Primer on Analysis of Experimental Data and Design of Experiments

Lecture 15. Conclusions and Outlook

Muhammad A. Alam
alam@purdue.edu

The topics covered …

Lectures 8-14

How to get a better f

Lectures 6-7

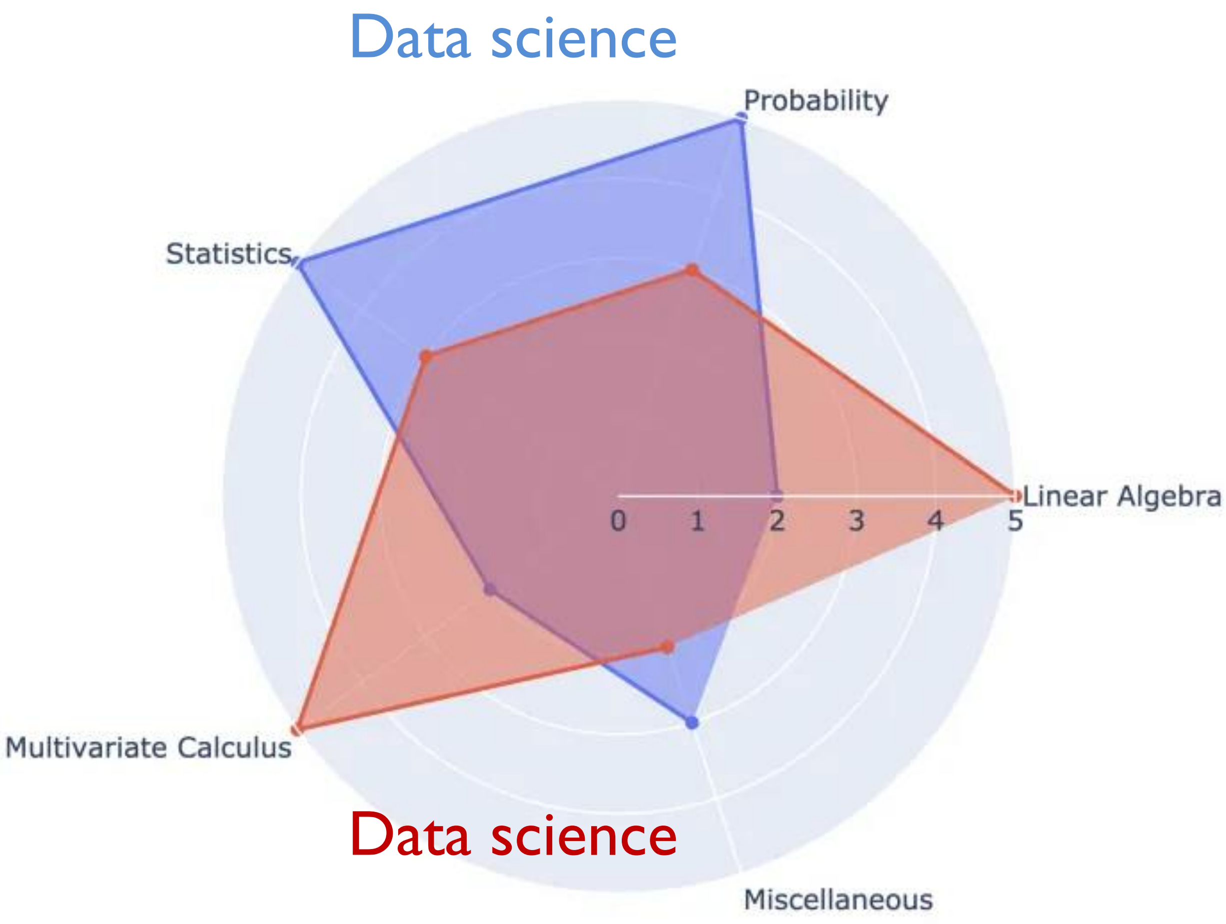
$\bar{x} = x_1, x_2, \ldots x_n$ $\longrightarrow$ $\bar{y} = f(\bar{x})$ $\longrightarrow$ $\bar{y} = y_1, y_2, \ldots y_m$

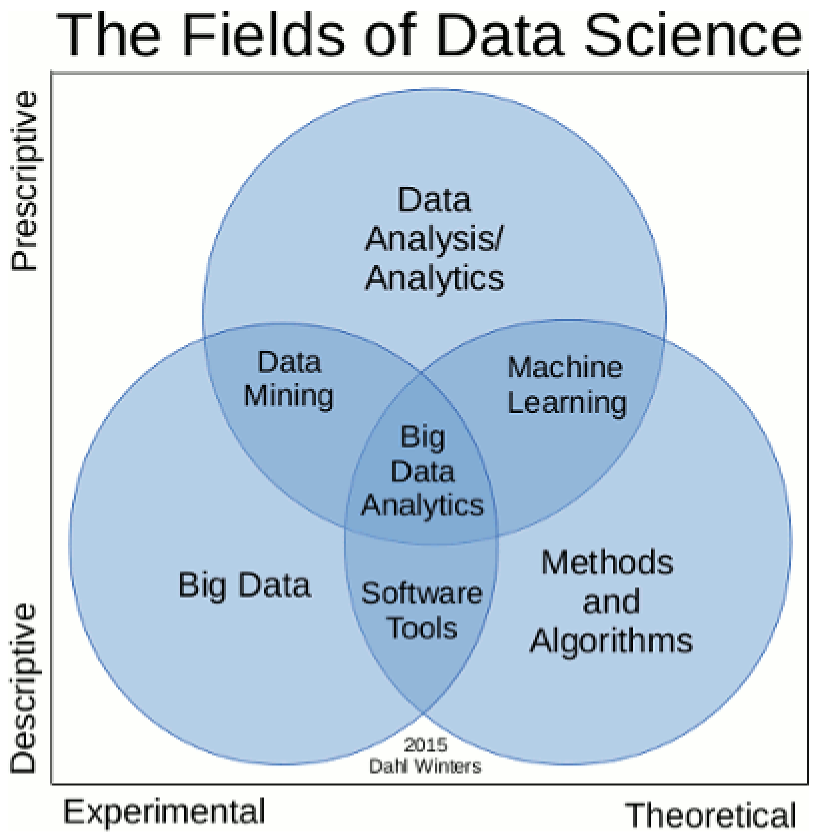
Lectures 1-2

Lectures 3-5

How to fit multiple hypothetical function f to the same y

Data science and Machine Learning



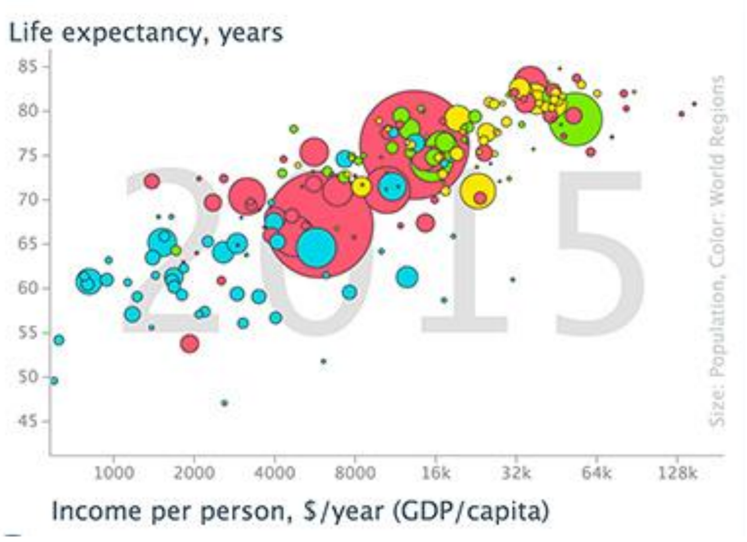


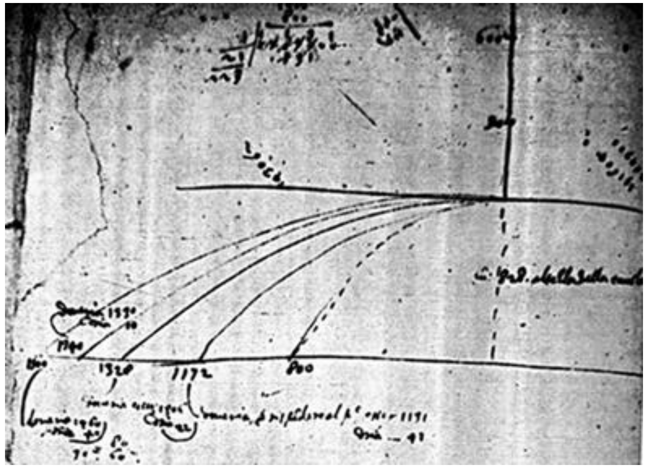
Course Outline

$$\overline{y} = f\,(\overline{x}\,) \qquad \overline{x} = x_1, x_2, \ldots x_n \qquad \overline{y} = y_1, y_2, \ldots y_m$$

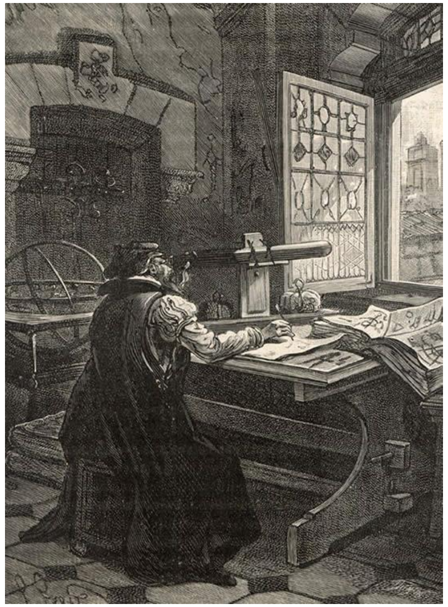
Lecture 1: Introduction
Lecture 2: Collecting and plotting $x_1, x_2, \ldots x_n$
Lecture 3: Physical and empirical $f, F, df/dx, \ldots$
Lecture 4: Model selection between $f_1, f_2, \ldots$
Lecture 5: Model Selection: Cross-validation and Bootstrapping method
Lecture 6: Scaling theory with known f, $f(\overline{x}) = f(\overline{X})$
Lecture 7: Scaling theory with unknown f, $\overline{x} \rightarrow X$
Lecture 8: Design of experiments to determine $\overline{y}_{\mathrm{max}} = f\,(\overline{x}\,)$
Lecture 9: DOE and ANOVA
Lecture 10: Principle component analysis for classifying $\{y\}$.
Lecture 11: Machine learning … Statistical approach to learn f
Lecture 12: Machine Learning …. Deep network, Karnaugh map, and other components
Lecture 13: Interpretable ML: Physics-based machine learning $f = f_{\mathrm{physics}} + \Delta f$
Lecture 14: Conclusions

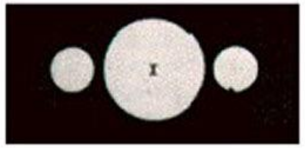
Lecture 1: A short history of data



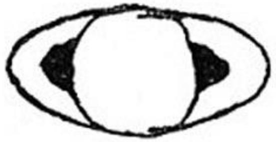


Small vs. big data

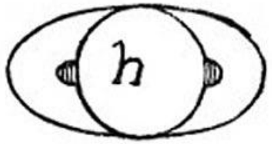




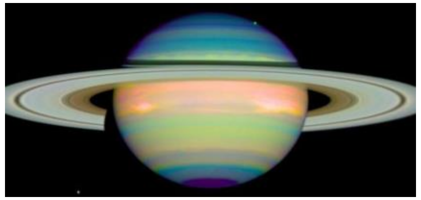
Galileo first sketch
1610



Better telescope
1616



Published etch
1623



Course Outline

$$\bar{y} = f(\bar{x}) \quad \bar{x} = x_1, x_2, \ldots x_n \quad \bar{y} = y_1, y_2, \ldots y_m$$

Lecture 1: Introduction

Lecture 2: Collecting and plotting $x_1, x_2, \ldots x_n$

Lecture 3: Physical and empirical $f, F, df/dx, \ldots$

Lecture 4: Model selection between $f_1, f_2, \ldots$

Lecture 5: Model Selection: Cross-validation and Bootstrapping method

Lecture 6: Scaling theory with known f, $f(\bar{x}) = f(\overline{X})$

Lecture 7: Scaling theory with unknown f, $\bar{x} \to X$

Lecture 8: Design of experiments to determine $\bar{y}_{\max} = f(\bar{x})$

Lecture 9: DOE and ANOVA

Lecture 10: Principle component analysis for classifying $\{y\}$.

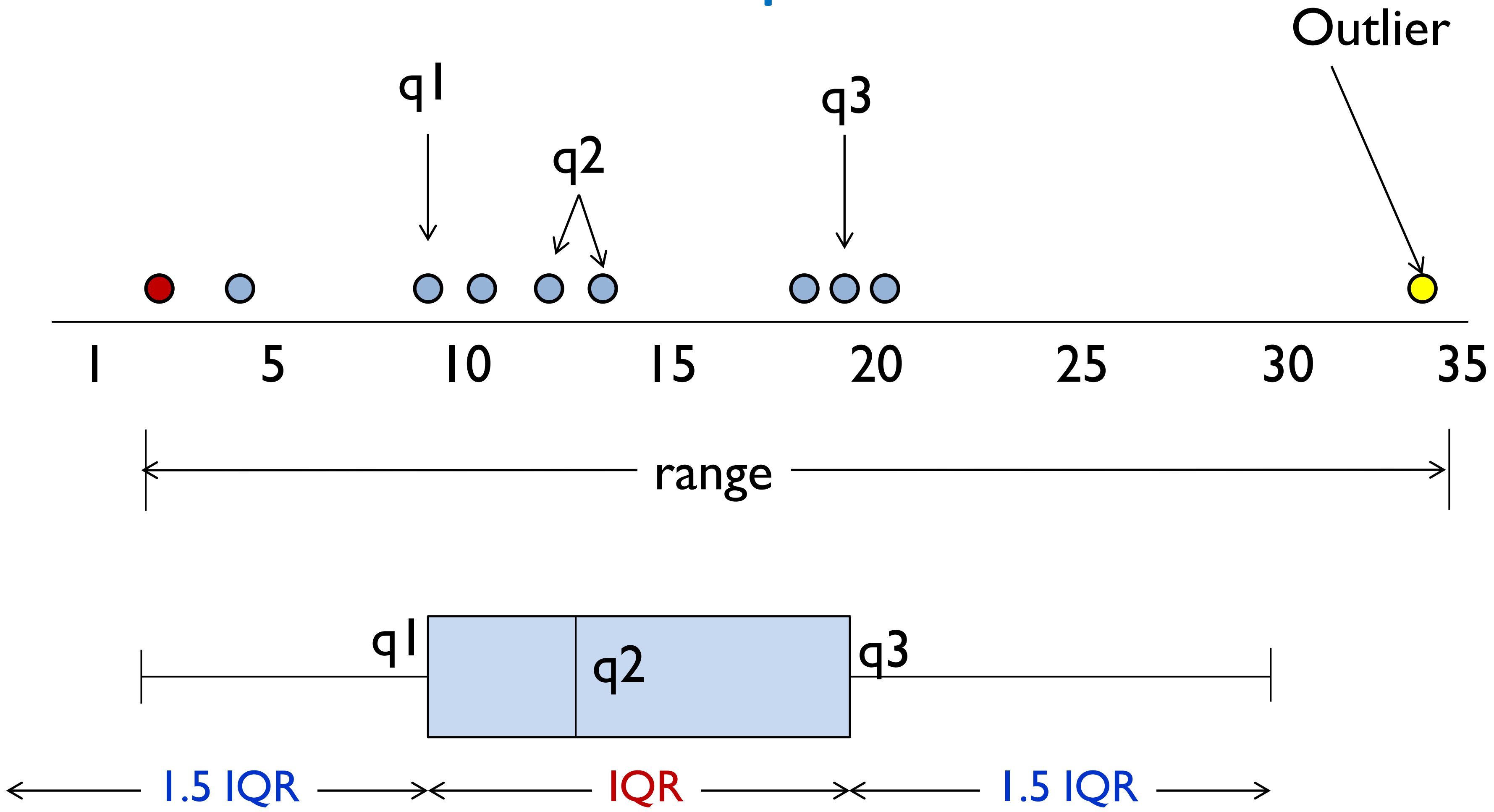
Lecture 11: Machine learning … Statistical approach to learn f

Lecture 12: Machine Learning …. Deep network, Karnaugh map, and other components

Lecture 13: Interpretable ML: Physics-based machine learning $f = f_{\text{physics}} + \Delta f$

Lecture 14: Conclusions

Box plot


Outlier
q1
q2
q3
1
5
10
15
20
25
30
35
range
q1
q2
q3
1.5 IQR
IQR
1.5 IQR

Stem and leaf display: Pre-histogram

Order data

44 46 47 49 63 64 66 68 68 72 72 75 76 81 84 88 106

n=17

4 | 4679 ← Leaf
5 |
6 | 34688
7 | 2256
8 | 148
9 |
10 | 6
↑
stem

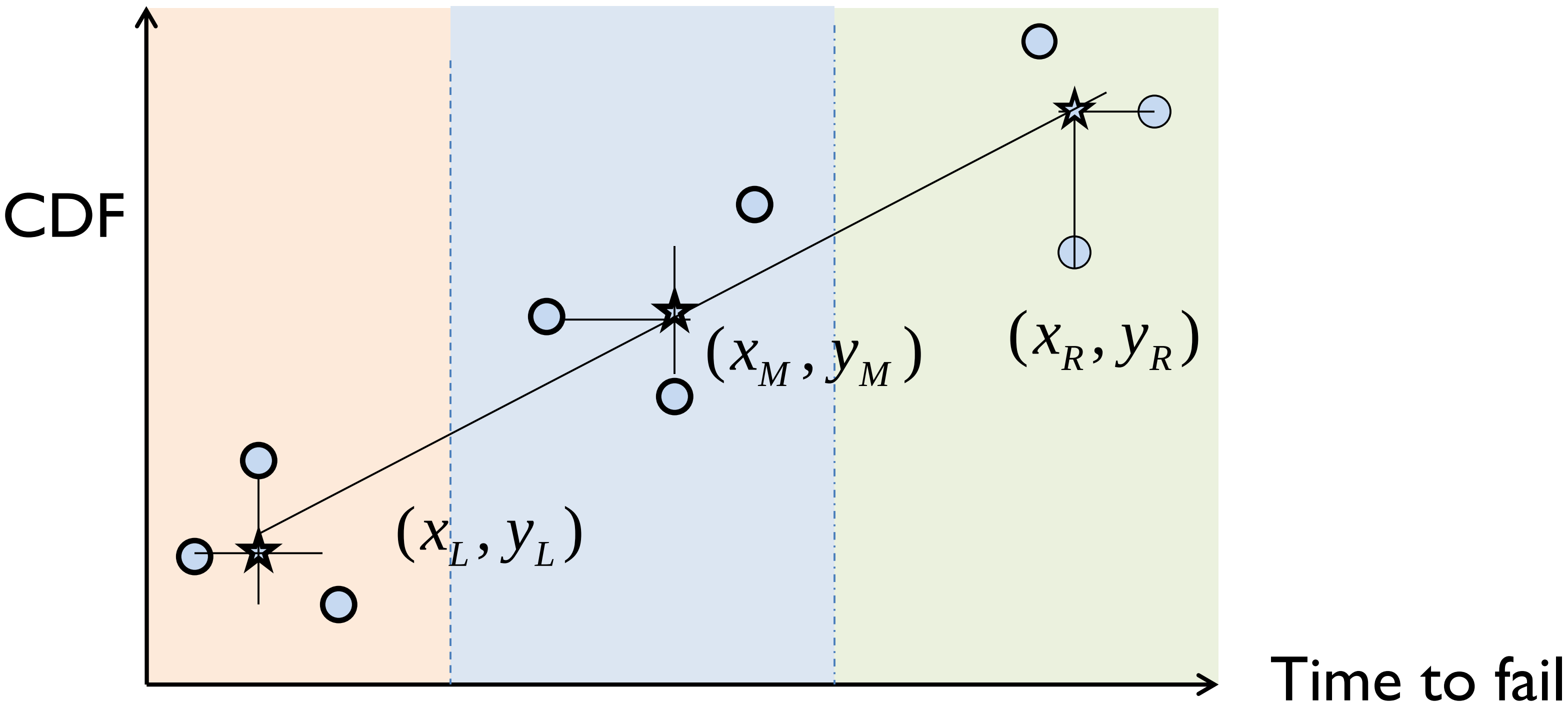
$L = [10 \times \log_{10} n] \sim 13$

h_n = (Range/L) to power of 10 (i.e. 4.77 → 10)

Therefore, 40, 50, 60 ...90, 100 are stem values

Should use the same approach for histogram
Histogram should not increase precision

Drawing lines resistant to outliers



$$y = b(x - x_M) + a$$

$$b_0 = (y_R - y_L)/(x_R - x_L)$$

$$3a_0 = [y_L - b_0(x_L - x_M)] + y_M + [y_R - b_0(x_R - x_M)]$$

$$r_i = y_i - [a_0 + b_0(x_i - x_0)]$$

$$a_1 = a_0 + \gamma_1 \quad b_1 = b_0 + \delta_1$$

For censored data

Assume that at time t_3, one sample is taken out of the experiments (censored)

n_{si} before t_i	5	4	3	2	1
n_{si} after t_i	4	3	2	1	0

$$F_1 = 1 - \frac{4+1}{4+2} = \frac{1}{6}$$

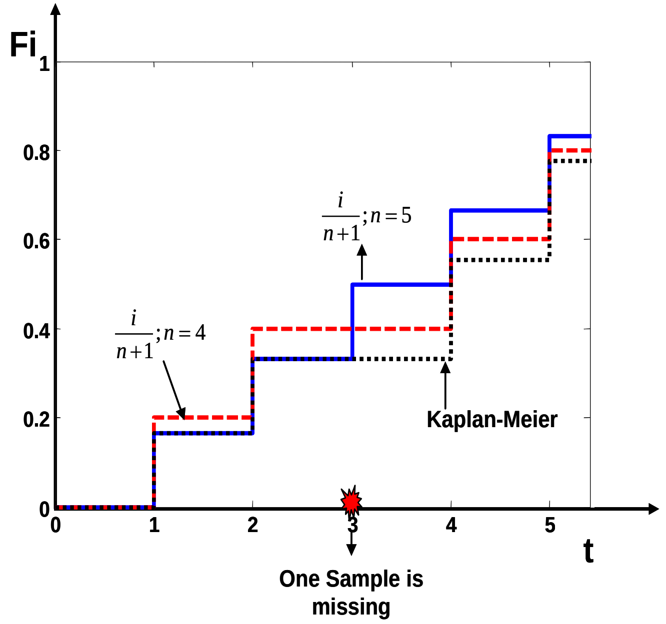
$$F_2 = 1 - \frac{4+1}{4+2} \bullet \frac{3+1}{3+2} = \frac{2}{6}$$

$$\cancel{F_3 = 1 - \frac{4+1}{4+2} \bullet \frac{3+1}{3+2} = \frac{2}{6}}$$

$$F_4 = 1 - \frac{4+1}{4+2} \bullet \frac{3+1}{3+2} \bullet \frac{1+1}{1+2} = 1 - \frac{5}{6} \bullet \frac{4}{5} \bullet \frac{2}{3} = \frac{5}{9}$$

$$F_5 = 1 - \frac{5}{6} \bullet \frac{4}{5} \bullet \frac{2}{3} \bullet \frac{1}{2} = \frac{7}{9}$$

¾ missing …



Lecture 2: Quiz

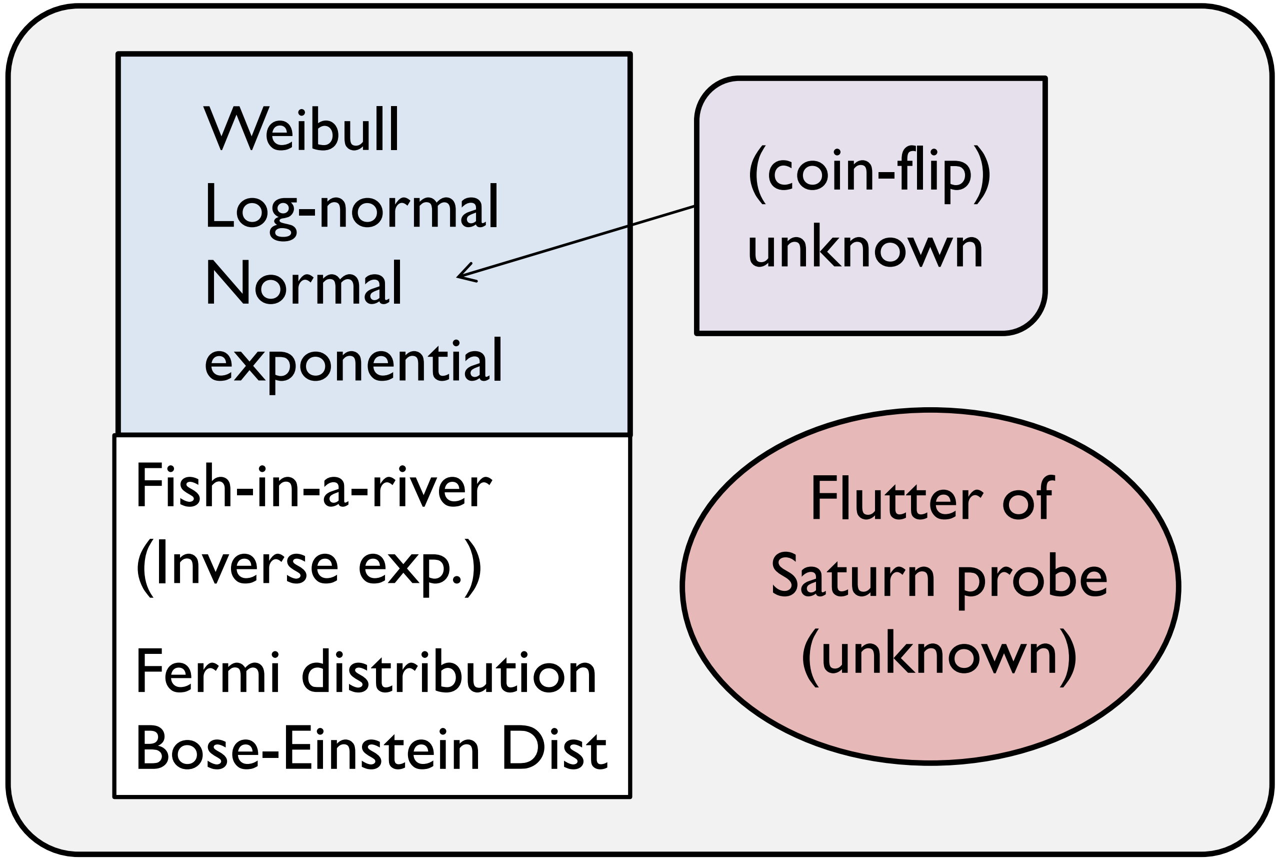
- Z parameter is used to infer the population metrics based on sample metrics for a normal distribution with certain confidence level.
- Bootstrap method is used to calculate distribution of sample metrics (average, standard deviation, etc.) when the population distribution is unknown.
- Any data that exceeds (median + 1.5 IQR) should be thrown out as an outlier.
- The stem-and-leaf intervals are calculated by including the outlier
- Kaplan-Meier formula is a generalization of the Hazen formula

Course Outline

$$\overline{y} = f(\overline{x}) \qquad \overline{x} = x_1, x_2, \dots x_n \qquad \overline{y} = y_1, y_2, \dots y_m$$

Lecture 1: Introduction
Lecture 2: Collecting and plotting $x_1, x_2, \dots x_n$
Lecture 3: Physical and empirical $f, F, df/dx, \dots$
Lecture 4: Model selection between $f_1, f_2, \dots$
Lecture 5: Model Selection: Cross-validation and Bootstrapping method
Lecture 6: Scaling theory with known f, $f(\overline{x}) = f(\overline{X})$
Lecture 7: Scaling theory with unknown f, $\overline{x} \rightarrow X$
Lecture 8: Design of experiments to determine $\overline{y}_{\max} = f(\overline{x})$
Lecture 9: DOE and ANOVA
Lecture 10: Principle component analysis for classifying $\{y\}$.
Lecture 11: Machine learning … Statistical approach to learn f
Lecture 12: Machine Learning …. Deep network, Karnaugh map, and other components
Lecture 13: Interpretable ML: Physics-based machine learning $f = f_{\text{physics}} + \Delta f$
Lecture 14: Conclusions

Statistical Distribution is Physical



If data belong to a well known family, large number of results are available.

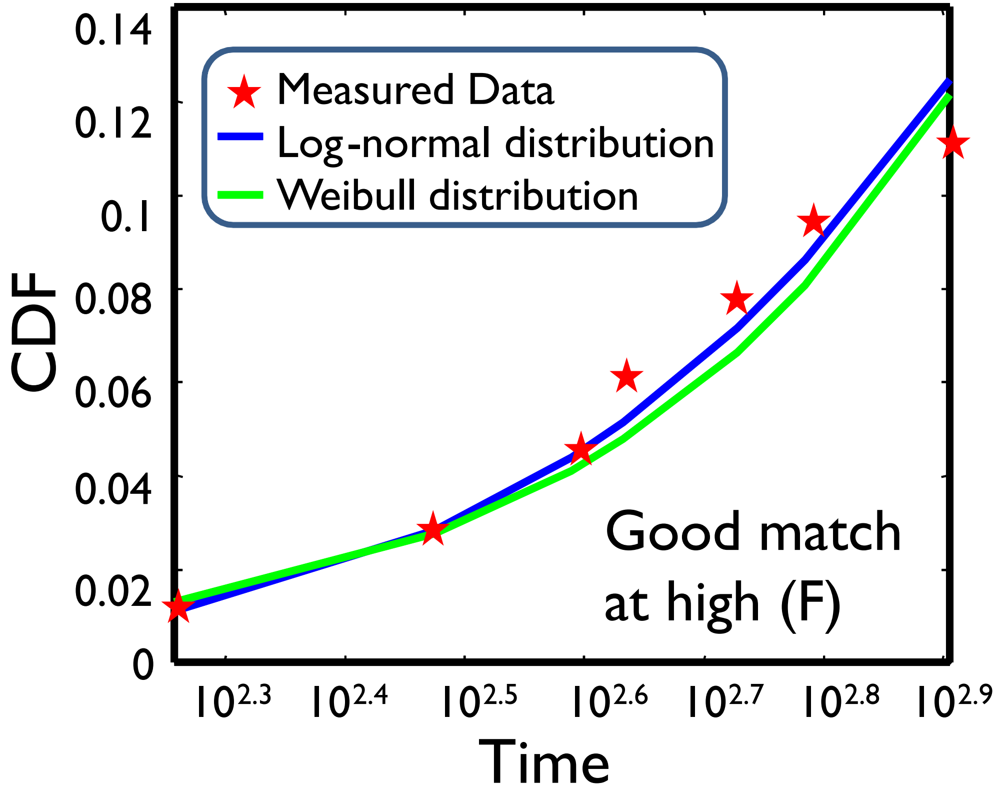
Lecture 3: Quiz

- All physical distributions must belong to Normal, Weibull, or log-normal distributions
- Weibull distribution is a three-parameter distribution
- Given the survival function R(t), we can always calculate the cumulative probability distribution, F(t)
- Moment-based fitting can uniquely identify the probability distribution the data came from.
- The BFRW is an example of a distribution which has finite first and second moments

Matching moments to distributions

Of 60 oxides, 7 failed in 1000 hrs

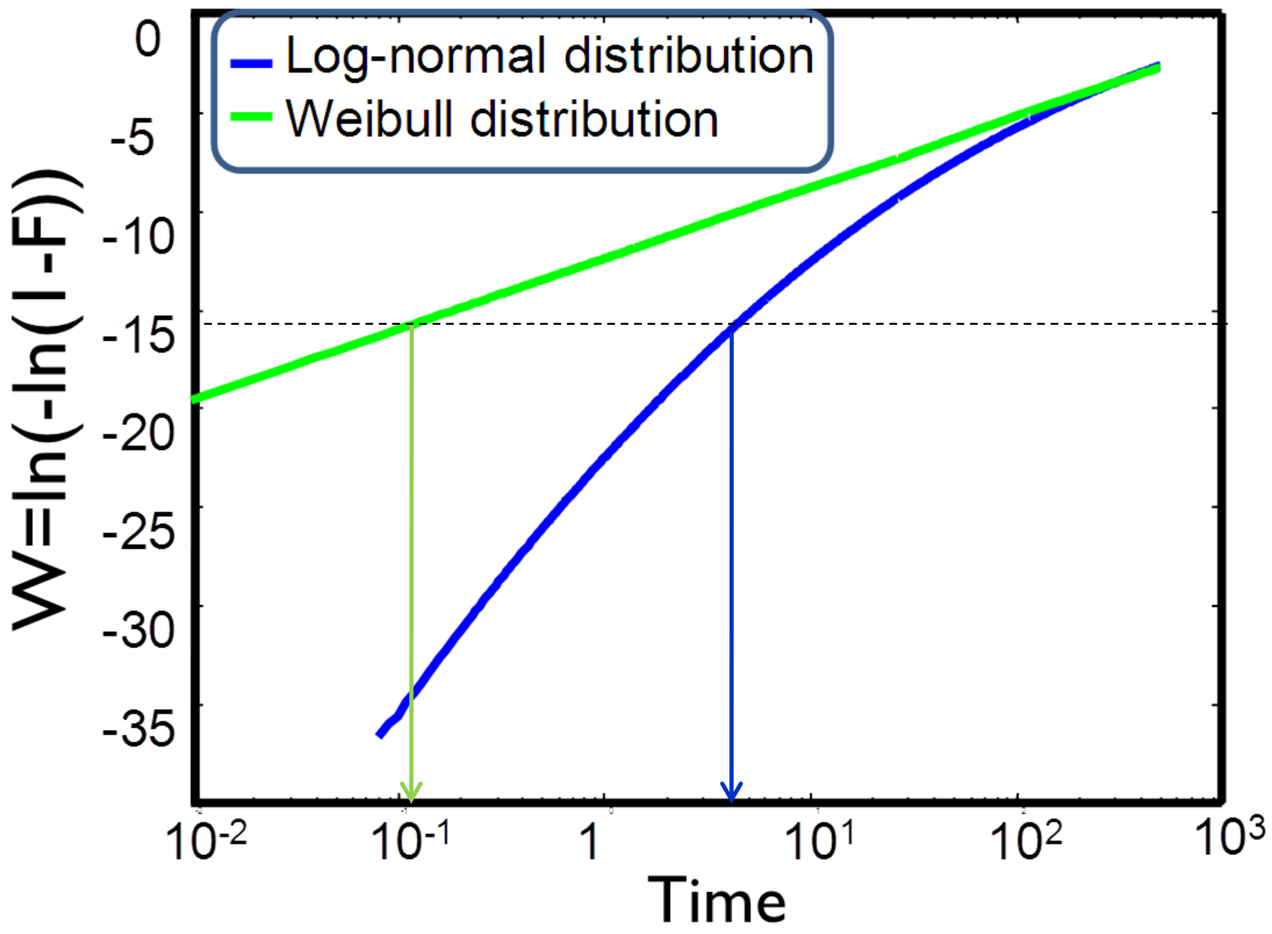
Rank	Lifetime	$F_i = (i - 0.3)/(n + 0.4)$
1	181	0.012
2	299	0.028
3	389	0.045
4	430	0.061
5	535	0.078
6	610	0.094
7	805	0.111



Weibull Distribution Parameters
When $t=\alpha$, $\ln(1-F(t))=-1$, $F(t)=0.632$, a=2990
β estimated using parameter fitting as 1.56

Log-Normal Distribution Parameters
$s=\ln(T_{50\%}/T_{15.9\%})$, $\sigma=\ln(3600/980)=1.30$
$\mu=\ln(T_{50\%})=\ln(3600)=8.19$

Problem of matching the moments



Log-normal distribution is considerably optimistic

Fisher's Maximum Likelihood Method

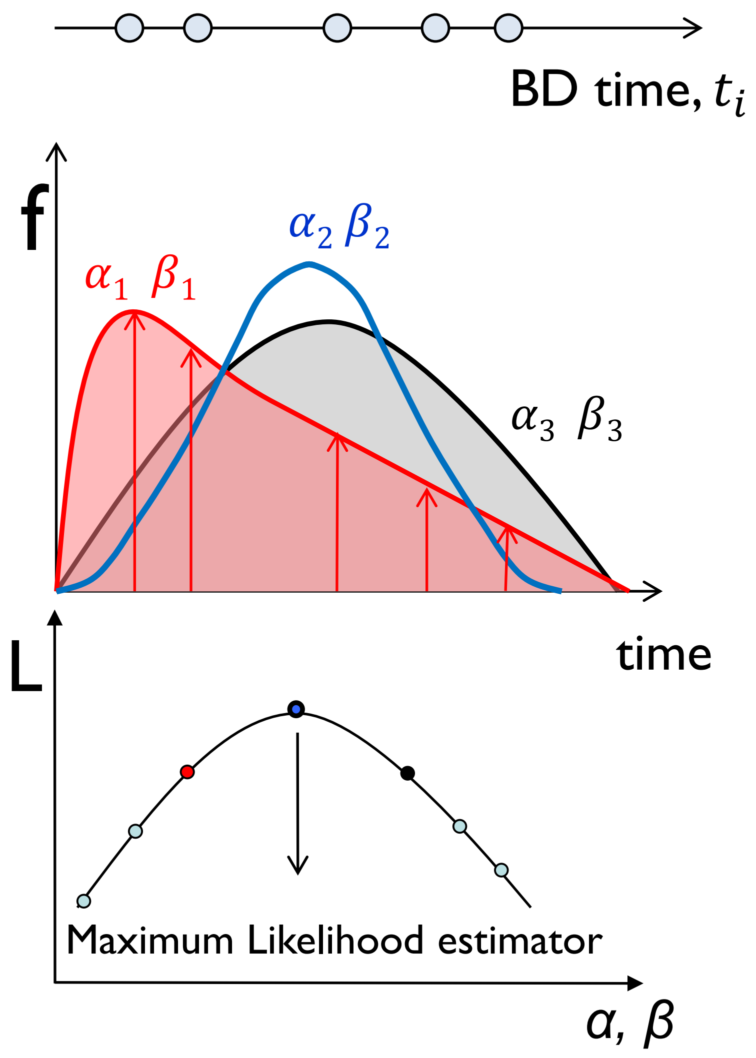
Towering figure, showed that Mendel manipulated data; MALAB fitdist functions

$$f(t_i, \alpha, \beta)$$

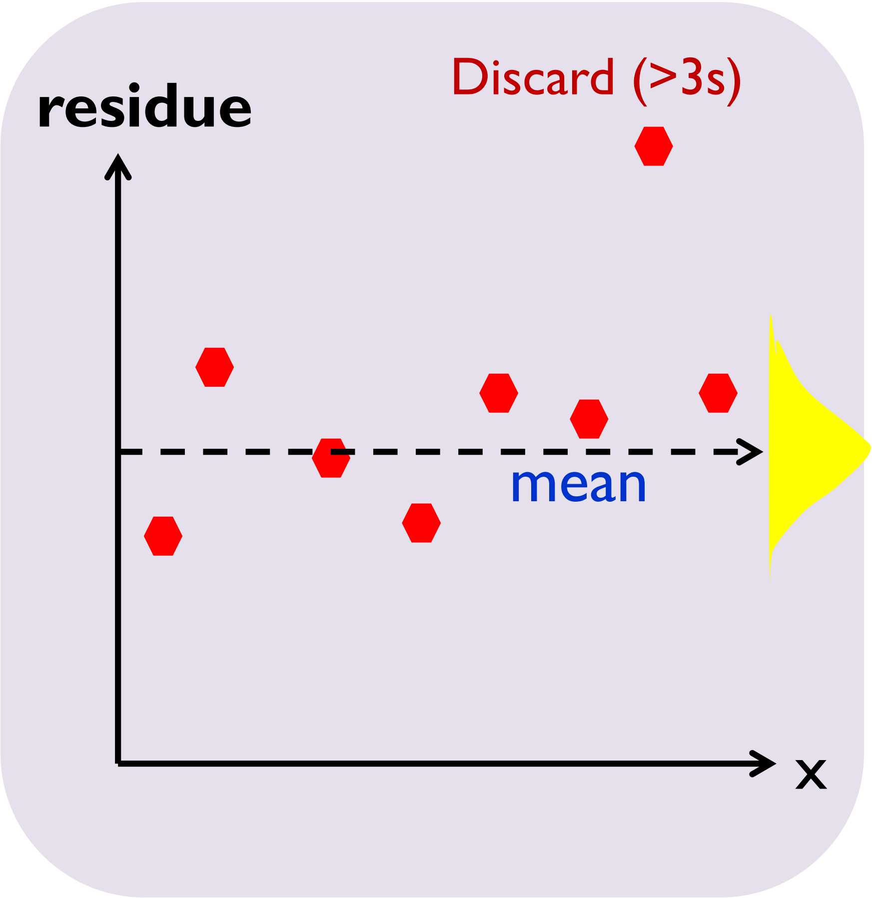
$$L = \prod_{i=1}^{n} f(t_i, \alpha, \beta)$$

$$\ln L = \sum_{i=1}^{n} \ln f(t_i, \alpha, \beta)$$

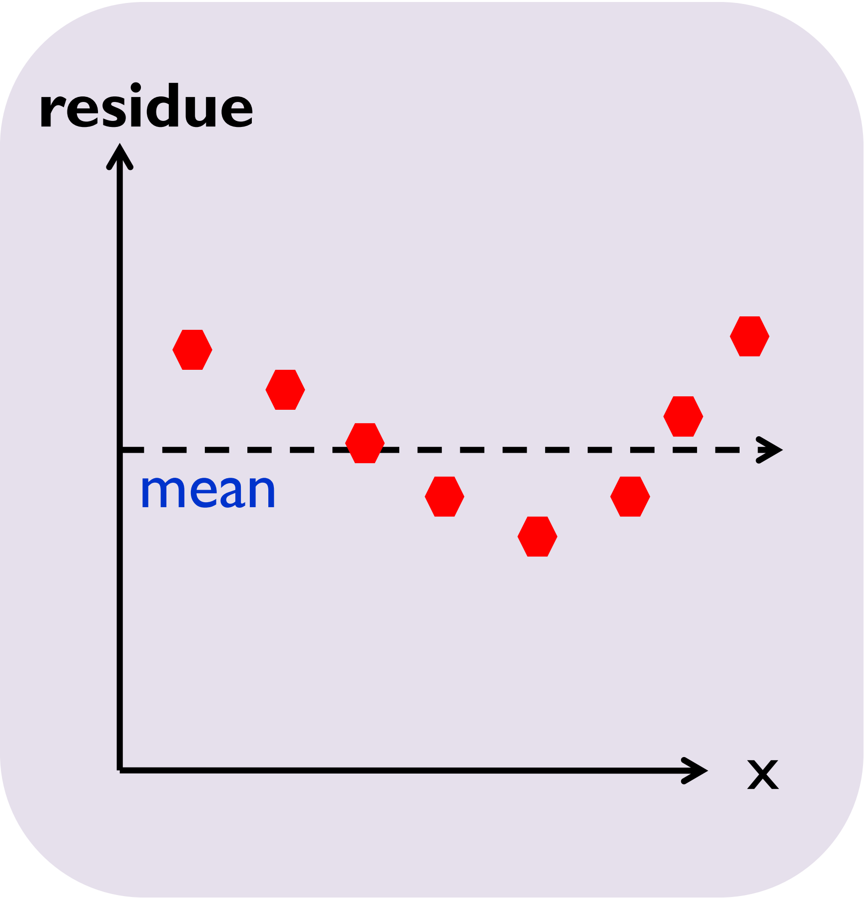
$$\frac{d \ln L}{d\alpha} = 0 \qquad \frac{d \ln L}{d\beta} = 0$$



Goodness of Fit: Residual method



A good fit (normal distribution of residue))



A bad fit (systematic distribution in residual)

Kolmogorov-Smirnov algorithm

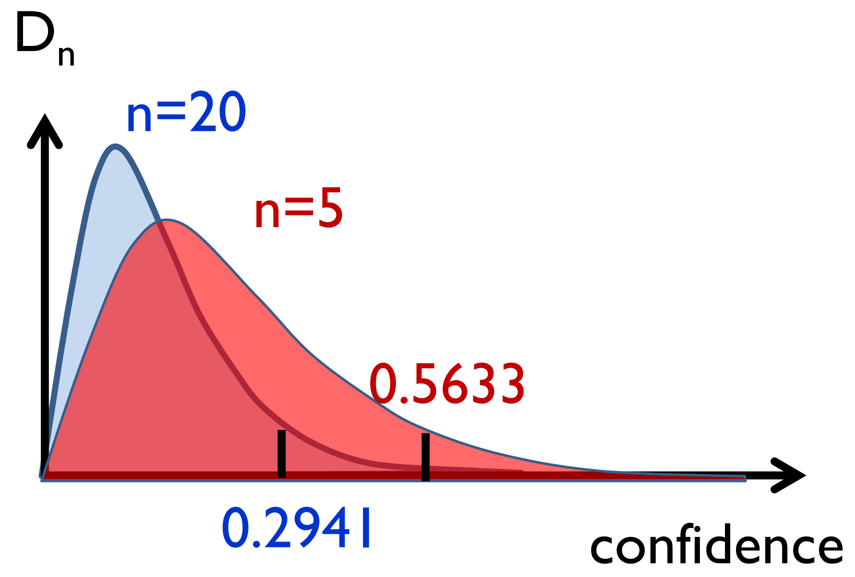
Compute … $D_n = \max \left| F_{obs}(t_i) - F_{theory}(t_i) \right|$

Sample size

If $D_n > D_n^{crit}$, fit is poor …

5% significance level

n	$D_{crit}(n)$
5	0.5633
10	0.4092
20	0.2941
50	0.1884



A famous example: Schon story

Degree of freedom=8-2-1=5

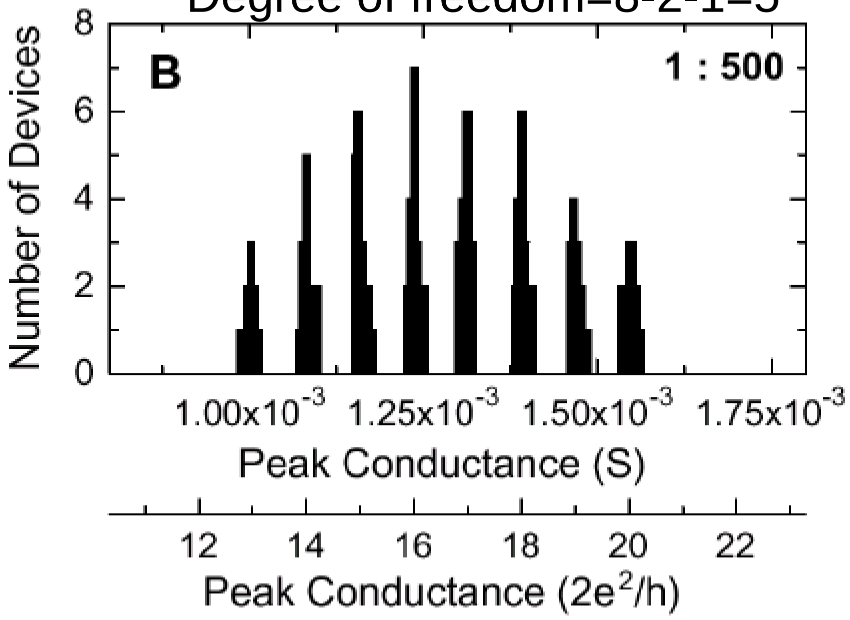


Figure 46. Figure 3(B) from "SingleMolecule" Paper (XIII), showing a histogram of conductances from diluted SAMFETs,

χ^2=0.125! Real data χ^2=1

Chance of this happening is 1.2%

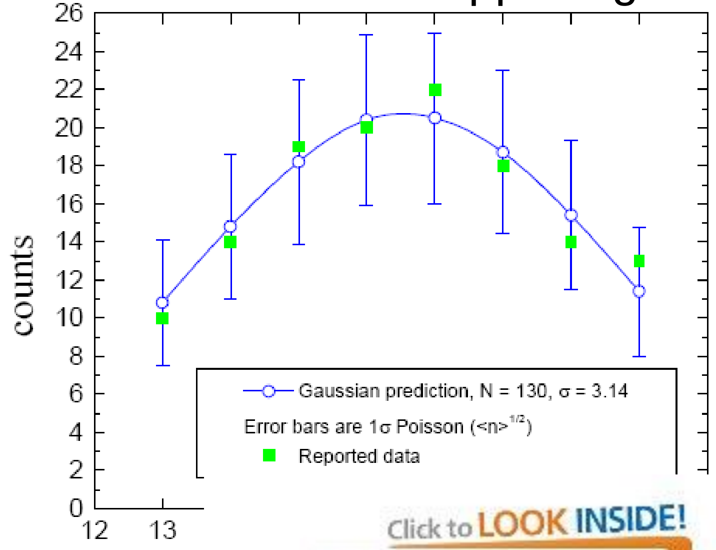
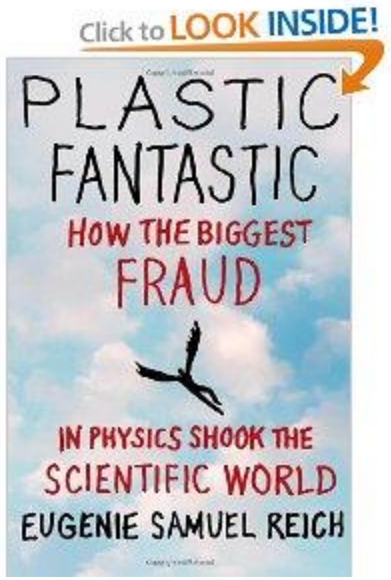


Figure 47.
devices in ea

The data indicating conductance quantization did not arise from an objec
measurement process. At a minimum, the assignment of conductance values was colo
by the expected shape of the final distribution. Such a biased process cannot prov
convincing evidence for quantization. The response to this concern appears to
deliberately deceptive, suggesting that this misrepresentation was intentional.

The preponderance of evidence indicates that Hendrik Schön committed scient
misconduct, specifically data fabrication, in this case.



Lecture 4: Quiz

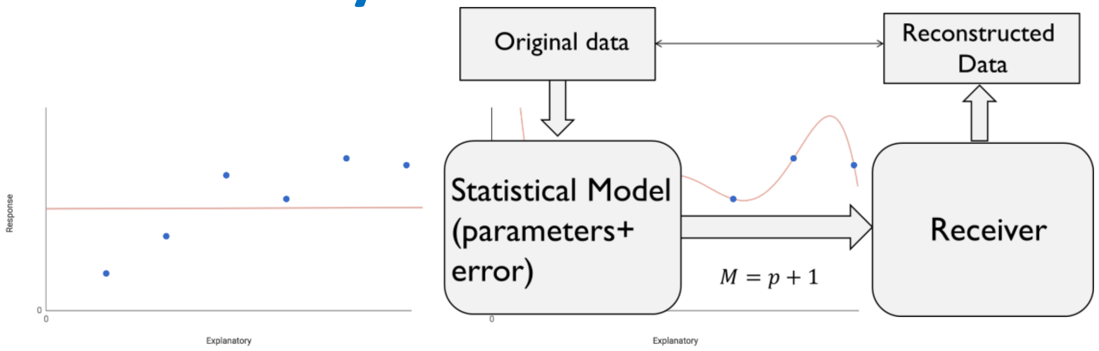
- Least square fitting assumes that the errors (the difference between the data and fitting function) have a Gaussian distribution
- MLE is a general method of data-fitting which is superior to Least-Square technique
- MLE parameter estimate for a log-normal distribution does NOT require iterative solution
- KS test checks on the sum of errors between the data and the fitting function
- Schon's fraud was discovered by the Q-Q method

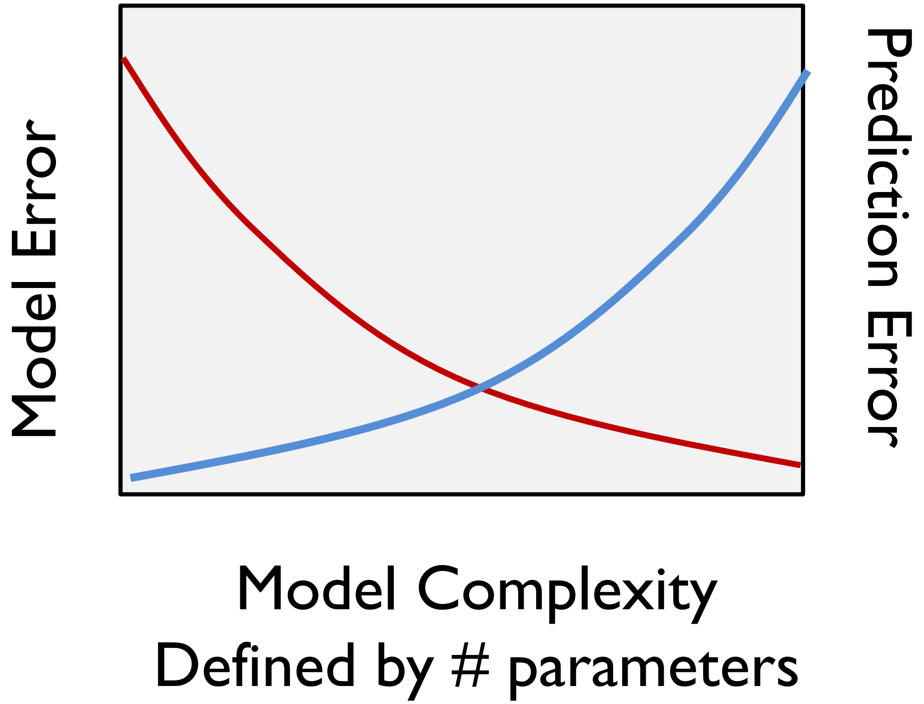
Principle of Parsimony

Aristotle: Nature operates in the shortest way possible.

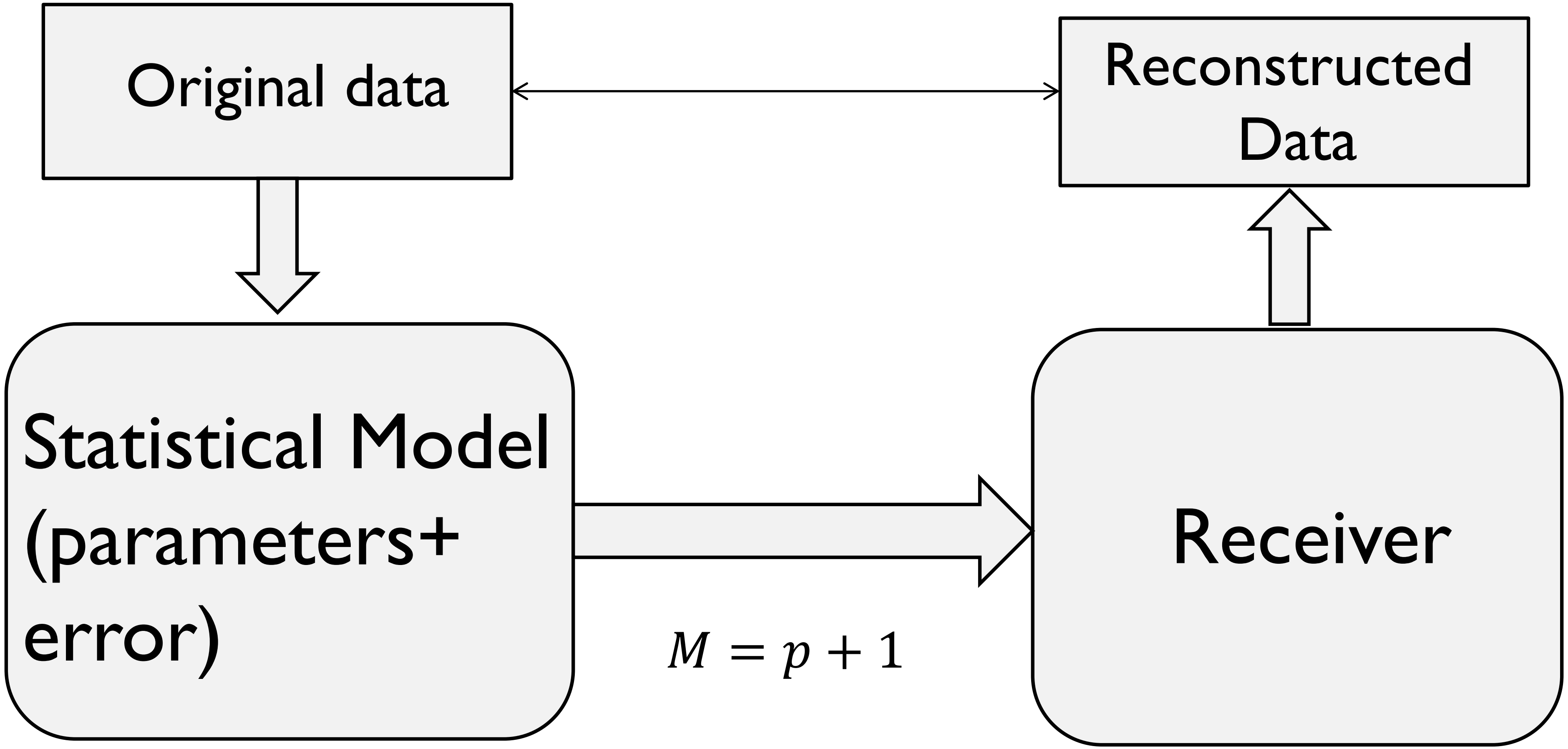
George Box: All models are wrong, but some are useful.

Occam's Razor: "given two or more equally acceptable explanations for a phenomenon, work with the one which introduces the fewest assumptions."





Essence of the information theoretic approach



Parameter number vs. goodness of fit

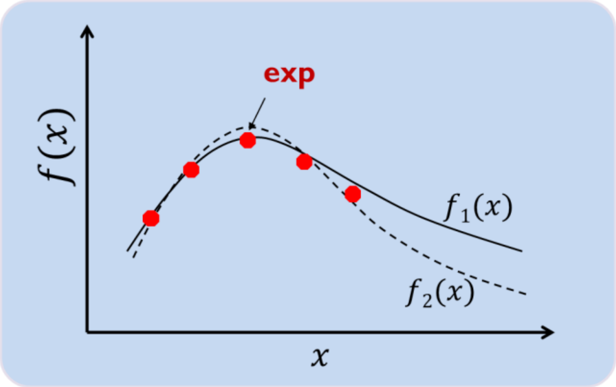
n = number of samples, M=number of parameters

$$R \equiv \sum_{n=1}^{n} \left(t_i - t_{i,fit}\right)^2$$

1) Method of adjusted residual …

$$R_{adj}^2 = \frac{(n-1)R^2 - (M-1)}{n-M}$$

$$M \rightarrow p+1$$



2) Akaike Information Criterion

$$AIC = n \times \ln\left(R^2/n\right) + 2M$$

2) Schwarz Information Criterion

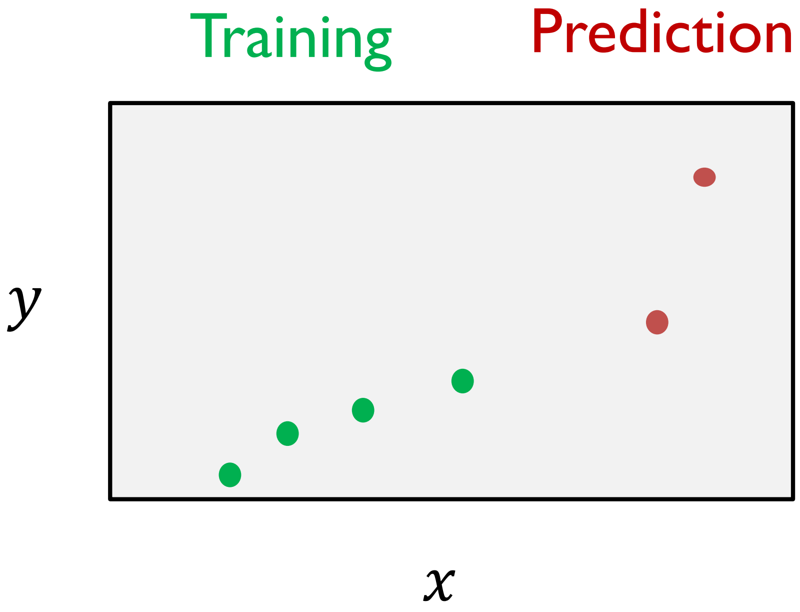
$$BIC = n \times \ln\left(R^2/n\right) + M \times \ln n$$

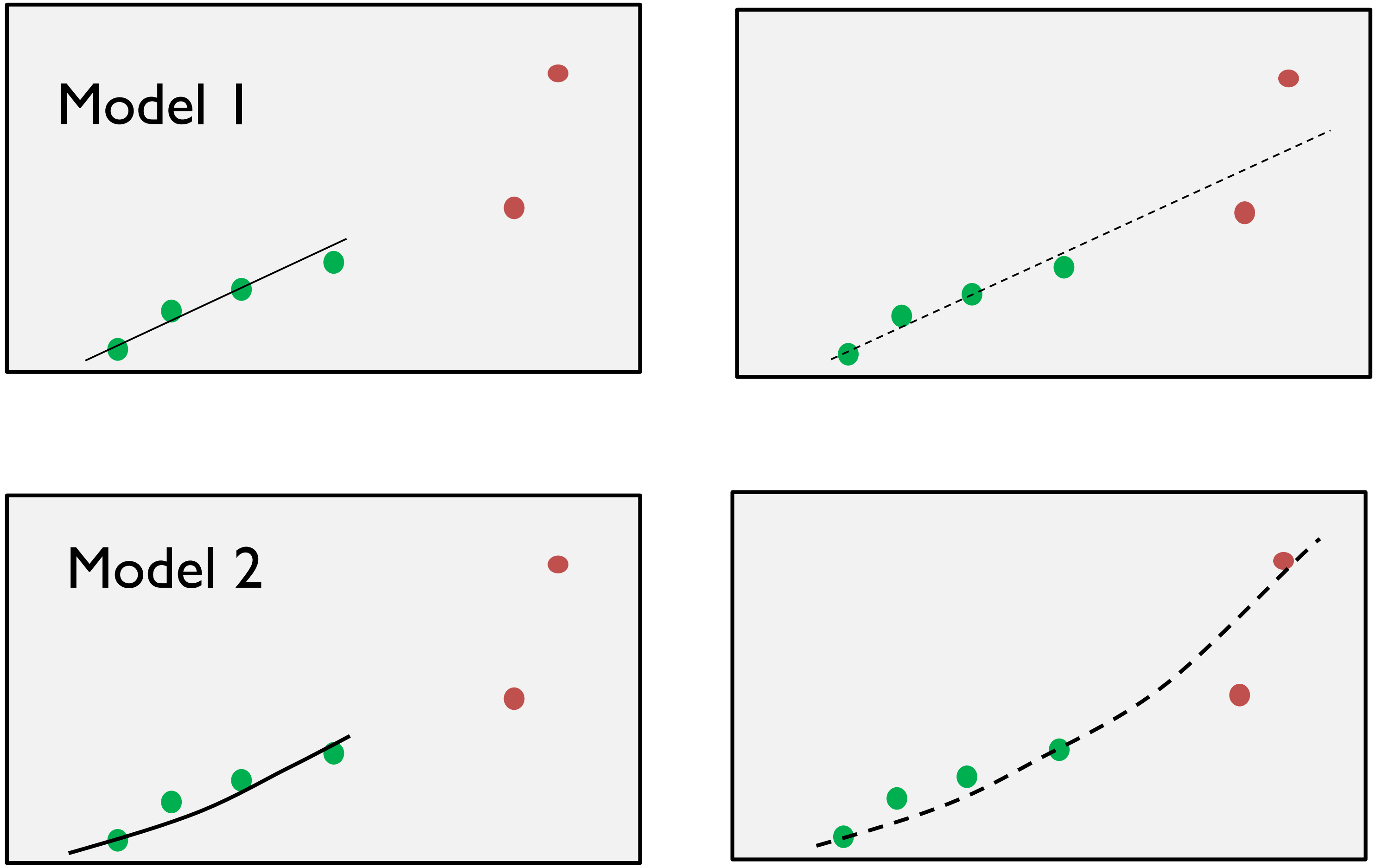
Error penalty

Parameter Penalty

Ref. Les Kirkup, Data Analysis with Excel, Cambridge Univ. Press. P. 304

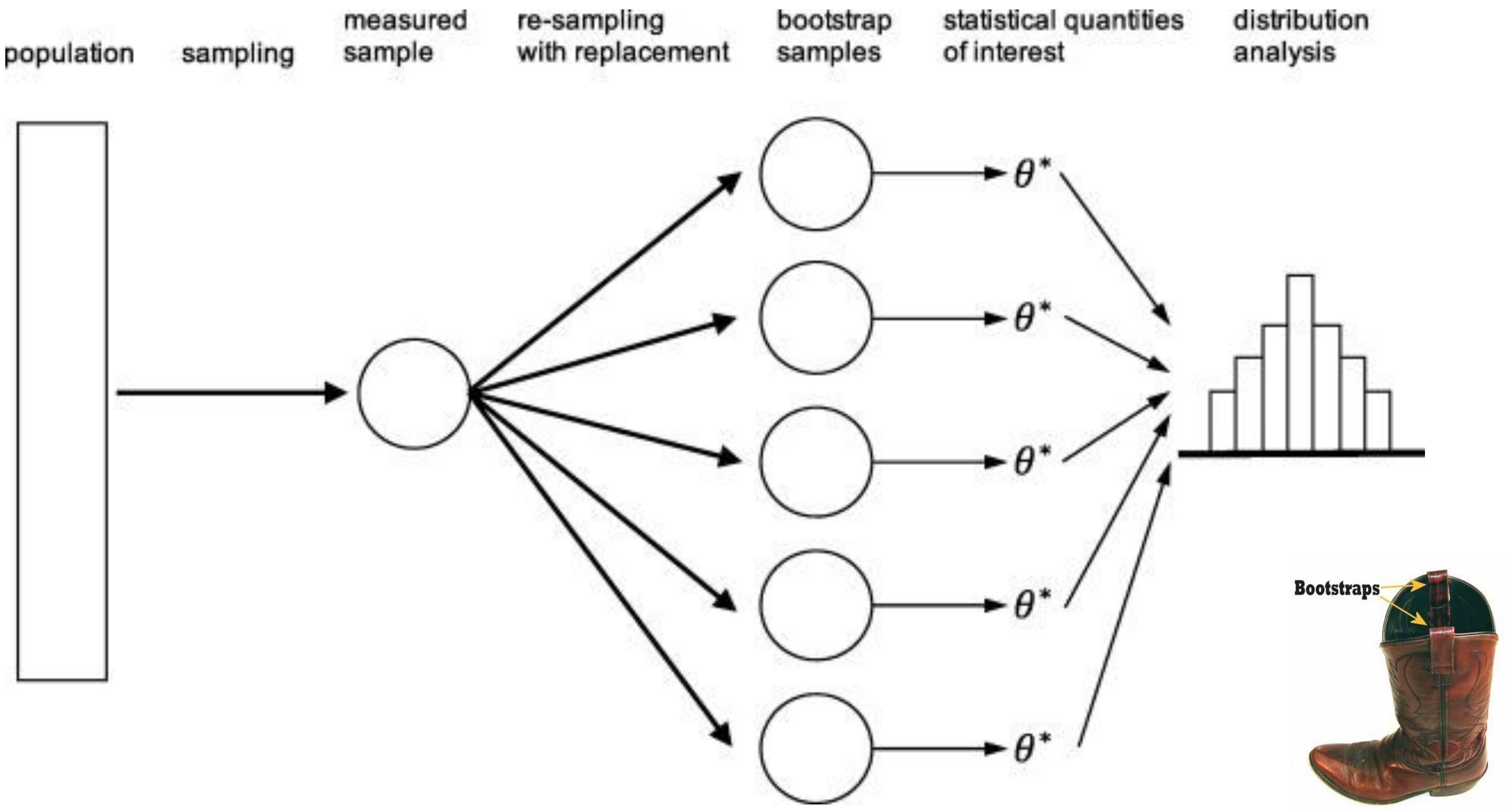
Cross validation method



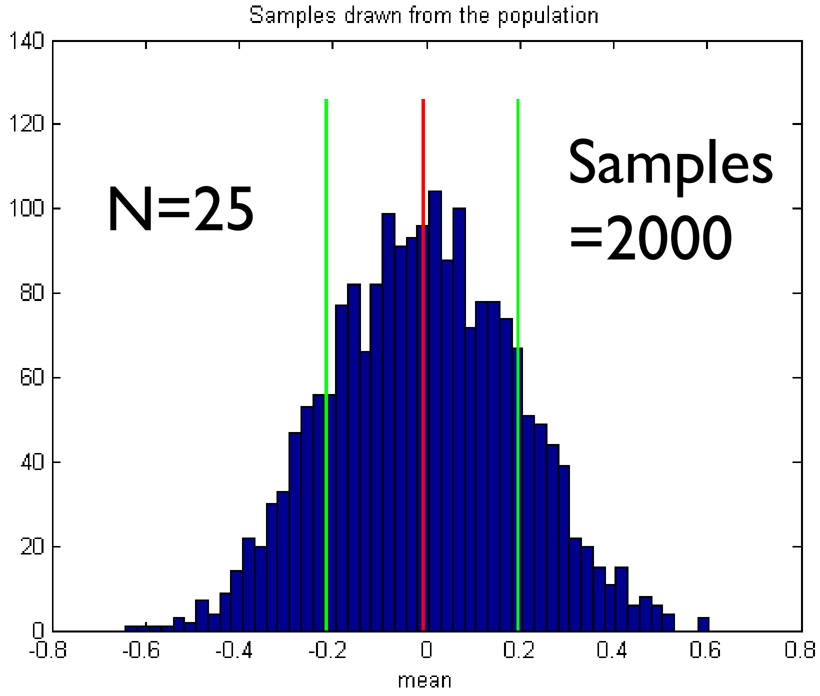


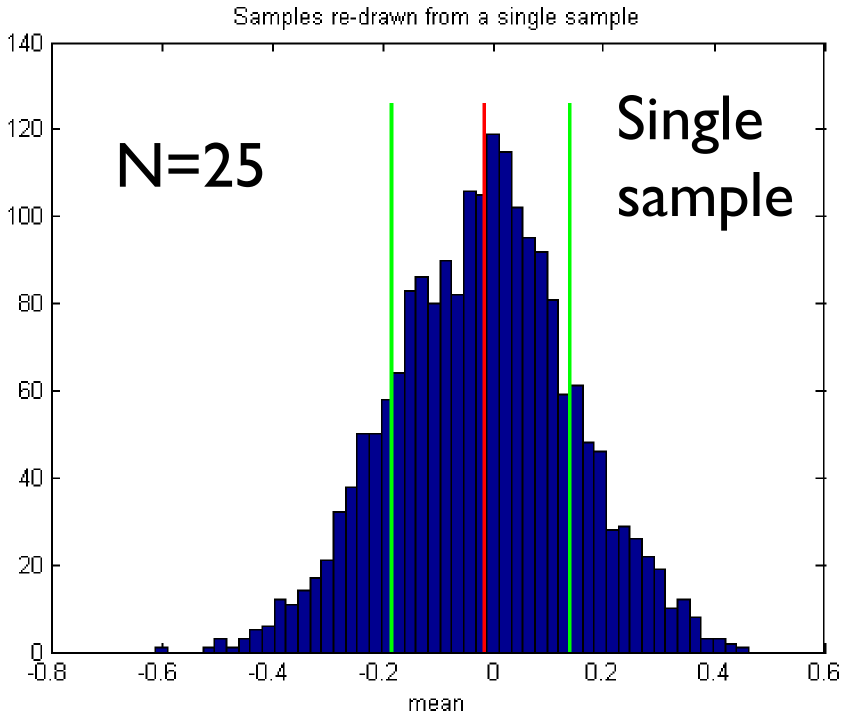
Vopnik-Chervonenkis (VC) dimension

Overall algorithm for bootstrapping



Multiple sample vs. single sample





Bootstrap average is not zero!
And yet, the s~0.18, just from a single sample.
The success of the method relies on precision measurement

Lecture 5: Quiz

- One must use "adjusted residual method" to compare the goodness-of-fit among the models
- AIC and BIC both involve penaltyies for increasing the number of parameters
- Cross-validation, Jack-knife and Bootstrap are computer-enabled method to quantify how well a model fits the data
- Cross validation is a special case of the Jack-knife method
- Parametric boot-strap method allows one to calculate the parameter uncertainty associated with the distribution function used to fit the dataset

Course Outline

$$\overline{y} = f\,(\overline{x}\,) \qquad \overline{x} = x_1, x_2, \dots x_n \qquad \overline{y} = y_1, y_2, \dots y_m$$

Lecture 1: Introduction
Lecture 2: Collecting and plotting $x_1, x_2, \dots x_n$
Lecture 3: Physical and empirical $f, F, df/dx$, ...
Lecture 4: Model selection between $f_1, f_2, \dots$
Lecture 5: Model Selection: Cross-validation and Bootstrapping method
Lecture 6: Scaling theory with known f, $f(\overline{x}) = f(\overline{X})$
Lecture 7: Scaling theory with unknown f, $\overline{x} \rightarrow X$
Lecture 8: Design of experiments to determine $\overline{y}_{\mathrm{max}} = f\,(\overline{x}\,)$
Lecture 9: DOE and ANOVA
Lecture 10: Principle component analysis for classifying $\{y\}$.
Lecture 11: Machine learning … Statistical approach to learn f
Lecture 12: Machine Learning …. Deep network, Karnaugh map, and other components
Lecture 13: Interpretable ML: Physics-based machine learning $f = f_{\mathrm{physics}} + \Delta f$
Lecture 14: Conclusions

Stress-induced cell death

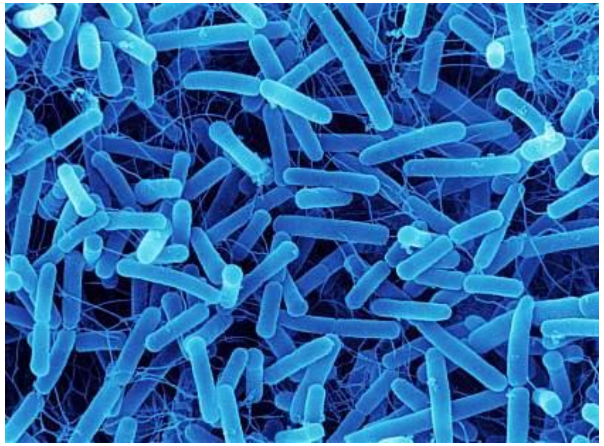
Equation: $\frac{dn}{dt} = -b(T)n$

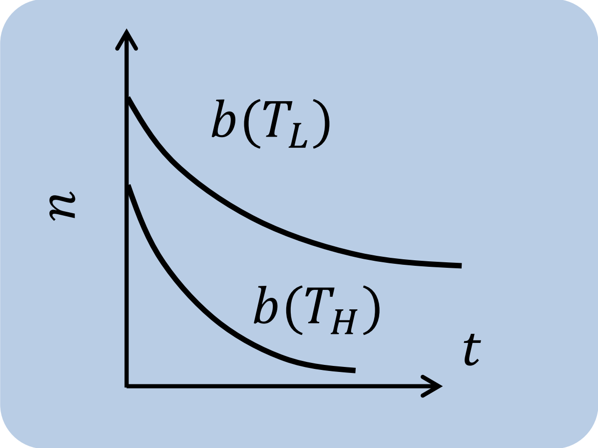
$\Rightarrow \quad n = n_0\, e^{-b(T)\,t} \equiv f(n_0, b, t)$

5 experiments each for n_0, b, t

... 125 measurements

If with multiple samples, hundreds of measurements required.





Buckingham PI Theorem

If the function g depends on parameters $q_1, q_2, \ldots\ q_n$, then

$$g(q_1, q_2, \ldots., q_n) = 0$$

The same expression can be expressed in terms of (n-m) independent dimensionless ration, or Π parameters.

$$G(\Pi_1, \Pi_2, \ldots., \Pi_{n-m}) = 0$$

m= minimum number of independent dimension typically given by r, where r is the rank of the matrix

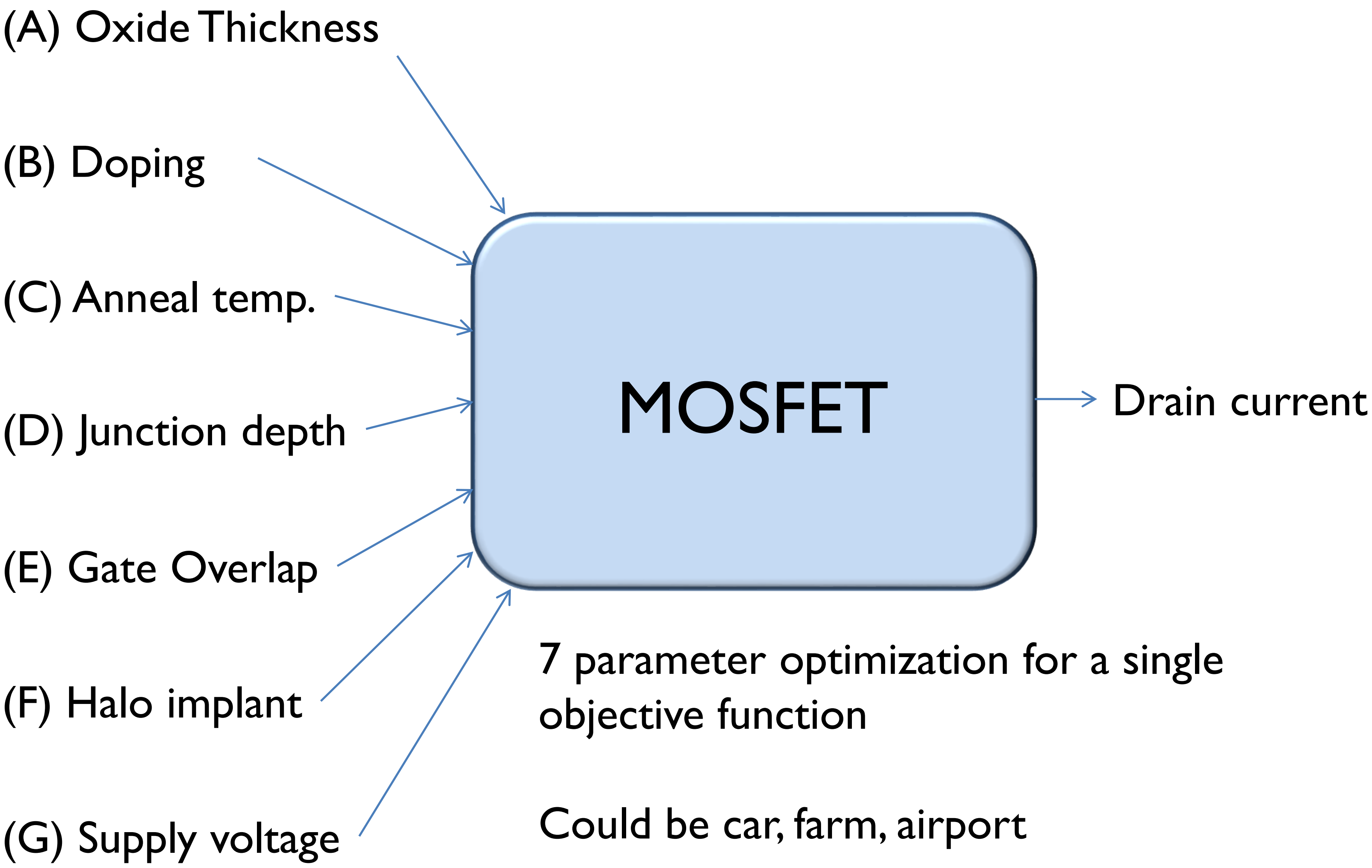
Lecture 6: Quiz

- The Buckingham-Pi parameters have dimension of length
- The Buckingham-Pi parameters are unique
- Buckingham-Pi theory is used to reduce the number of experiments
- Reynold's number is an example of dimensionless Buckingham-Pi parmeter

Lecture 7: Quiz

- Consider a two variable equation (one dependent, one independent). The number of scaled coefficient to be set to unit is 2.
- One only needs to scale the differential equation, not the boundary conditions
- Buckingham-Pi approach is a general form of scaling that can be used even when the differential equation is not known
- The scaling principle cannot be used to analyze coupled different equation
- The goal of the scaling theory is to reduce the number of experiments necessary to quantify a model

Fisher's design of experiment



7 parameter optimization for a single objective function

Could be car, farm, airport

Taguchi table: Continued

$L_4(2^3)$

Run	Columns		
	1	2	3
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1

$L_8(2^7)$

Run	Columns						
	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

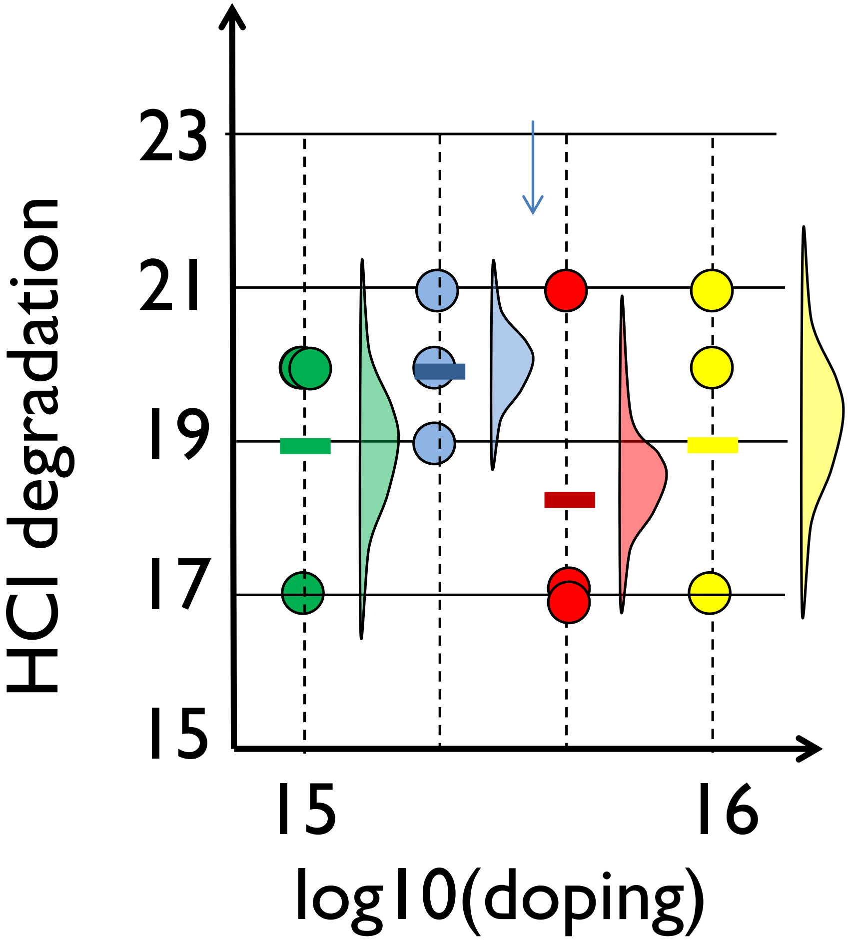
$L_{12}(2^{11})$

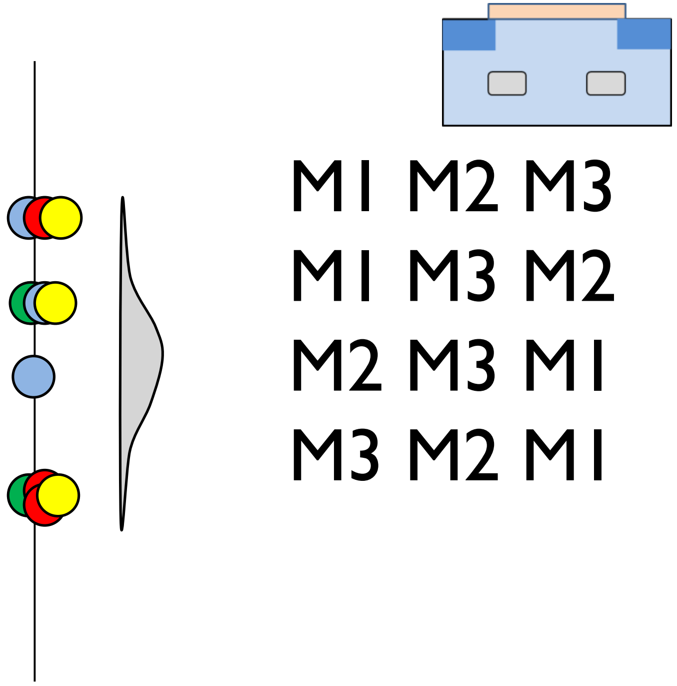
Run	Columns										
	1	2	3	4	5	6	7	8	9	10	11
1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	2	2	2	2	2	2
3	1	1	2	2	2	1	1	1	2	2	2
4	1	2	1	2	2	1	2	2	1	1	2
5	1	2	2	1	2	2	1	2	1	2	1
6	1	2	2	2	1	2	2	1	2	1	1
7	2	1	2	2	1	1	2	2	1	2	1
8	2	1	2	1	2	2	2	1	1	1	2
9	2	1	1	2	2	2	1	2	2	1	1
10	2	2	2	1	1	1	1	2	2	1	2
11	2	2	1	2	1	2	1	1	1	2	2
12	2	2	1	1	2	1	2	1	2	2	1

$L_N(S^M)$

M ..Variables
S … Levels
N … experiment > DOF = 1 + M(S-1)

Analysis of Variance: Single/multiple factors Analysis





M1 M2 M3
M1 M3 M2
M2 M3 M1
M3 M2 M1

Important to cancel out any machine specific issue

Single factor ANOVA: Wood Treatment

replicates →

treatments ↓

	1	2	3	4	5	6
5	7	8	15	11	9	10
10	12	17	13	18	19	15
15	14	18	19	17	16	18
20	19	25	22	23	18	20

s-avg	(s-avg-AVG)^2
10.00	35.50174
15.67	0.085069
17.00	1.085069
21.17	27.12674
15.96	63.79861

$$\sum(data - AVG)^2 = 512$$

$$6 \times\ 63.8\ = 382.8$$

Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	382.7917	3	127.5972	19.60521	3.59E-06	4.938193
Within Groups	130.1667	20	6.508333			
Total	512.9583	23				

Lecture 8: Quiz

- The person credited with developing design of experiment is ?
- One-factor-at-a-time is an efficient DOE technique
- A full-factorial DOE involving 4 factors and 3 levels is given by ?
- Taguchi tables are used when the experiments numbers are (multiples of 2, 4, 6, 8).
- Fractional factorial experiments are defined by Setting a higher order correlation term to 1

Lecture 9: Quiz

- If the levels of a factor is 7 and the number of replicates is 6, then the Degree of freedom is … (7*6 – 1) = 41
- Mean-squared error for the previous problem is given by = SS_error/**N**, where N is (41 – (7-1)) = 35.
- The ANOVA significance depends on (p level desired, number of level associated with a factor, number of replicates, **all of the above**)
- ANOVA analysis can be used to analyze the data for (Full factorial design, fractional factorial design, Taguchi table, P-B design, **All of the above**)

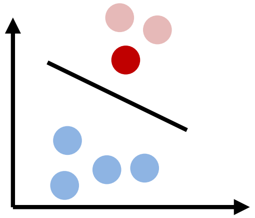
Course Outline

$$\bar{y} = f(\bar{x}) \quad \bar{x} = x_1, x_2, \ldots x_n \quad \bar{y} = y_1, y_2, \ldots y_m$$

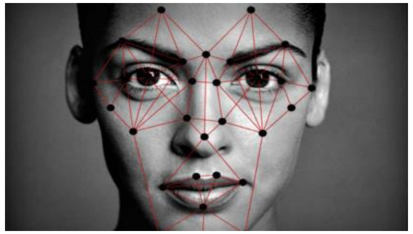
Lecture 1: Introduction
Lecture 2: Collecting and plotting $x_1, x_2, \ldots x_n$
Lecture 3: Physical and empirical $f, F, df/dx, \ldots$
Lecture 4: Model selection between $f_1, f_2, \ldots$
Lecture 5: Model Selection: Cross-validation and Bootstrapping method
Lecture 6: Scaling theory with known f, $f(\bar{x}) = f(\overline{X})$
Lecture 7: Scaling theory with unknown f, $\bar{x} \rightarrow X$
Lecture 8: Design of experiments to determine $\bar{y}_{\max} = f(\bar{x})$
Lecture 9: DOE and ANOVA
Lecture 10: Principle component analysis for classifying $\{y\}$.
Lecture 11: Machine learning … Statistical approach to learn f
Lecture 12: Machine Learning …. Deep network, Karnaugh map, and other components
Lecture 13: Interpretable ML: Physics-based machine learning $f = f_{\text{physics}} + \Delta f$
Lecture 14: Conclusions

Classification problem in big data

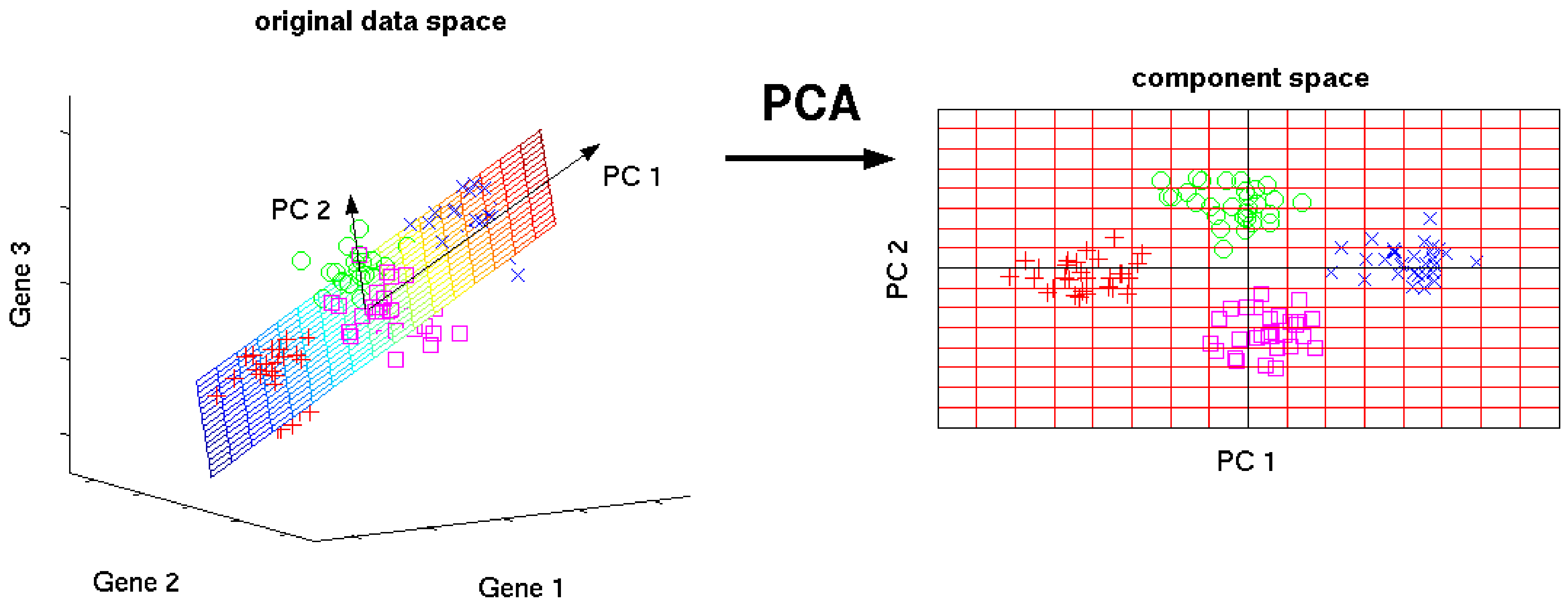
Advertisement
Recommendation



Facial Recognition
Voice Recognition
Spam Filtering



Principle Component Analysis: Classification

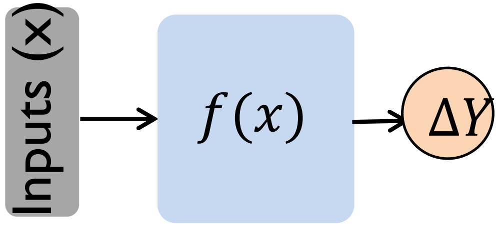


If you like this book, you will also like that book (because you belong to the same category)

Lecture 10: Quiz

- Which of the following ML algorithm involves unsupervised learning (PCA, ANN, Clustering, Logistic regression, decision trees)
- SVD technique is used to (cluster the data, reduce the data for subsequent clustering, train ML algorithm)
- The Sigma matrix determines … Weight of the principle components.
- The V matrix determines …. Directions of the new axes
- The columns of the U*Sigma matrix determines … Projection along the principal components

Statistical Machine Learning

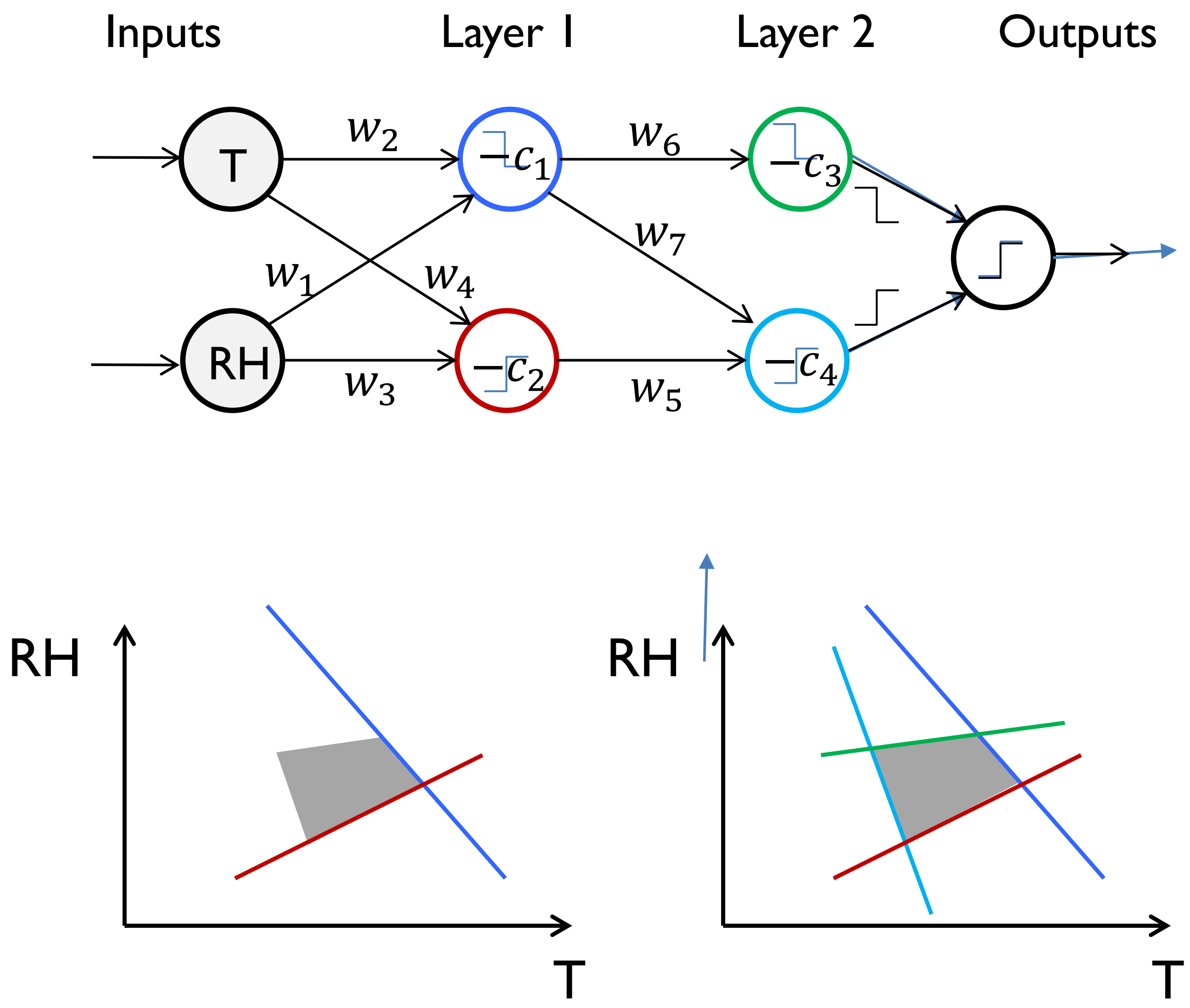


$$y = f(x)$$

y: pass, fail
y: A, B, C, D, E
Y= grade points.

$f(x)$ … Physics
$f(x)$ … Statistical curve fitting
$f_{\max}(x)$ …. Design of expt

Deep network



Lecture 11: Quiz

- Which of the following is NOT an activation function (sigmoid, tanh, recLu, Gaussian, shifted sigmoid)
- The cost function or loss-function is related to (MLE, SVD, PCA, Node number, Front Propagation)
- In a N-dimensional parameter space, every bubble in the ANN signifies (A circle, a sigmoid, a (N-1) dimensional hyperplane, a neuron, none of the above)
- A support vector machine is used to: create a more robust demarkation among the clusters

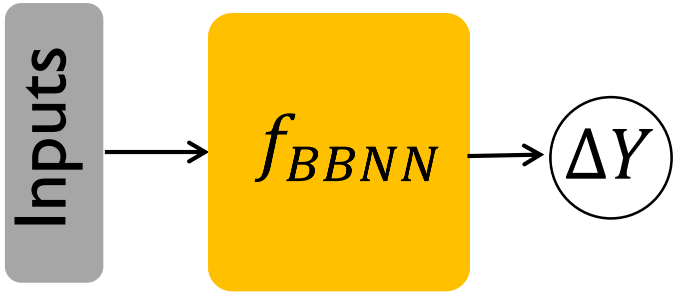
Lecture 12: Quiz

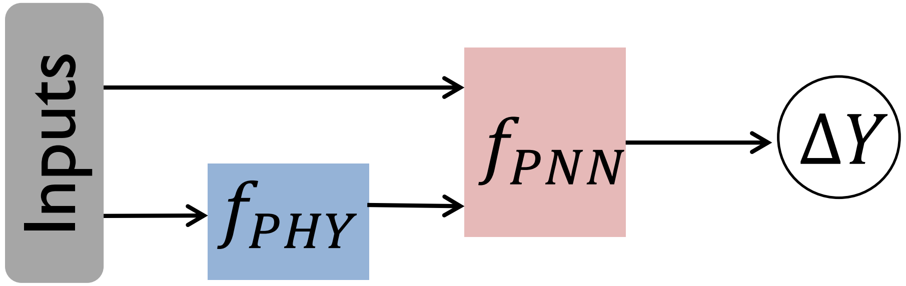
- A deep neural network involves (supervised learning, a special type of ANN with many layers, can integrate subparts for better classification, used for image recognition, all of the above).
- Backpropagation is iterative way of: Fitting the coefficients of an ANN)
- An epoch is defined by: (1000 years, a single back-propagation cycle)
- TensorFlow (can create ANN, has multiple activation energy, can work with many layers of deep neural network, all of the above)

A falling ball with gravity and resistance

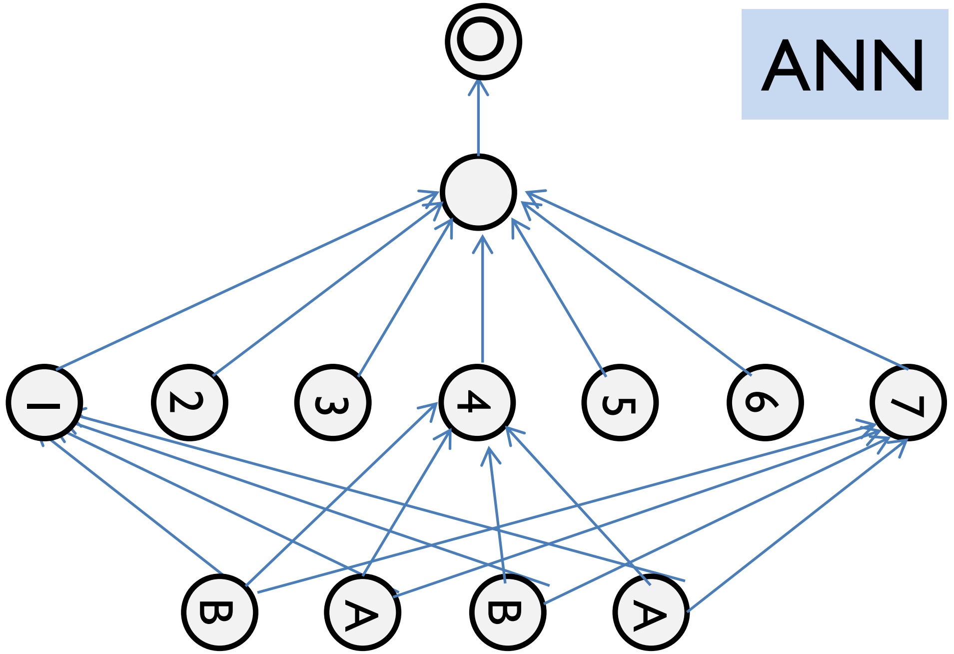
$$\frac{dv}{dt} = -\frac{g}{(1+z/R)^2} + bv^2$$

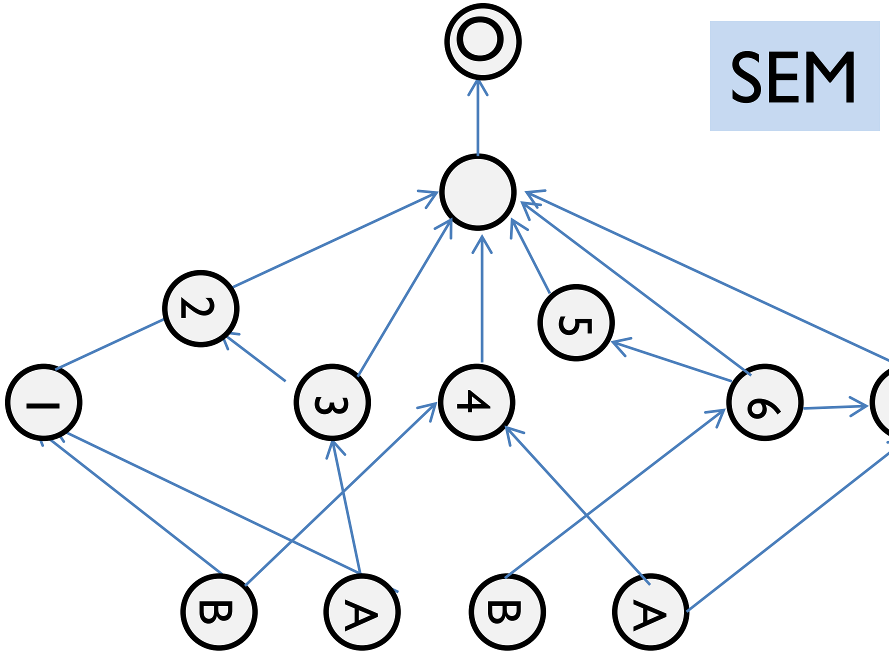
A numerical/perturbative solution may not be possible. In practice, b(z) is unknown function of humidity, temperature, etc. A machine learning approach is preferred.





Structural Equation modeling: Motivation





Nodes defined statistically
not appropriate for extrapolation

Nodes defined physically
Interconnects are nonlinear
Extrapolation possible

Lecture 13: Quiz

- ML algorithm: (essentially curve-fitting procedure, training is expensive, coefficients are unphysical, needs large data, all of the above)
- Physics-based ML assigns the dominant effect to: (Physics of the problem, statistics of the problem, deep neural network)
- The ANN associated with physics-based ML is: (small than, similar to, or larger than) standard ML.
- Structural equation modeling is a type of: (Deep learning network, domain-knowledge based network, clustering network)
- The weights of structural equation modeling involves: (single linear function, multiple linear function, any function, only constant terms)

Conclusions

1. Data is the lifeblood of modern science and technology. Learning to treat data with respect is an essential skill.

2. Modern data analysis is easy because they are embedded in systems. Once we become aware of the treasure-trove of functions available, data analysis is considerably simplified.

3. Design of experiments is a powerful approach for modern manufacturing. The scaling theory, Taguchi techniques, ANOVA all help reduce the dimensionality of the problem.

4. Machine learning is a modern way of curve-fitting, powered by modern computers. Its applications in advertisement and classification have been remarkable. One must develop the technique further for applications in science and engineering.

Review Questions

1. Name three concepts that you are going to remember from this course. Explain why they were particularly important for you.
2. What is the essential different between classical and modern statistics?
3. What are the topics from this course you can immediately apply to your research? What specific problems would you be able to apply them?
4. What are your over-all impression about the course? Did you find the concepts coherent or was the diversity of concepts distracting?
5. What additional topics would you have preferred that we cover in this course?