x,ln(y)) data (log-transform) and do a linear fit
  - Fit a model y=Cxᵃ is the same as fitting a model ln(y)=ln(C)+a ln(x) power
    - Replace the (x,y) data with the (ln(x),ln(y)) data (log-log-transform) and do a linear fit

---

Data transformations

- When doing data transformations the concept of distance is also transformed
  - Fitting a transformed model to minimize the deviations may not yield a final model with minimum deviation to the original data
  - Always verify the final model against the original data!

Data transformations

$\min \sum_i | y_i - f(x_i) |$ "the unique criteria"

$\neq \min \sum | \ln y_i - (\ln C + \frac{1}{x_i}) |$

- Example
  - Fit a model y=Ce1/x
  - Log-transform of the data: ln(y)=ln(C)+1/x
    - Data gets squeezed together
    - Absolute deviations appear small
    - Not all good fits on the log-scale will be good fits on the original scale

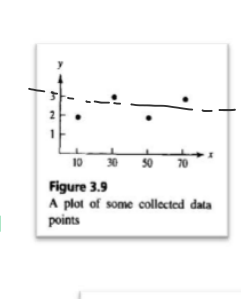


Figure 3.9 A plot of some collected data points

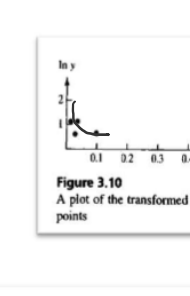


Figure 3.10 A plot of the transformed data points

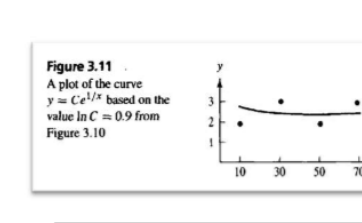


Figure 3.11 A plot of the curve $y = Ce^{1/x}$ based on the value $\ln C \approx 0.9$ from Figure 3.10

Giordano et al. A first course in mathematical modeling. (3rd edition) Figs 3.9-3.11, page 104-105

Data transformations

- Data may get squeezed
  - Modeler may be tricked in selecting a poor model based on the transformed data
    - Very important to remember when comparing alternative models
  - Always compare using the original data
  - Note: computer-based environments may use (hidden) implicit data transformations
    - Check how the indicators of model fit are computed

ANALYTIC MODEL FITTING

Analytic model fitting

- Methods for judging the fitness of a model
  - Chebyshev approximation criterion
  - Sum of absolute deviations
  - Least-squares criterion
  - …

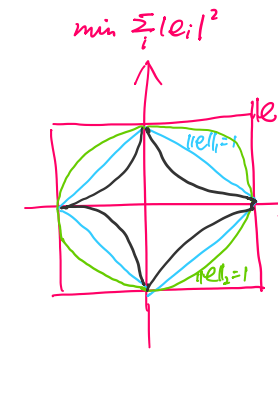
$\min \max_i |e_i|$ $\min \|\vec{e}\|_\infty$

$\min \sum_i |e_i|$ $\min \|\vec{e}\|_1$

$\min \sum_i |e_i|^2$ $\min \|\vec{e}\|_2$

$\vec{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$

$n = 2$



First criterion for goodness of a fit: Chebyshev approximation

- Chebyshev criterion:
  - Given a data set (xᵢ,yᵢ), 1≤i≤m, and a model y=f(k,x), with k the vector of parameters to be fit, select those parameter values which minimize

    max {|yᵢ-f(xᵢ)|, 1≤i≤m}

  - In other words: minimize the largest absolute deviation

HW4. $f(x) = ax + b$. $\{x_i, y_i\}_{i=1}^n$

Find a, b. s.t.

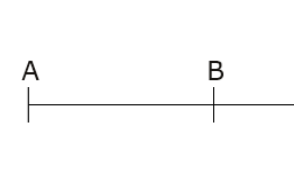
① $\min \|\vec{e}\|_\infty$ a. b.

② $\min \|\vec{e}\|_1$ a. b

③ $\min \|\vec{e}\|_2$ a. b ✓

④ $\min \|\vec{e}\|_0$ a. b

Example



- We are given a segment AC and three measurements: of AC itself, of AB and of BC, where B is a point on the segment AC
  - Example: AC=19, AB=13, BC=7
- Problem: find the values for the length of each segment that give the best fit using the Chebyshev criterion
  - Notation: let x₁, x₂, x₁+x₂ be the lengths of the segments AB, BC, and AC, resp.
  - The deviations are: x₁-13=r₁, x₂-7=r₂, x₁+x₂-19= r₃.
- Problem formulation: find the minimal r such that:
  - |r₁|≤r, |r₂|≤r, |r₃|≤r
  - Equivalently: -r≤ r₁≤ r, -r≤ r₂≤ r, -r≤ r₃≤ r
  - Still equivalently: r-r₁≥0, r+r₁≥0, r-r₂≥0, r+r₂≥0, r-r₃≥0, r+r₃≥0
  - In our example:
    - r-x₁+13≥0
    - r+x₁-13≥0
    - r-x₂+7≥0
    - r+x₂-7≥0
    - r-x₁-x₂+19≥0
    - r+x₁+x₂-19≥0
  - Linear programming!
    - Solved through the simplex method

$x_1 = AB$ $r_1 = x_1 - 13$ ✓

$x_2 = BC$ $r_2 = x_2 - 7$ ✓

$x_1 + x_2 = AC$ $r_3 = x_1 + x_2 - 19$ ✓

$\min \max_{i=1}^{3} |r_i|$

$|r_i| < 1$

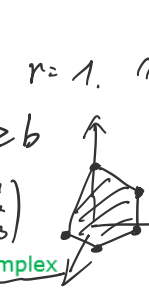
6×3

r = 1.

x_1, x_2, x_3

$\Leftrightarrow A x \geq b$

$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$



linprog(A,b)

Chebyshev approximation

- Note: in general, the resulting optimization problem may not be linear!
  - Example: f(x)=sin kx
  - for this reason, the criterion is not much used in practice

$\min \max_i | y_i - \sin k x_i |$

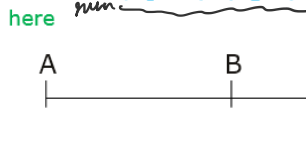
non-smooth + nonlinear

Second criterion for goodness of a fit: sum of absolute deviations

- Criterion:
  - Given a data set (xᵢ,yᵢ), 1≤i≤m, and a model y=f(k,x), with k the vector of parameters to be fit, select those parameter values which minimize

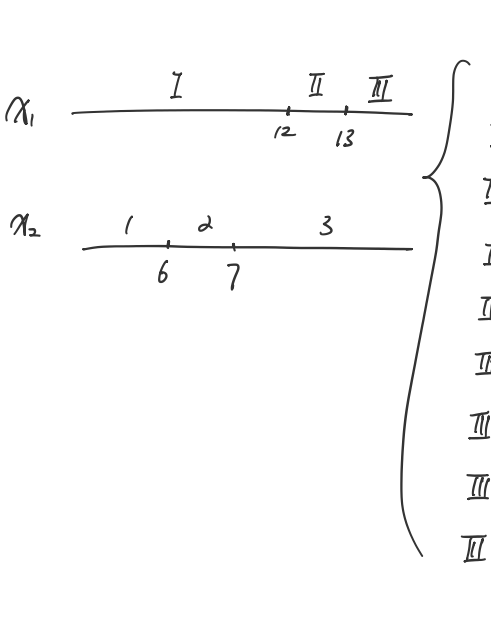
    $\sum_{1\leq i\leq m} |y_i - f(x_i)|$

  - In other words: minimize the sum of absolute deviations
  - General approach: differentiate the sum with respect to each parameter, solve the 0-equations to find the critical points
    - Problem: because of the modules, the derivatives may not be continuous
- Example on the previous slide:
  - Find x₁,x₂≥0 such that |x₁-13|+|x₂-7|+|x₁+x₂-19| is minimal min
    - not solved here



3. Study role of sparse regularization and the LASSO method!

$\min \| \vec{y} - (a\vec{X} + b) \|_1$



I ×1

II ×1

III ×1

I ×2

II ×2

III ×2

III ×1

III ×2

III ×3

Third criterion for goodness of a fit: Sum of least-squares

- Criterion:
  - Given a data set (xᵢ,yᵢ), 1≤i≤m, and a model y=f(k,x), with k the vector of parameters to be fit, select those parameter values which minimize

    $\sum_{1\leq i\leq m} |y_i - f(x_i)|^2$

  - In other words: minimize the sum of squares of absolute deviations

---

Third criterion for goodness of a fit: least-squares

- Widely used criterion because the resulting problem can be easily solved using calculus: if f is mathematically "well-behaved" function (say, analytical), then so is the sum of squares
- Example on the previous slide: quadratic form must admit a unique minimizer!
  - Find x₁,x₂≥0 such that (x₁-13)² +(x₂-7)² +(x₁+x₂-19)² is minimal
  - Solution: differentiate (partial derivative) with respect to the two parameters
    - 2(x₁-13)+2(x₁+x₂-19)=0, 2(x₂-7)+2(x₁+x₂-19)=0,
    - Equivalently, 2x₁+x₂=32, x₁+2x₂=26
    - Solution: x₁=12+2/3, x₂=6+2/3, x₁+x₂=19+1/3

$\frac{\partial}{\partial x_1} = \dots = 0$

$\frac{\partial}{\partial x_2} = \dots = 0$

Example: fitting a straight line

- Fitting a straight line with the least squares criterion
  - look for a model y=Ax+B → $y = Ax^2 + Bx + C$
  - data set: (xᵢ,yᵢ), 1≤i≤m
  - Denote the least-squares solution by y=ax+b:
    - the minimization of $\sum_{1\leq i\leq m} (y_i - ax_i - b)^2$
    - necessary conditions: the partial derivatives of the sum with respect to a and with b are zero:

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

$$\frac{\partial S}{\partial b}(a,b) = -2\sum_{i=1}^{m}(y_i - ax_i - b) = 0$$

$$a = \frac{m\sum_{i=1}^{m} x_i y_i - \sum_{i=1}^{m} x_i \sum_{i=1}^{m} y_i}{m\sum_{i=1}^{m} x_i^2 - \left(\sum_{i=1}^{m} x_i\right)^2}$$

$$b = \frac{\sum_{i=1}^{m} x_i^2 \sum_{i=1}^{m} y_i - \sum_{i=1}^{m} x_i y_i \sum_{i=1}^{m} x_i}{m\sum_{i=1}^{m} x_i^2 - \left(\sum_{i=1}^{m} x_i\right)^2}$$

ompmod/

$\min \sum_{i=1}^{m} (y_i - (Ax_i + B))^2 \sim F$

$\frac{\partial F}{\partial A} = -2\sum_i x_i (y_i - (Ax_i + B)) = 0$

$\frac{\partial F}{\partial B} = -2\sum_i (y_i - (Ax_i + B)) = 0$

$A = \dots$

$B = \dots$

Example: fitting a power curve

$ax^n + bx^{n-1}$ or $ax^n + b$

- Fitting a power curve with the least squares criterion
  - Goal: fit a curve of the form y=axⁿ, where n is fixed, to a given data set
  - Minimize the following sum: $S = \sum_{i=1}^{m} (y_i - f(x_i))^2 = \sum_{i=1}^{m} (y_i - ax_i^n)^2$
  - The only parameter here is a
  - Necessary condition: the derivative dS/da(a)=0:

$$\frac{dS}{da}(a) = -2\sum_{i=1}^{m} x_i^n (y_i - ax_i^n) = 0$$

$$a = \frac{\sum_{i=1}^{m} x_i^n y_i}{\sum_{i=1}^{m} x_i^{2n}}$$

Relating the 3 criteria

- Geometric intuition
  - Chebyshev criterion: minimize the largest absolute deviation l_∞ − norm
    - more weight given to the worst point
  - Minimize the sum of absolute deviations l_1 − norm
    - tends to treat each data point equally and to average the deviations
  - Least-squares l_2 − norm
    - somewhat in-between

Relating the 3 criteria

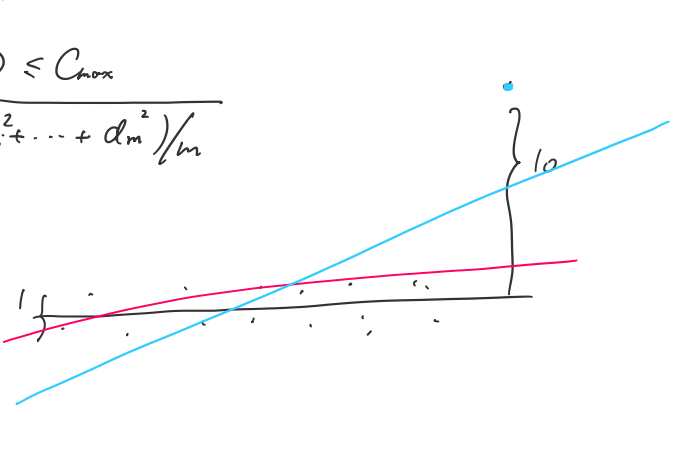
- Analytical relationship between Chebyshev and least-squares
  - Let f₁(x) be the solution given by the Chebyshev criterion.
    - Let cᵢ=|yᵢ-f₁(xᵢ)|
    - Let c_max be the largest cᵢ: the largest absolute deviation
    - f₁ is calculated so that c_max is minimal
  - Let f₂(x) be the solution given by the least-squares criterion
    - Let dᵢ=|yᵢ-f₂(xᵢ)|
    - f₂ is calculated so that d₁²+ d₂²+… dₘ² is minimal
    - Let d_max be the largest dᵢ
  - Optimality of the Chebyshev solution: c_max ≤d_max
  - Optimality of the least-squares solution: d₁²+ d₂²+… dₘ²≤ c₁²+ c₂²+…+ cₘ²≤mc_max²
  - Define D=sqrt((d₁²+ d₂²+… dₘ²)/m) (square root)
  - Conclusion: **D≤c_max≤d_max**
- Note: least-squares is more widely used; if the difference between D and d_max is however considerable, consider using Chebyshev instead

$d_1^2 + \dots + d_m^2 \leq m\, c_{max}^2$

$d_i \leq c_{max}$

$D \leq c_{max}$

$D = \sqrt{(d_1^2 + \dots + d_m^2)/m}$



Choosing a best model

- For the same data set and the same type of model different results may be obtained depending on the type of fit
  - other methods for model fitting also exist, potentially giving different results for the same data set (nondeterministic methods)
- Question: which model to choose as being "best"?
  - if one were allowed to change the type of math model (e.g., look for a cubic, rather than a quadratic polynomial) even better fits may be found
  - The answer should be given depending on the purpose of the model, the required precision, accuracy of the data, etc.
    - a preset sum of squares may be enough to judge a model "good"
  - Careful about applying the numerical criteria blindly
  - Example: for the following 4 data sets, the model y=x yields the same sum of squared deviations

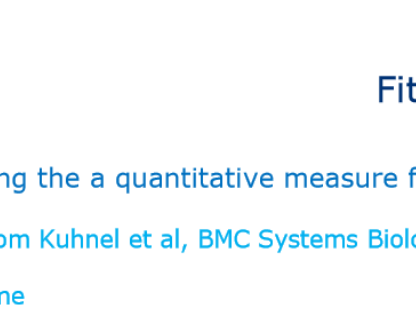


Figure 3.15 In each of these graphs the model $y = x$ has the same sum of squared deviations

$y = (ax)$

$= f(x) =$

"distributed"

Fit quality

- Various methods for defining the a quantitative measure for the quality of a model fit
  - Here present just one, from Kuhnel et al, BMC Systems Biology (2008)
  - Only one data set at a time
  - Gives a measure of the average deviation of the model prediction from the experimental data, normalized by (the average of) the absolute values of the model prediction
  - This measure of fit quality does not discriminate against models aiming to explain experimental data with large absolute values
  - Let exp be the experimental data; m the number of experimental points

$$qual(\exp) = \frac{\sqrt{\dfrac{sum\_of\_squared\_deviations}{m}}}{mean\_of\_predicted\_values} \cdot 100\%$$

  - Rule of thumb (Kuhnel et al): lower than 20% value for qual(exp) can be considered as a good fit

---

Learning objectives

- Understand the concepts of model fitting and data interpolation
- Indicate several possible sources of errors in modeling
- Understand the concept of confidence interval
- Ability to formulate the following 3 fitting criteria for a given model:
  - Chebyshev approximation
  - Sum of absolute deviation
  - Least squares