

Lecture 1

TAYLOR DEVET MASC.

PHD. CANDIDATE BIOLOGICAL AND BIOMEDICAL ENGINEERING

MCGILL UNIVERSITY

SHRINERS HOSPITAL FOR CHILDREN

Todays Aims...



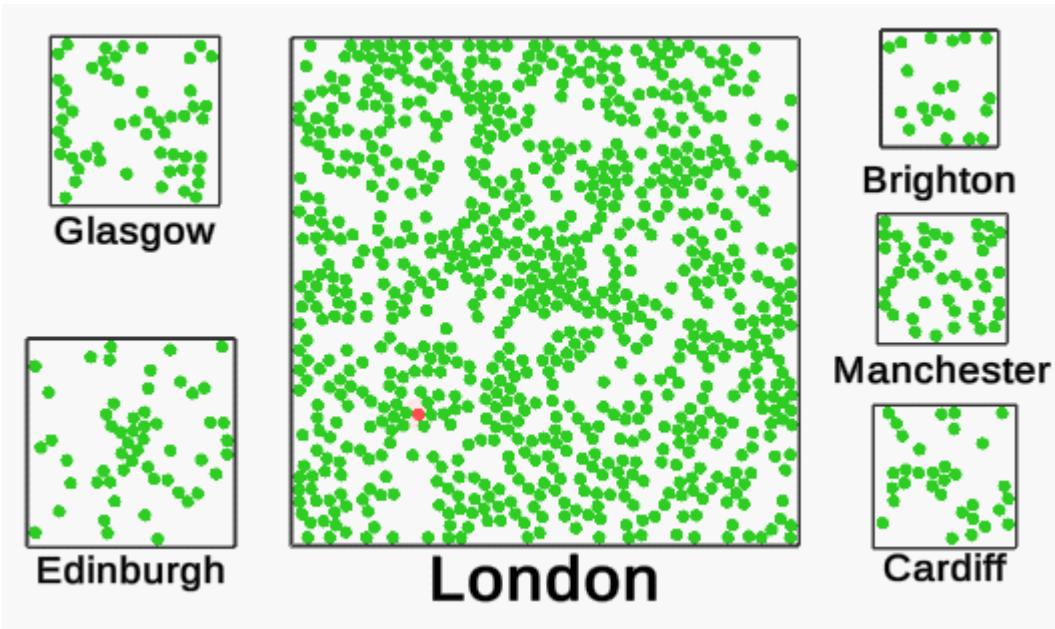
What is Biological
Modelling?



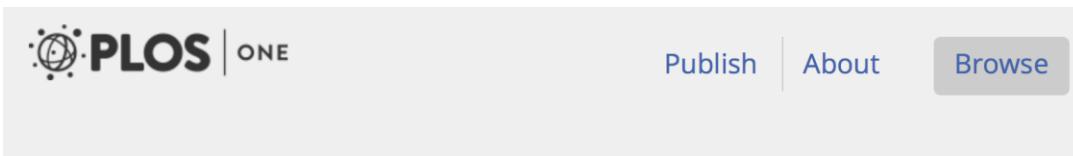
Why do we do it?



How do models compare
to the real data?



<https://www.alanzucconi.com/2020/03/30/mathematics-epidemics/>



OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Fractal Analysis of Brain Blood Oxygenation Level Dependent (BOLD) Signals from Children with Mild Traumatic Brain Injury (mTBI)

Olga Dona, Michael D. Noseworthy, Carol DeMatteo, John F. Connolly

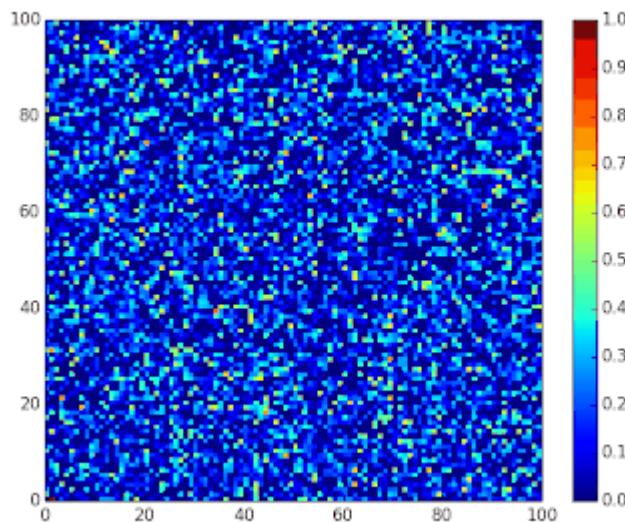
Published: January 10, 2017 • <https://doi.org/10.1371/journal.pone.0169647>

Special Issue

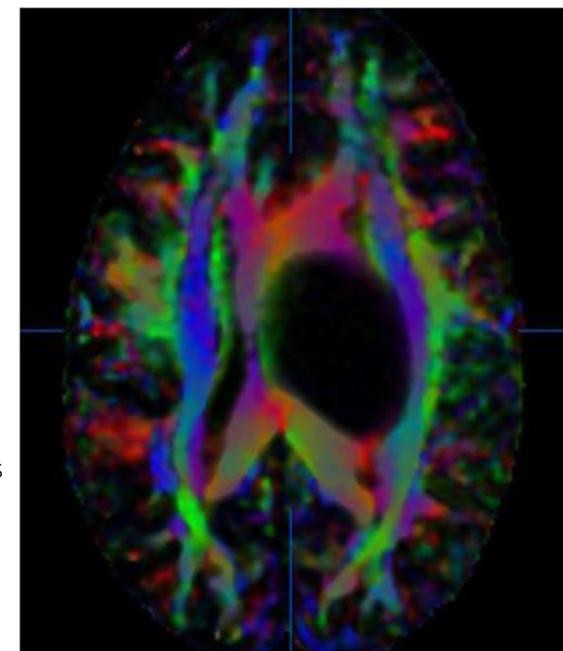
Received: 12 September 2007, Revised: 11 December 2007, Accepted: 1 February 2008, Published online in Wiley InterScience: 14 May 2008
www.interscience.wiley.com DOI: 10.1002/cem.1143

Dynamic contrast-enhanced MRI diagnostics in oncology via principal component analysis

Mark-John Bruwer^a, John F. MacGregor^{a*} and Michael D. Noseworthy^{b,c,d}

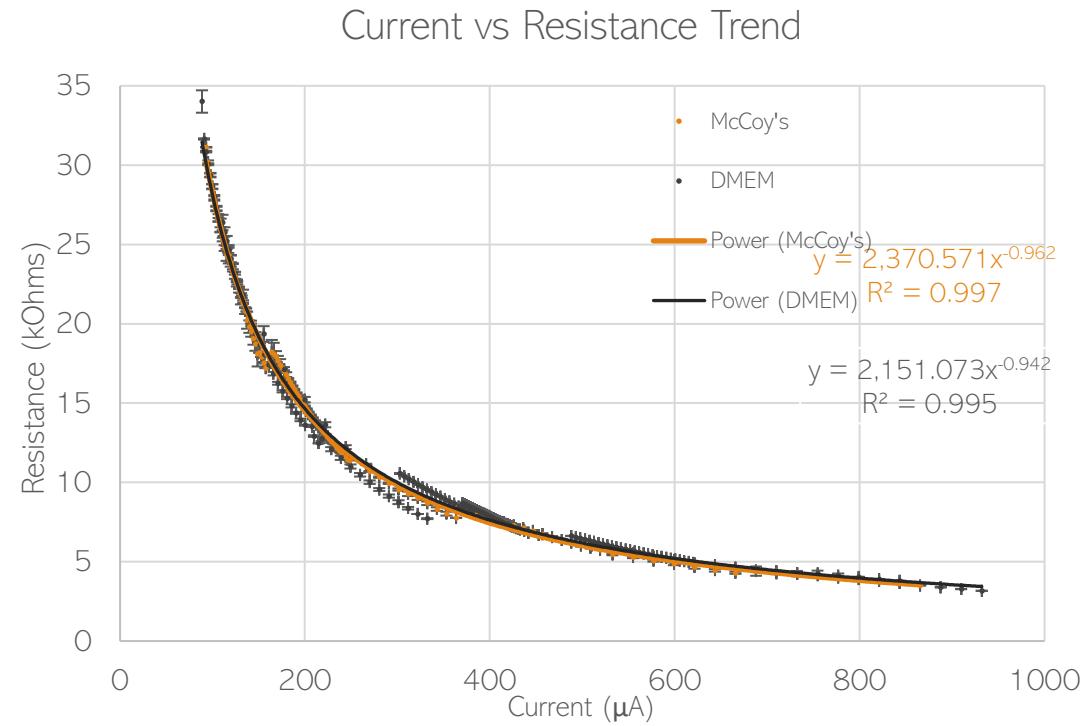
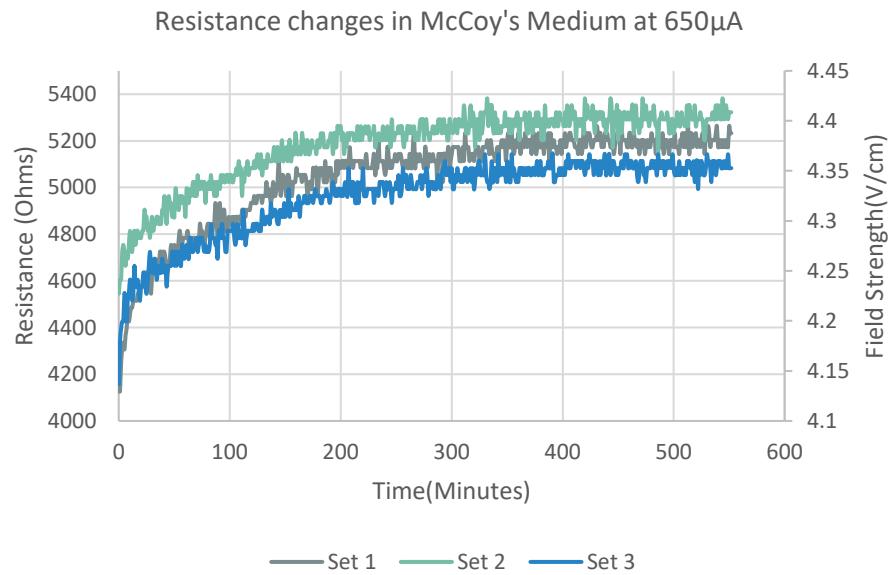


https://en.wikipedia.org/wiki/Epidemic_models_on_lattices



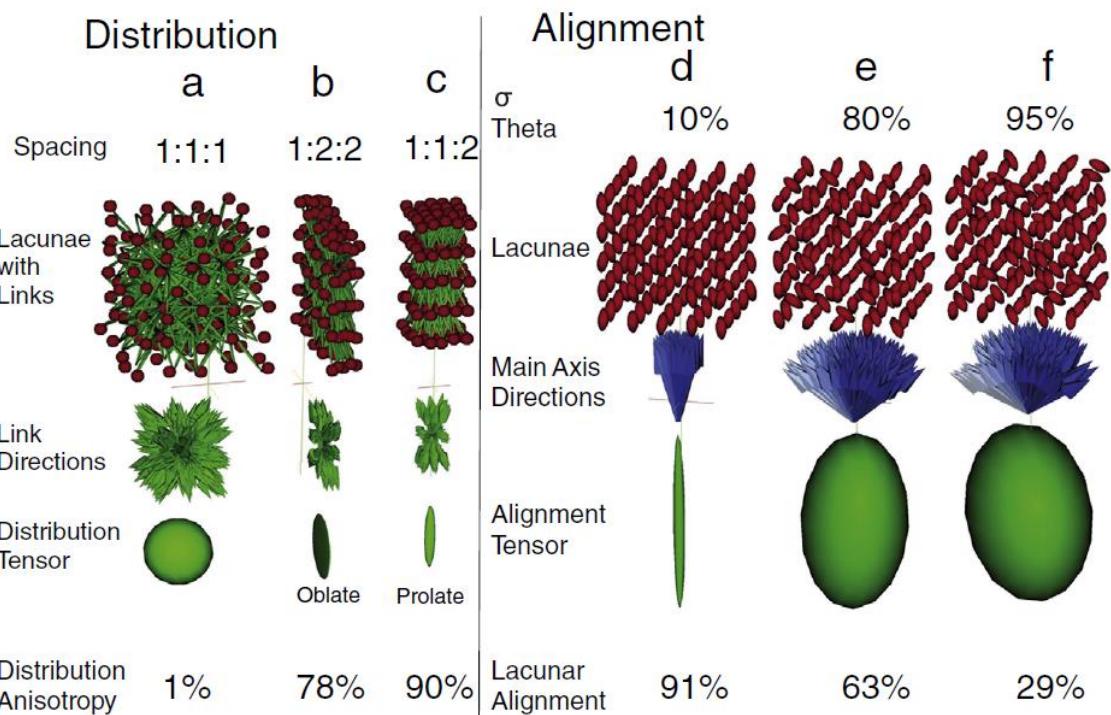
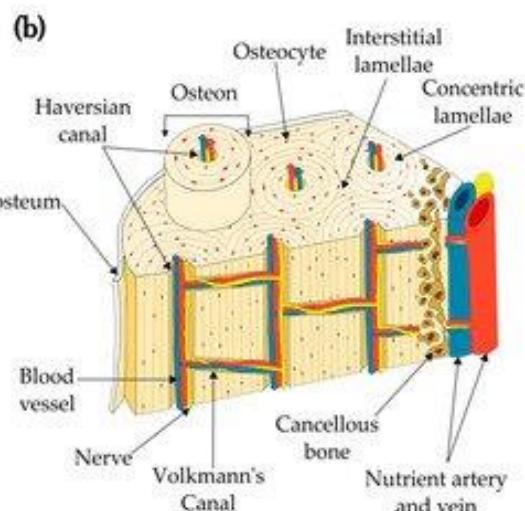
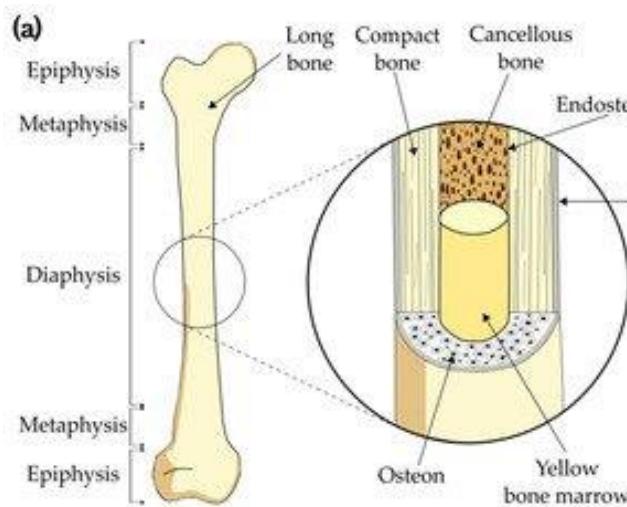
<https://www.frontiersin.org/articles/10.3389/fneur.2017.00660/full>

Electrical Properties of Cell Culture Media



Modelling of Osteocyte-Lacunocanicular network

K.S. Mader et al. / Bone 57 (2013) 142–154



Model Types

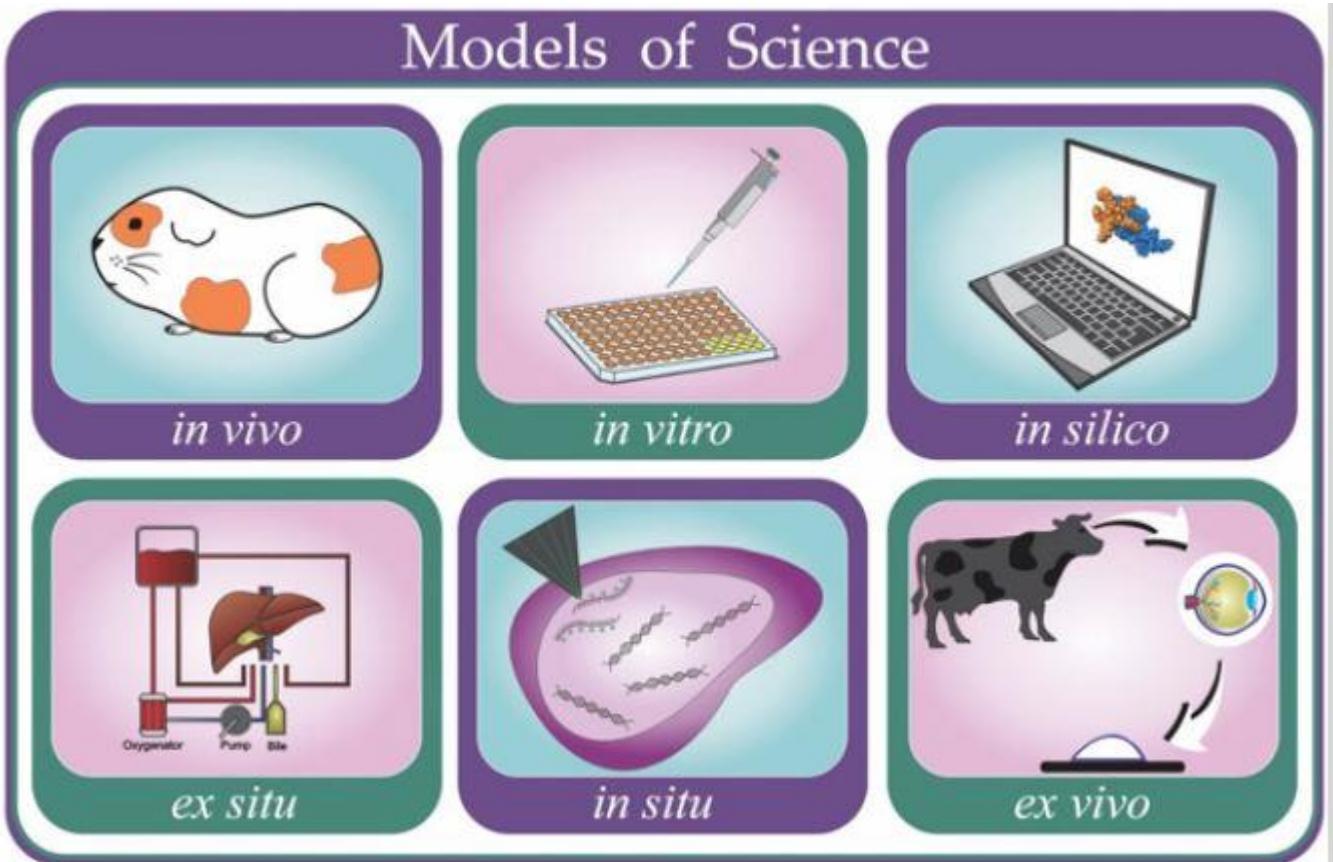
In Silico – “in silicon” - computer model

In Vitro – “in the glass” – cell model

In Vivo – “in the living” – whole organisms - plants animals

In Situ – “on site”

Computer model --> Cell model -->
Small Animal Model--> Large Animal
Model--> Human Trials



1) What do you want to study?? Create a model

- based on some *a priori* understanding of the system to test
- build a model based on physical/physiological/chemical characteristics
- consider inclusion of assumptions and constraints to help simplify things

2) Test the model (part 1)

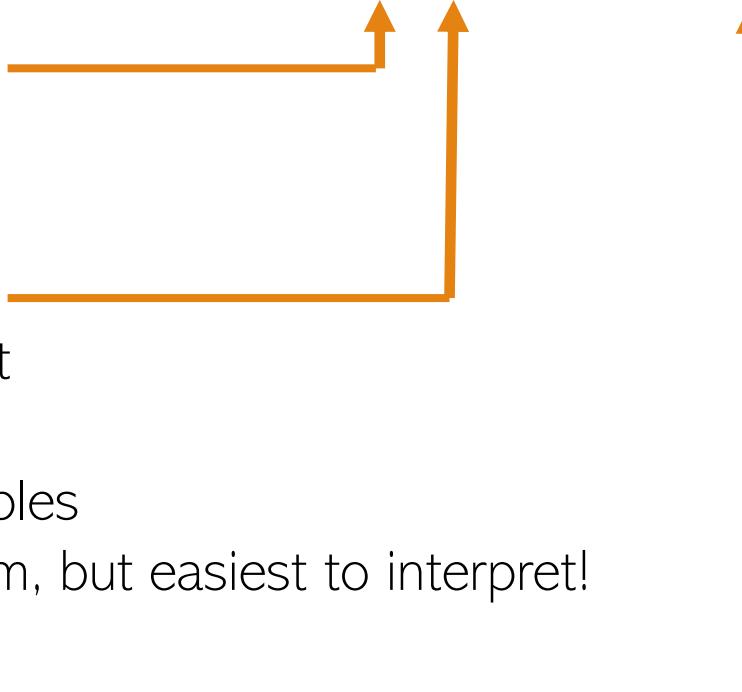
- computer simulation
- programming

3) Test the model (part 2)

- build the system to test
- mock system
- can control many variables
- hardest to get data from, but easiest to interpret!

4) Test the model (part 3)

- the REAL thing → your biological system!!
- the easiest to get data from, but hardest to interpret!



ELEMENTS

WHAT ARE THE ODDS WE ARE LIVING IN A

Elon Musk says we may live in a simulation. Here's how we might tell if he's right

Scientists are looking for ways to put this mind-bending idea to the test.



The posthuman future has never been easier to imagine—especially for those who work at the forefront of technology.

PHOTOGRAPH BY JANUS FILMS/EVERETT

Simulation

- a comprehensive method for studying systems
- refers to an entire process which includes:
 - choosing a model
 - finding a way to efficiently implement the model on a computer
 - calculating output of the algorithm
 - and visualizing and studying the resultant data

Simulation Types

Equation Based Simulation

A set of equations that describe the system

Executing the model is solving the series of equations

Commonly ordinary differential equations (ODE) or over time and space partial differential equations (PDE)

i.e. simulation of fluid flow

Agent Based Simulation

Simulates the interactions of autonomous agents

Used in behavioural and social sciences where studying network interactions of individuals

i.e. n-many discrete particles

Simulation Types

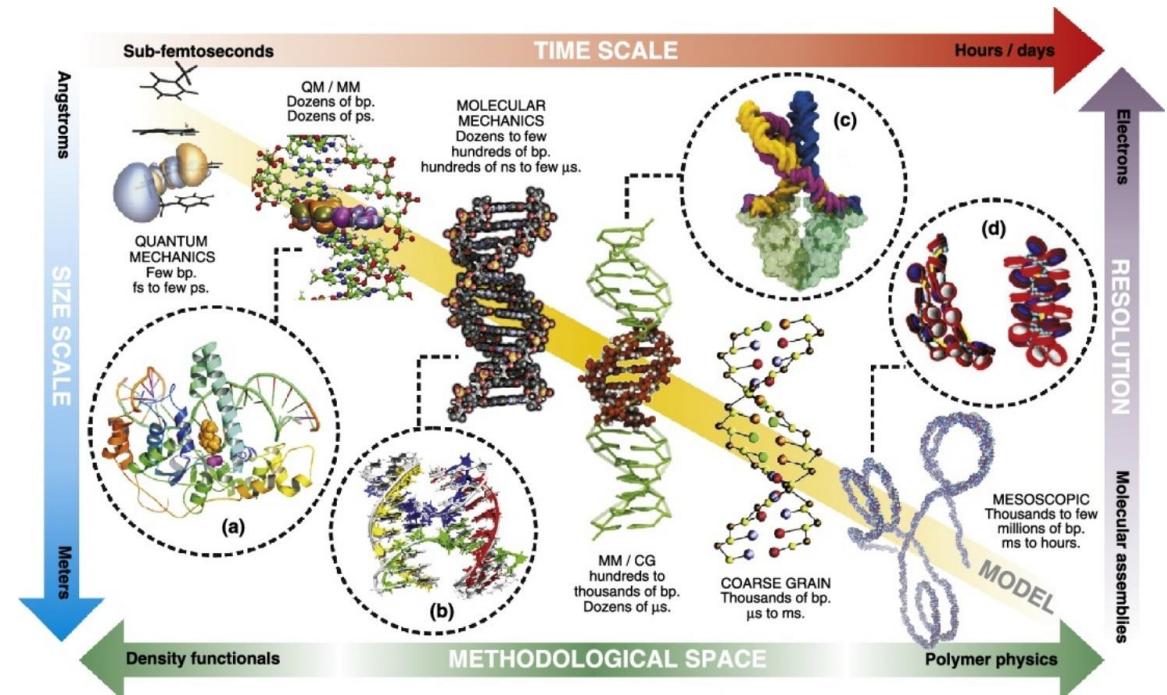
Multi-scale simulations

Multiple models at various scales are used together to describe the system

Monte Carlo Simulations

System that uses randomness to generate a model/outcome

Used to look at predictability of models



<https://www.sciencedirect.com/science/article/abs/pii/S0959440X15001761>

Computer Simulations

- assist in the design, creation, and evaluation of complex systems to understand and evaluate 'what if' case scenarios.

Benefits:

- Gain better understanding of a process or group of interconnected processes. Also used to identify problem areas
- Evaluate effect of systems or process changes such as demand, resources, supply, and constraints. This could help in cost assessment, disease prediction, etc.
- Identify actions needed upstream or downstream relative to a given operation, organization, or activity to either improve or mitigate processes or events
- Evaluate impact if a change in the system occurs

Computer Simulations

Cons:

Have to have a very thorough understanding of our system before starting

- Initial conditions
- Parameters
- Variability?

Can make mistakes while creating the model

Have to find a way to interpret the results

Research Ethics

Essentiality

- Scrutinized by external body (Animal committee, ethics board etc.)

voluntariness, informed consent, and community agreement

- Participants should be made aware of risks and benefits

non-exploitation

- Participants should be made aware of all danger

privacy and confidentiality

- Records are kept confidential

precaution and risk minimization

- Minimal risk to participants at all phases of the study

Research Ethics cont.

professional competence

- Conducted by qualified people

accountability and transparency

- Research done in a fair, honest, impartial, and transparent manner

maximization of the public interest and of distributive justice

- Research should benefit all, not just socially better off

public domain

- Findings should be public domain

Should all be considered when doing any biological study

Research Ethics - HeLa

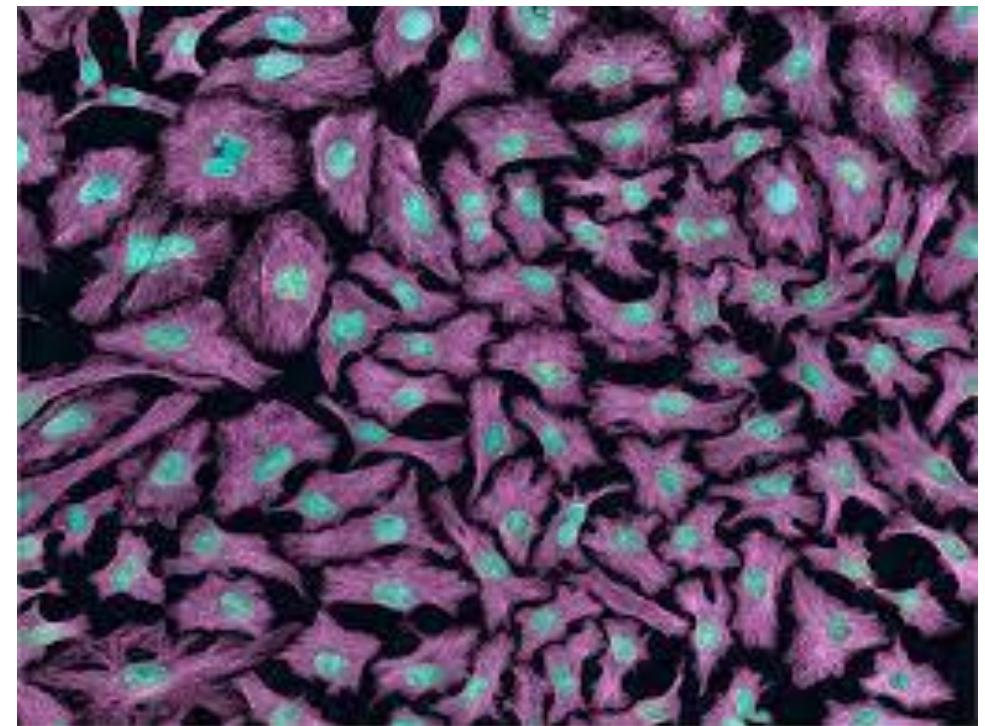
Most used immortal cell line

1st successful human in vitro line

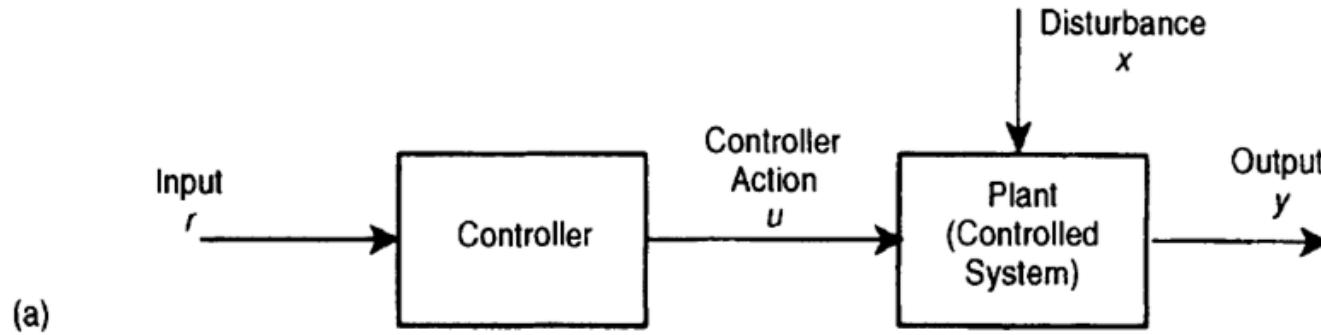
Taken from cervical cancer of Henrietta Lacks

No consent was given, ante or post-mortem...

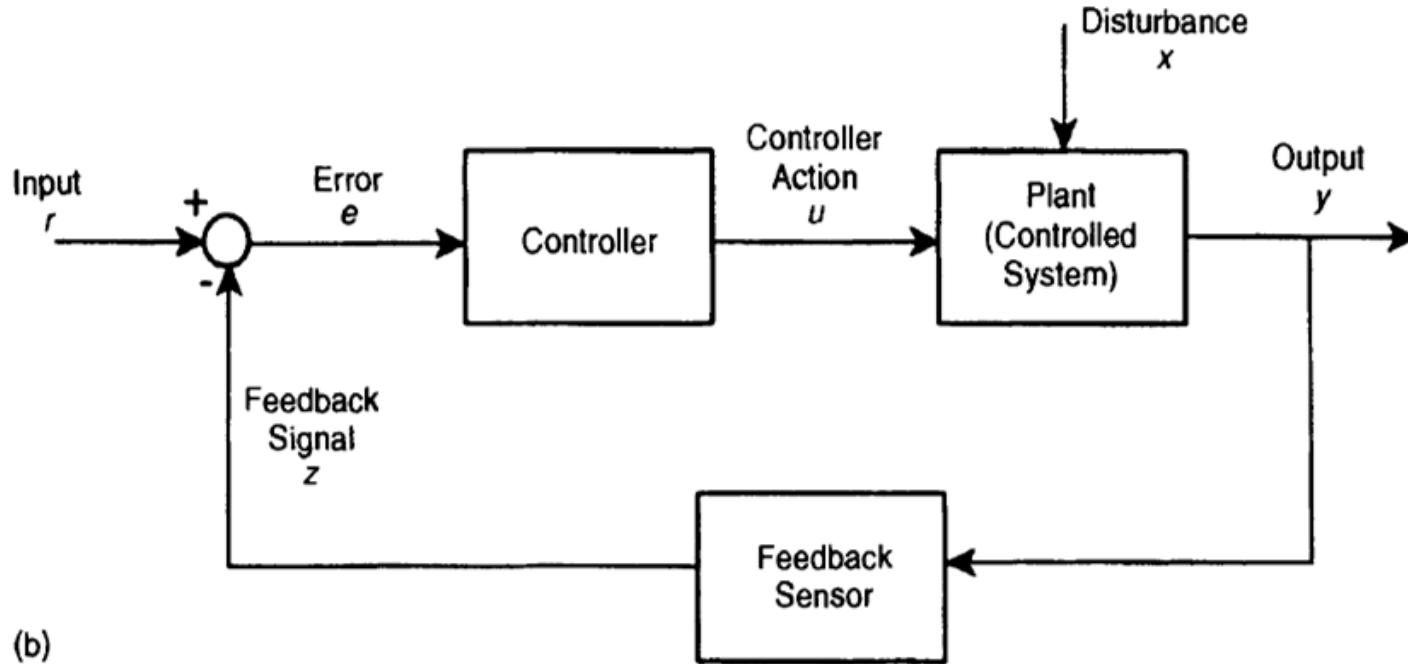
No anonymity



Biological Systems as Controlled System: Open Loop

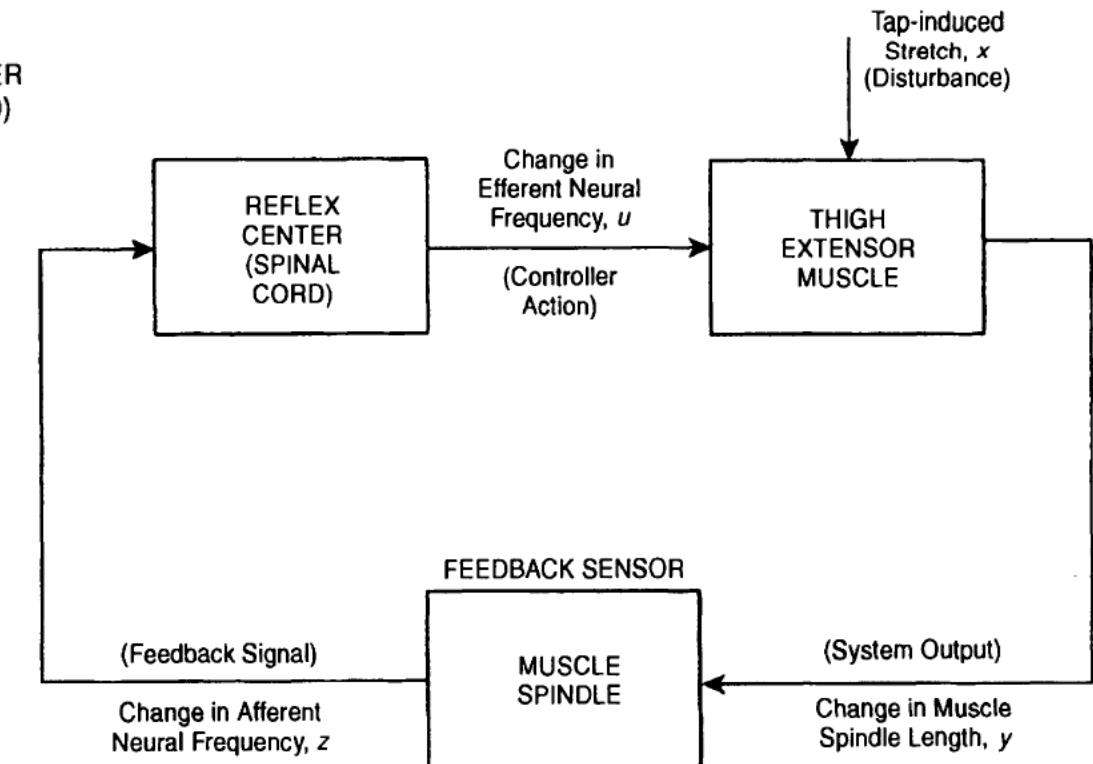
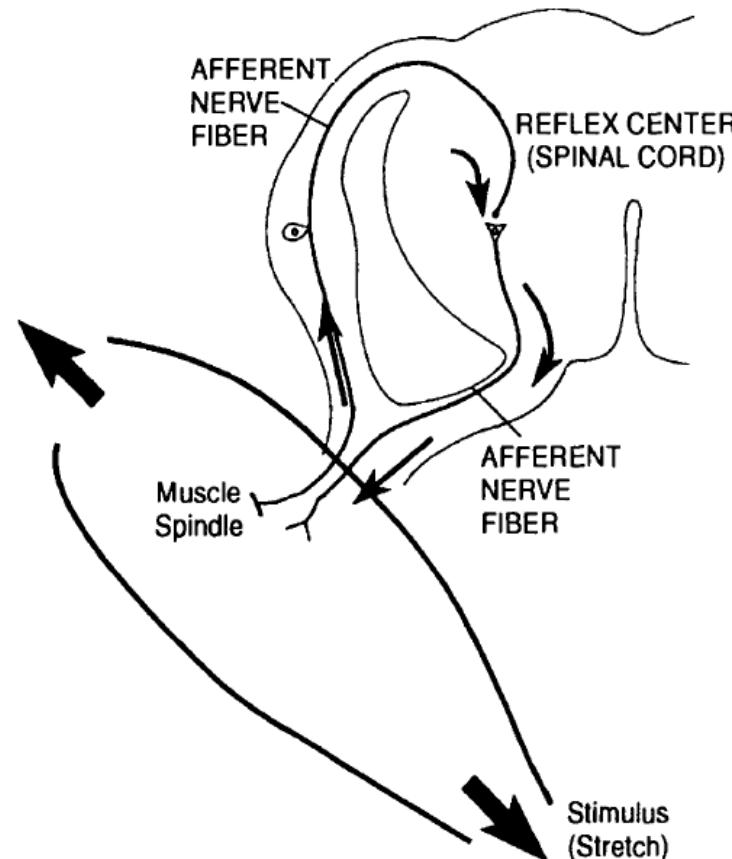


Biological Systems as Controlled System: Closed Loop

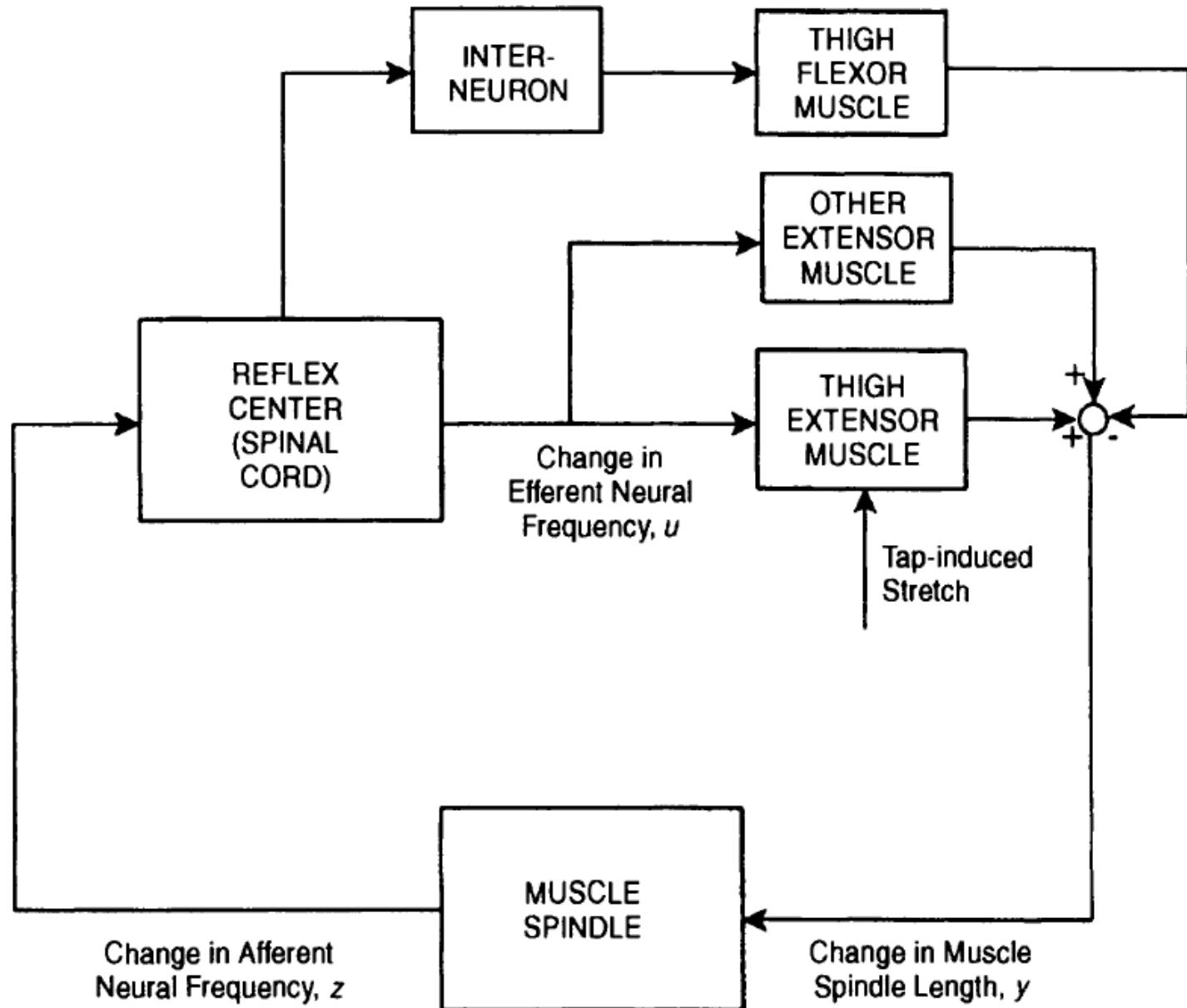


Biological Examples

- body temperature
(homeothermic vs.
poikilothermic)
- heart rate
(sympathetic vs.
parasympathetic)



More thorough



Engineering Control vs. Biological Control System

First Difference

Engineering

- accomplish a defined task with extensively optimized, fine-tuned parameters.
- Will perform its task "optimally" manner (at least, under the circumstances in which it is tested).

Biological

- built for versatility and may be capable of serving several different functions.
- e.g. primary purpose of the respiratory system is gas exchange, a secondary but also important function is to facilitate the elimination of body heat.

Engineering Control vs. Biological Control System

Second Difference

Engineering

- built by the designer
- characteristics of various components are generally known.

Biological

- consists of components that are unknown and difficult to analyze.
- Need to apply system identification techniques to determine how these various subsystems behave before proceeding to analyzing the overall control system.

Engineering Control vs. Biological Control System

Third Difference

Engineering

- Tend to be as straightforward as possible

Biological:

- extensive degree of cross-coupling or interaction among different physiological control systems.
- e.g. cardiovascular system has a large dependence on interactions with the respiratory, renal, endocrine, and other organ systems.

Engineering Control vs. Biological Control System

Fourth Difference

Biological:

- Physiological control systems, in general, are adaptive.
- Thus a system may be able to offset any change in output not only through feedback but also by allowing the controller or plant characteristics to change.

Engineering:

- Can be designed to be adaptive
- In general, conditions are set and system follows those conditions

Engineering Control vs. Biological Control System

Fifth Difference

Biological:

- Feedback in biological systems is often embedded.
- Not always a simple sensor that provides negative feedback

Engineering

- In engineering this is often a sensor that imposes –ve feedback (say through subtraction) to the system.,

Engineering Control vs. Biological Control System

Sixth Difference

Biological:

- Generally nonlinear,

Engineering:

- Can be linear or nonlinear.
- Frequently, an engineer prefers to design or use linear system components since they have properties that are well-behaved and easy to predict.

Before we model...
what does our Data
look like?

Data Types

Different Types of data demand different types of analysis

Quantitative

Discrete (**integer, ordinal**)
(e.g. # of petals on flower)

Continuous (**interval**)
(weight, height, depth, etc.)

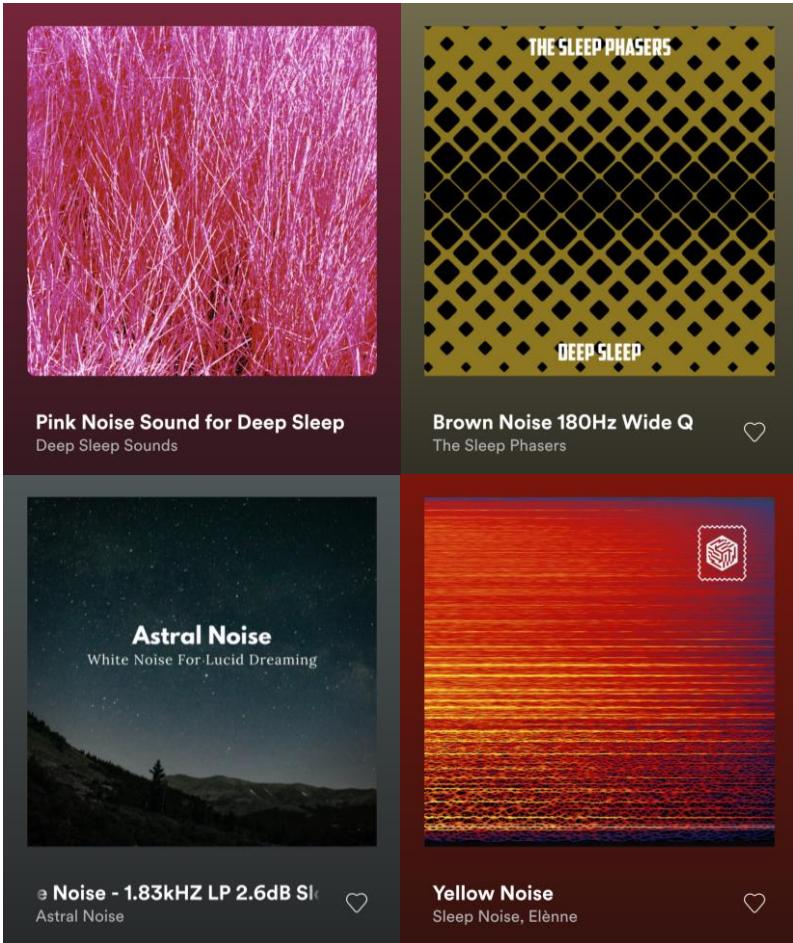
Qualitative

Discrete (**nominal**)
(classified, bins, ethnic
background, sex)

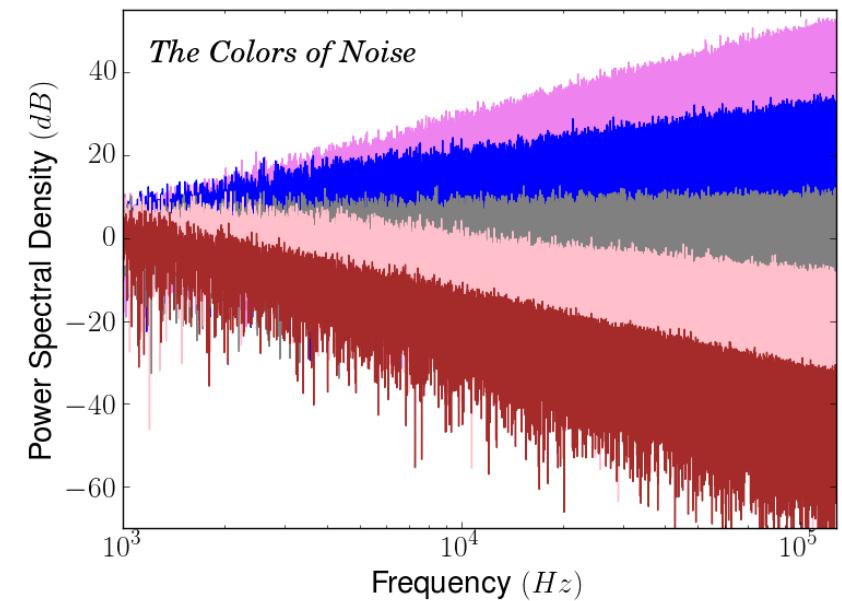
Noise

- An obstacle/interference to your message/signal/data
- Any unwanted data that doesn't help explain the relationship or feature you're looking for
- Types of biomedical noise
 - physiological variability
 - environmental noise or interference
 - transducer-induced noise
 - electronic noise

Noise Colour



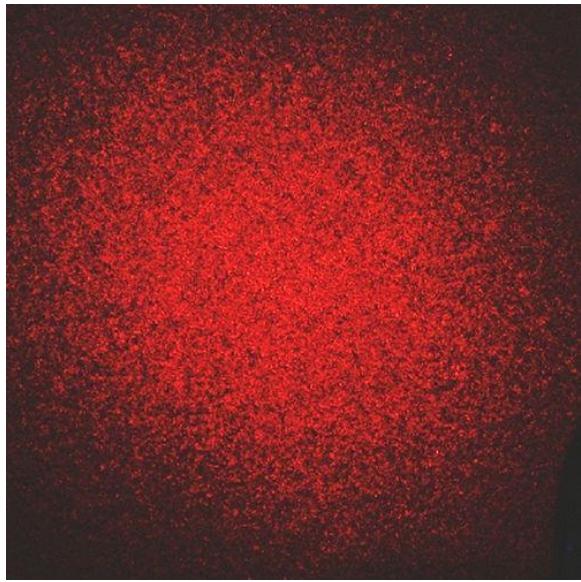
- White noise – uniform noise, random error
- Pink noise – $1/f$ pattern – intensity decreases with frequency
- Red noise – more low frequency than average
- Blue – more high frequency than average
- Most follows normal distribution, but it doesn't have to



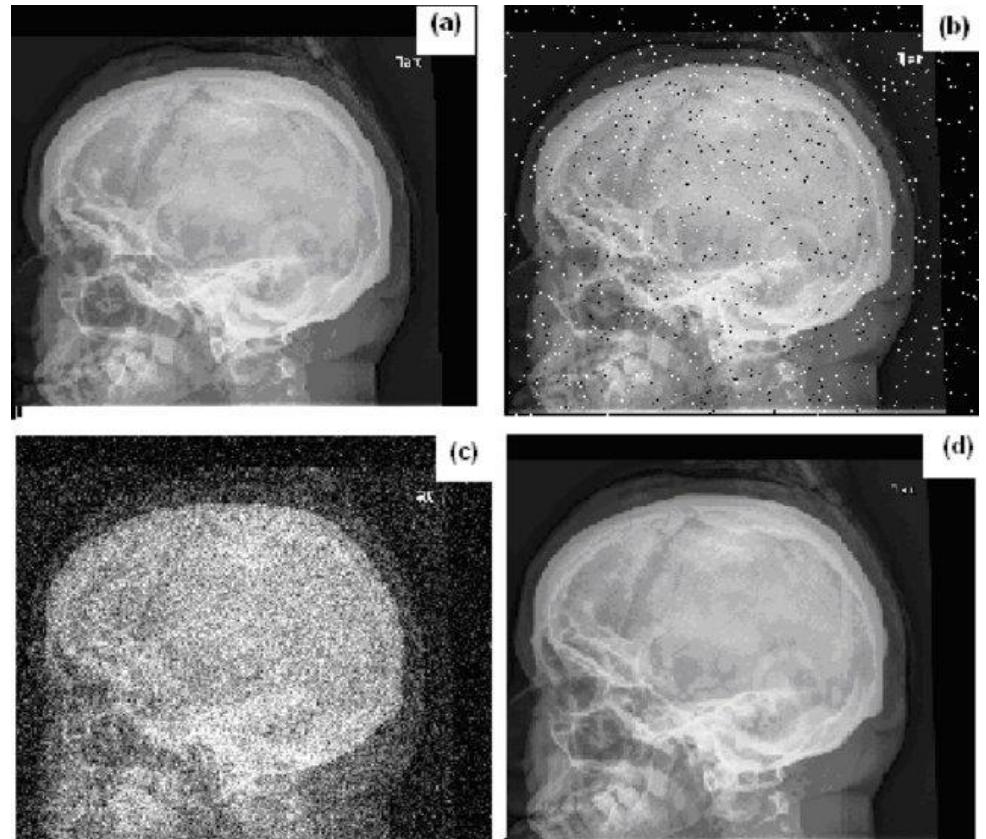
https://en.wikipedia.org/wiki/Colors_of_noise

Image Noise

Speckle noise - ultrasound



- (a) Original Image,
- (b) Salt and Pepper Noisy Image
- (c) Gaussian Noisy Image
- (d) Poisson Noisy Image



https://www.researchgate.net/publication/239735322_Computed_radiography_skull_image_enhancement_using_Wiener_filter

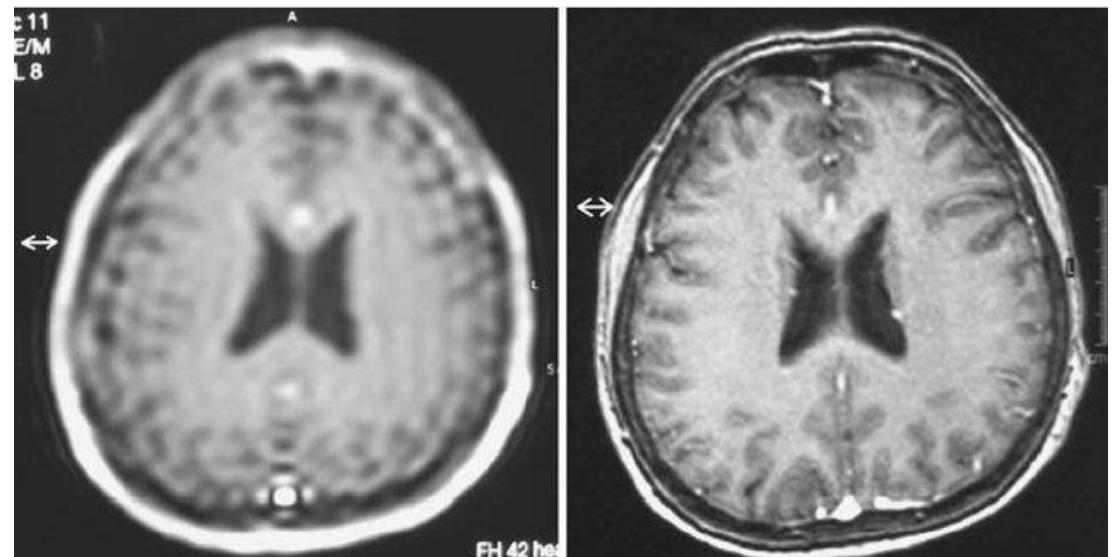
Artifact vs noise

Artifact

- appears to be a feature in the image
- Error in perception of information

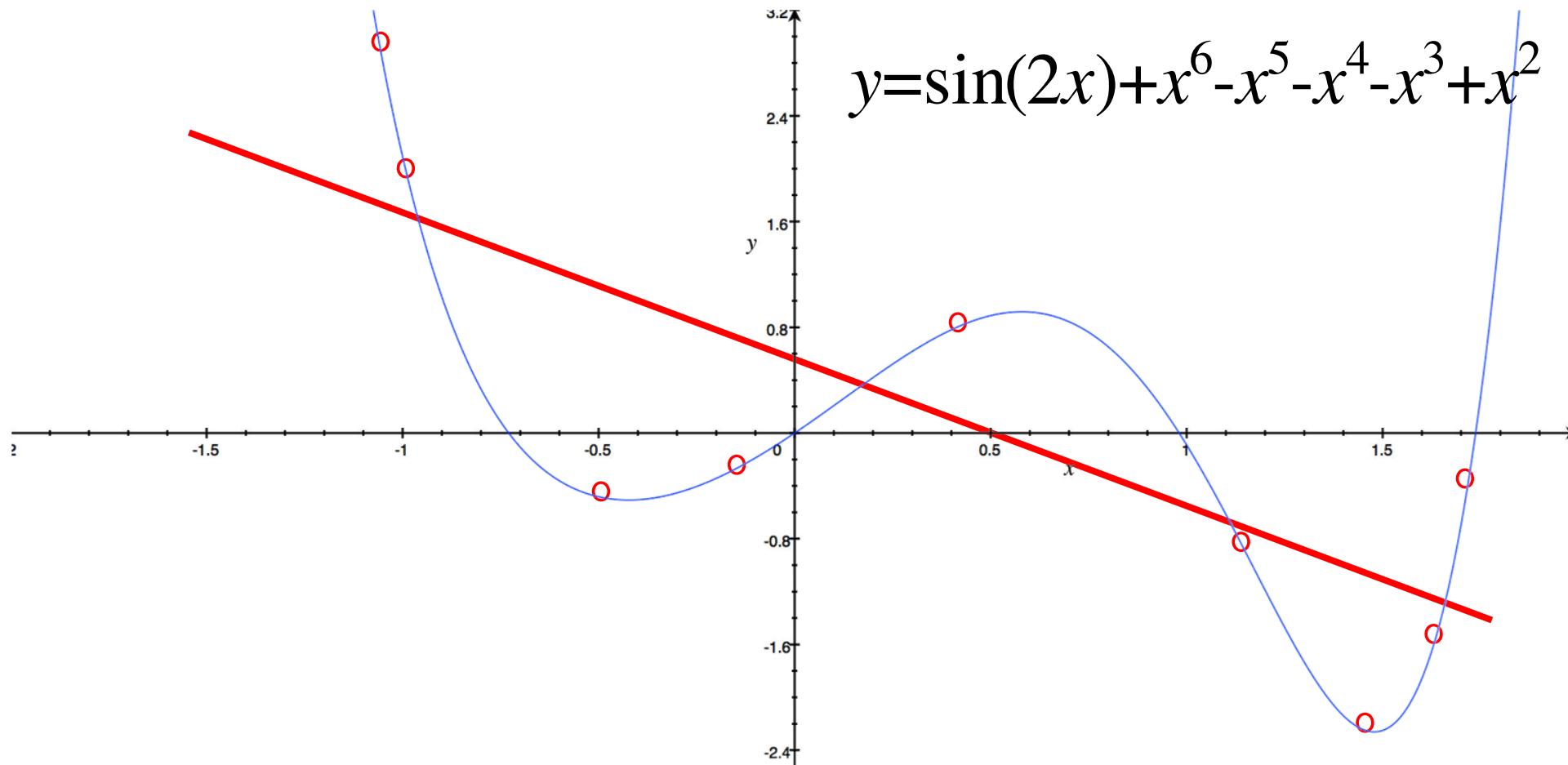
Noise

- obscures features
- generally random over frequencies or across image



https://www.researchgate.net/publication/6649032_Artifacts_in_body_MR_imaging_Their_appearance_and_how_to_eliminate_them/figures?lo=1

What is the Appropriate Mathematical Model: How to choose??



How do we prove our model is good?

Statistics?

- 1) Computer simulation
 - do we need statistics?

- 2) Mock System
 - do we need statistics?

- 3) The Biological System
 - so we need statistics?

Consider adding noise (colour?) or tolerance to see how model behaves

YES! How do we deal with multiple devices? Errors add in quadrature

YES! Absolutely essential- biological variability AND device tolerance + system noise demands the need for statistics

Types of Statistical Analysis

- Parametric vs. Non-parametric
- regression; minimizing least squares
- fit analysis/quality, e.g. R^2 , χ^2 tests
- test differences between a number of specified treatments

Errors

The difference between the observed and the true value

Illegitimate errors

- true mistakes
- ballpark errors, wrong equations etc.

Systematic errors

- difficult to spot
- faulty calibration, observer bias, observation parallax etc.

Random errors

- Experimental accuracy, precision

Error Propagation

Also known as Propagation of Uncertainty

Effect the uncertainty of a variable has on the uncertainty of the function it is used in

Due to measurement errors/observation errors

Calculus used to calculate combined error from multiple components

Used to estimate true error/uncertainty in a system

Why is error propagation important?

Kersten Blunder

- Vigor space probe
- Inches to mm conversion was off (0.9mm)
- Probe completely missed Venus and was lost in space

Space Mountain Tokyo

- Roller-coaster axle broke mid ride due to an error converting from imperial to metric

Air Canada Gimli Glider

- Measurement error by pilot for fuel requirement for the flight
- Used up all fuel halfway through flight

Error in using a tape measure?

Human error

Instrument error

- Tick error
- Hook error



TAPE MEASURES TOLERANCE MM ACCORDING TO EU STANDARDS

	ACCURACY CLASS I	ACCURACY CLASS II	ACCURACY CLASS III
3 METRES	±0,4	±0,9	±1,8
5 METRES	±0,6	±1,3	±2,6
8 METRES	±0,9	±1,9	±3,8

Example: Dimensions of a box

Example: What is the volume of a box of dimensions L, W and H ?

Easy:

But how do the uncertainties in these measurements result in uncertainty of the final result, V_0 ?

Note: ignore higher order terms in expansion which is equivalent to neglecting the fact the partial derivatives are not constant over the ranges of L, W, and H. But if errors are large need to include 2nd partial derivatives and cross derivatives!

Example: Dimensions of a box

Assuming someone is building a 1m³ box of wood and assuming they know how to use a tape measure...

Error Propagation Equations

They look worse than they are

Look for ways to simplify

- Neglect insignificant terms that make negligible contributions to the final uncertainty
- Rule of thumb: ignore error terms that make final contributions that are less than 10% of the largest contribution.
- But take care with this as many small uncertainties can add up!

Understanding what can be ignored takes practice!

Error Analysis Continued

→ Consider: we want to determine a quantity x that is a function of at least 2 measured variables u and v

$$x = f(u, v, \dots)$$

Although not exact consider the most probable value for x is:

$$\bar{x} = f(\bar{u}, \bar{v}, \dots)$$

The uncertainty in x can be determined by considering the spread of values in x resulting from the individual measurements in u_i and v_i :

$$x_i = f(u_i, v_i, \dots)$$

Error Propagation Equation

In the limit of infinite measurements the mean of the distribution coincides with x with variance described by:

Recall volume calculation example, where V was a function of the deviations in the calculated dimensions, express deviations in x in terms of deviations in u and v :

Error Propagation Equation

Each partial derivative is evaluated with all the other variables fixed at their mean values.

Combine Equations:

$$\sigma_x^2 = \lim_{N \rightarrow \infty} \left[\frac{1}{N} \sum (x_i - \bar{x})^2 \right]$$

$$x_i - \bar{x} \approx (u_i - \bar{u}) \left(\frac{\partial x}{\partial u} \right) + (v - \bar{v}) \left(\frac{\partial x}{\partial v} \right) + \dots$$

Error Propagation Equation

This allows expression of variance for x in terms of variances for the variables u, v, \dots (i.e. whatever was measured)

$$\begin{aligned}\sigma_x^2 &\simeq \lim_{N \rightarrow \infty} \frac{1}{N} \sum \left[(u_i - \bar{u}) \left(\frac{\partial x}{\partial u} \right) + (v_i - \bar{v}) \left(\frac{\partial x}{\partial v} \right) + \dots \right]^2 \\ &\simeq \lim_{N \rightarrow \infty} \frac{1}{N} \sum \left[(u_i - \bar{u})^2 \left(\frac{\partial x}{\partial u} \right)^2 + (v_i - \bar{v})^2 \left(\frac{\partial x}{\partial v} \right)^2 \right. \\ &\quad \left. + 2(u_i - \bar{u})(v_i - \bar{v}) \left(\frac{\partial x}{\partial u} \right) \left(\frac{\partial x}{\partial v} \right) + \dots \right]\end{aligned}$$

Error Propagation Equation

The first 2 terms are simple:

$$\sigma_u^2 = \lim_{N \rightarrow \infty} \left[\frac{1}{N} \sum (u_i - \bar{u}_i)^2 \right] \quad \sigma_v^2 = \lim_{N \rightarrow \infty} \left[\frac{1}{N} \sum (v_i - \bar{v}_i)^2 \right]$$

The 3rd term may be replaced by a covariance term:

$$\sigma_{uv}^2 \equiv \lim_{N \rightarrow \infty} \left[\frac{1}{N} \sum [(u_i - \bar{u})(v_i - \bar{v})] \right]$$

Error Propagation Equation

Substituting 3 terms back into original equation:

$$\sigma_x^2 \approx \sigma_u^2 \left(\frac{\partial x}{\partial u} \right)^2 + \sigma_v^2 \left(\frac{\partial x}{\partial v} \right)^2 + \dots + 2\sigma_{uv}^2 \left(\frac{\partial x}{\partial u} \right) \left(\frac{\partial x}{\partial v} \right) + \dots$$

- the first 2 terms are the averages of the squares of deviations weighted by the squares of the partial derivatives

Error Propagation Equation

$$\sigma_x^2 \approx \sigma_u^2 \left(\frac{\partial x}{\partial u} \right)^2 + \sigma_v^2 \left(\frac{\partial x}{\partial v} \right)^2 + \dots + 2\sigma_{uv}^2 \left(\frac{\partial x}{\partial u} \right) \left(\frac{\partial x}{\partial v} \right) + \dots$$

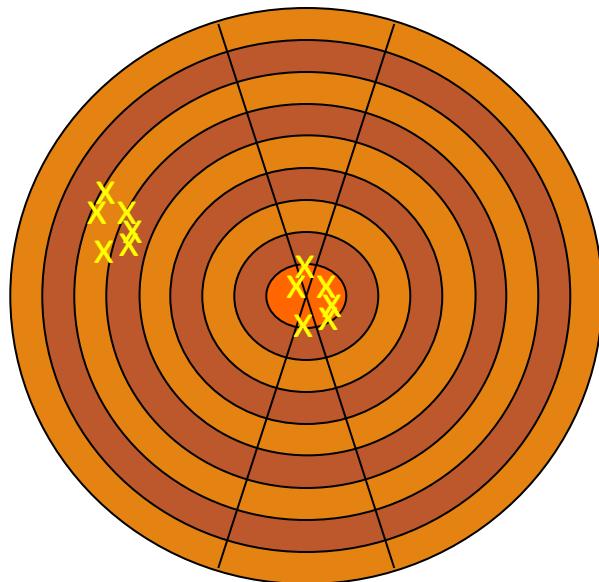
- If fluctuations in u and v are uncorrelated, then one would expect equal distribution of +ve and -ve values for this term and hence it would vanish with increased numbers of randomly selected observations. The equation then reduces to:

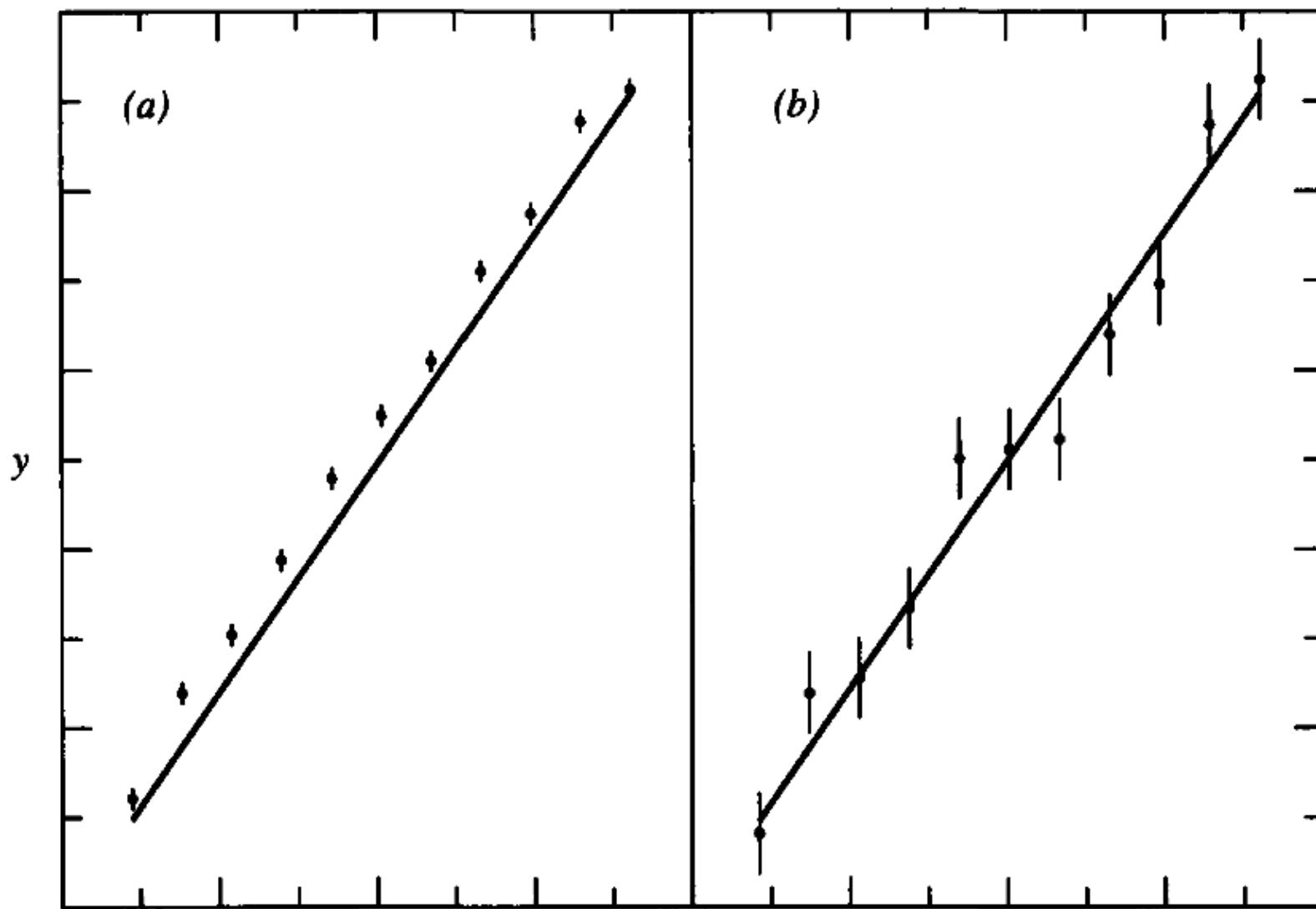
$$\sigma_x^2 \approx \sigma_u^2 \left(\frac{\partial x}{\partial u} \right)^2 + \sigma_v^2 \left(\frac{\partial x}{\partial v} \right)^2 + \dots$$

Accuracy versus Precision

Accuracy: how close to the true value did we get.

Precision: how exact are the measurements
repeatability





Precise but inaccurate

“TRUE” = straight line

Accurate but imprecise

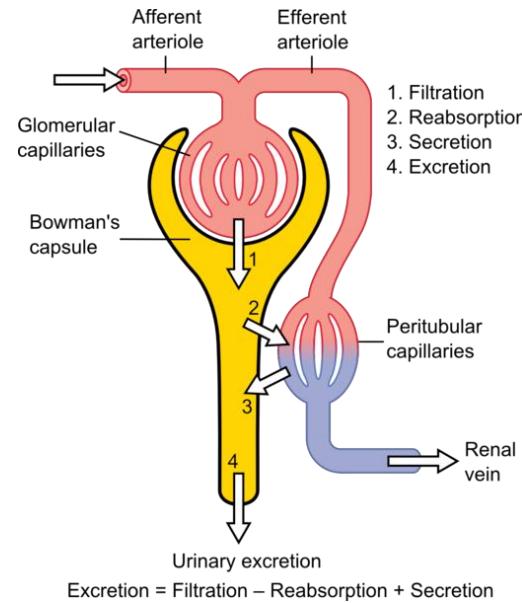
x

Is this Important to Know for Biological Modelling??

Consider measuring Glomerular Filtration Rate (GFR) an indication of renal function

$$GFR = 141 \cdot \min(SCr/k, 1)^a \cdot \max(SCr/k, 1)^{-1.209} \cdot 0.993^{Age}$$

[1.018] ↑ [1.159]



If female If African

SCr = serum creatinine (mg/dL)

k = 0.7 (females) or 0.9 (males)

a = -0.329 (females) or -0.411 (males)

min = minimum of SCr/k or 1 (whichever is least)

max = maximum of SCr/k or 1 (whichever is maximum)



Integrated Biomedical
Engineering & Health
Sciences Program

IBEHS - 4QZ3
Modelling of Biological Systems

Lecture 2

TAYLOR DEVET MASC.

PHD. CANDIDATE BIOLOGICAL AND BIOMEDICAL ENGINEERING

MCGILL UNIVERSITY

SHRINERS HOSPITAL FOR CHILDREN

Todays Aims...



Statistics Review



Linear Regression



Experimental Design

Reminders

Quizzes

- Friday – Sunday

Group Project

- Let Noor and Andrew know your groups so we can sort remaining people into groups

Assignment 1 due October 3rd

Statistics Review

Math

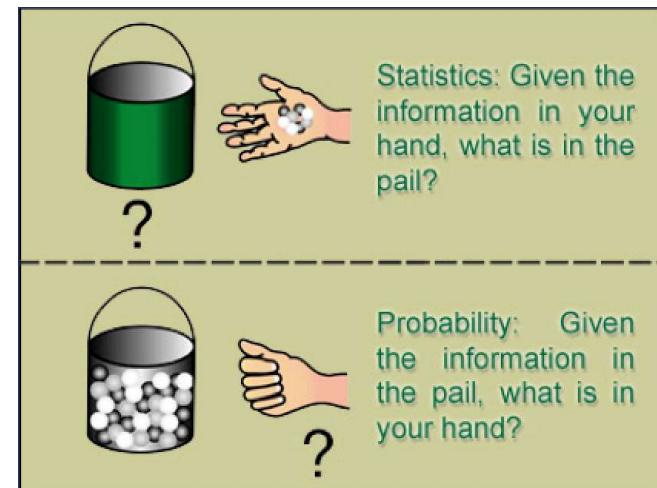
- study of space, change, structure, quantity
- Science or order, structure and relationships
- Given data, make model

Statistics

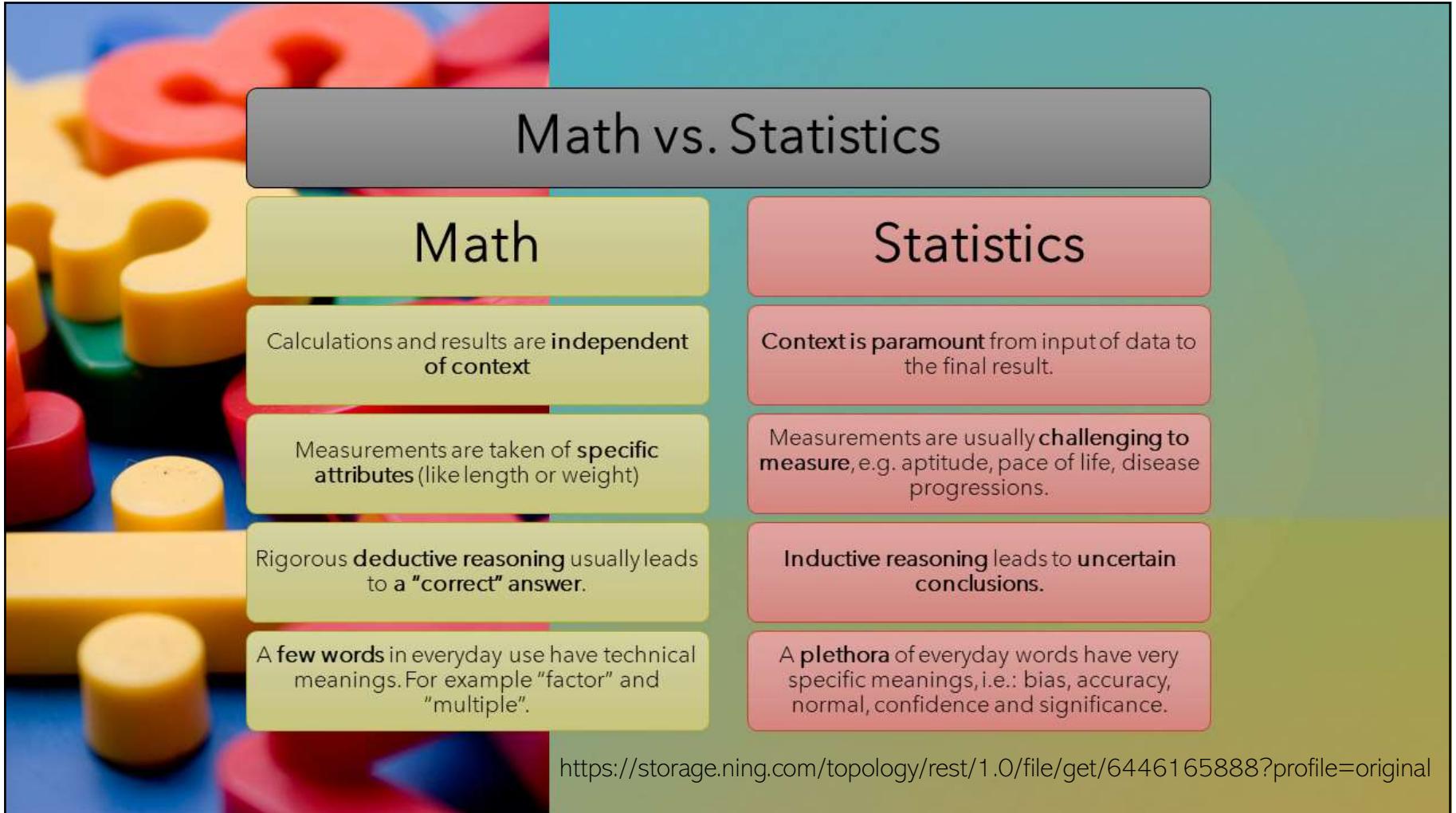
- Collection, analysis, explanation, interpretation of data
- Type of mathematical science
- Given data, predict model

Generally

- Determine population to look at OR model to study
- Collect data using survey or experiment
- Analyze data to look for significance



<https://mathprojects.com/tag/statistics/>



Math vs. Statistics

Math	Statistics
Calculations and results are independent of context	Context is paramount from input of data to the final result.
Measurements are taken of specific attributes (like length or weight)	Measurements are usually challenging to measure , e.g. aptitude, pace of life, disease progressions.
Rigorous deductive reasoning usually leads to a "correct" answer.	Inductive reasoning leads to uncertain conclusions .
A few words in everyday use have technical meanings. For example "factor" and "multiple".	A plethora of everyday words have very specific meanings, i.e.: bias, accuracy, normal, confidence and significance.

<https://storage.ning.com/topology/rest/1.0/file/get/6446165888?profile=original>

Branches of Statistics

Descriptive Statistics

- Summarize data using standard deviation, mean, median, mode etc
- Look at shape of data, skewness, kurtosis etc
- Focuses on learning about a sample rather than population
- Generally used for non parametric data
- Take large amount of data and describe it using parameters

Inferential Statistics

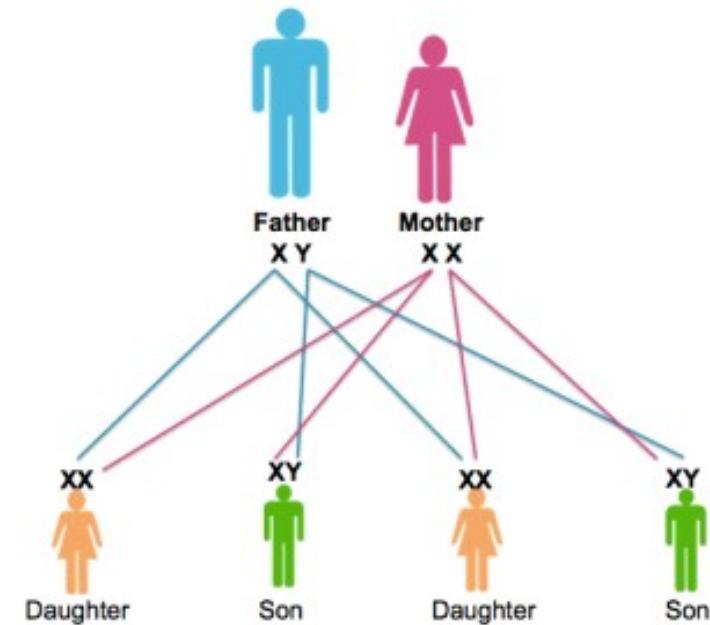
- Make conclusions about data that have random variance
- Uses data to make predictions
- I.e. use sample mean to make inferences about population mean
- Hypothesis testing to answer research question

Probability

Branch of mathematics

Math to describe how likely something is to happen

How probable it is that a statement is true



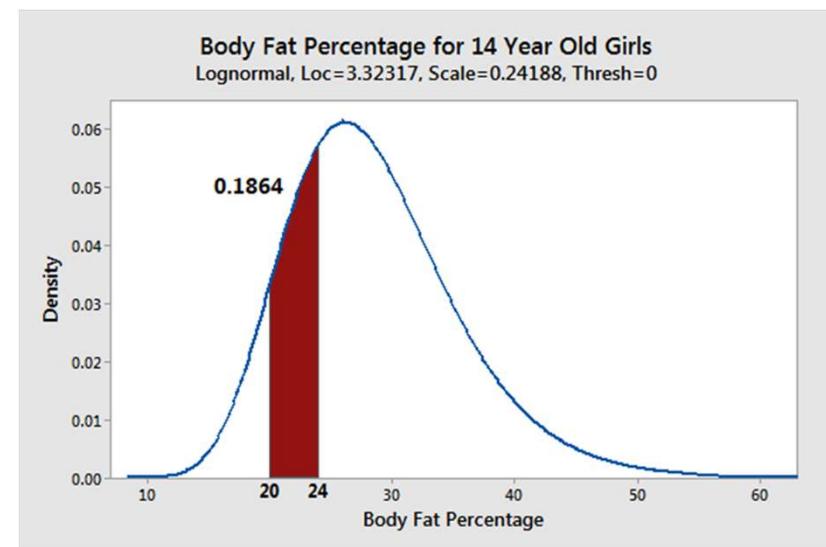
<https://allinonehighschool.com/using-punnett-squares-to-predict-offspring/>

Probability Distributions

Display the likelihood of obtaining a value given all the possibilities of a random variable

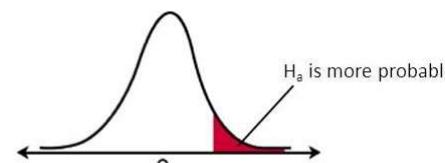
$P(x)$

- function that shows the likelihood that a random variable will be the specific value of x

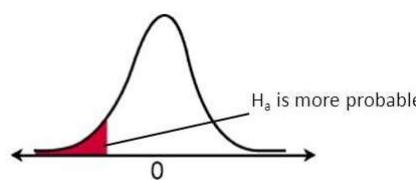


Hypothesis Testing

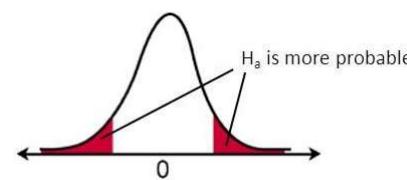
- Hypothesis tests look at 2 mutually exclusive statements regarding the population and determine which is more true
- Null Hypothesis (H_0)
 - No difference or effect
 - Accepting this leads to no change
- Alternate Hypothesis (H_a)
 - There is some difference or effect
 - Accepting this leads to a change
- Critical region
 - Region of value that corresponds to rejection of H_0



Right-tail test
 $H_a: \mu > \text{value}$



Left-tail test
 $H_a: \mu < \text{value}$



Two-tail test
 $H_a: \mu \neq \text{value}$

<https://towardsdatascience.com/everything-you-need-to-know-about-hypothesis-testing-part-i-4de9abebc8a>

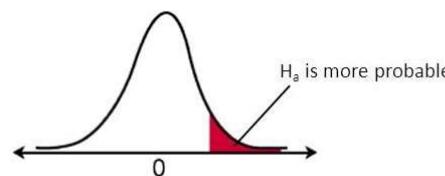
Hypothesis Testing

One tailed Test

- Critical area is one sided
- Critical area is either greater or less than critical value, not both
- If sample falls into area, H_a is accepted

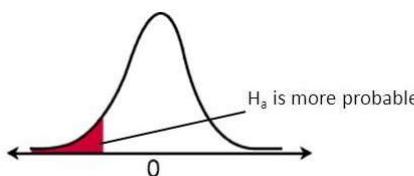
Two Tailed Test

- Critical area is two sided
- Critical area is greater than or less than critical values
- If sample falls into either of the critical areas, H_a is accepted



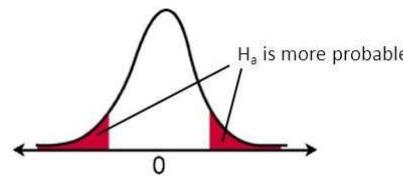
Right-tail test

$$H_a: \mu > \text{value}$$



Left-tail test

$$H_a: \mu < \text{value}$$



Two-tail test

$$H_a: \mu \neq \text{value}$$

<https://towardsdatascience.com/everything-you-need-to-know-about-hypothesis-testing-part-i-4de9abebc8a>

Steps in Hypothesis Testing

1. Formulate Hypothesis
2. Select Test type
3. Pick significance level
4. Collect data
5. Determine Critical value of test statistic
6. Determine if Test statistic falls into rejection region
7. Reject or don't reject H_0

Test Statistic

A measure of how close the sample comes to the null hypothesis

Gives information relevant to deciding if H_0 should be rejected

Each distribution uses its own test statistic

- Z test – Z Statistic
- T-test – t-statistic
- ANOVA – F Statistic
- Chi-square – Chi-Square statistic

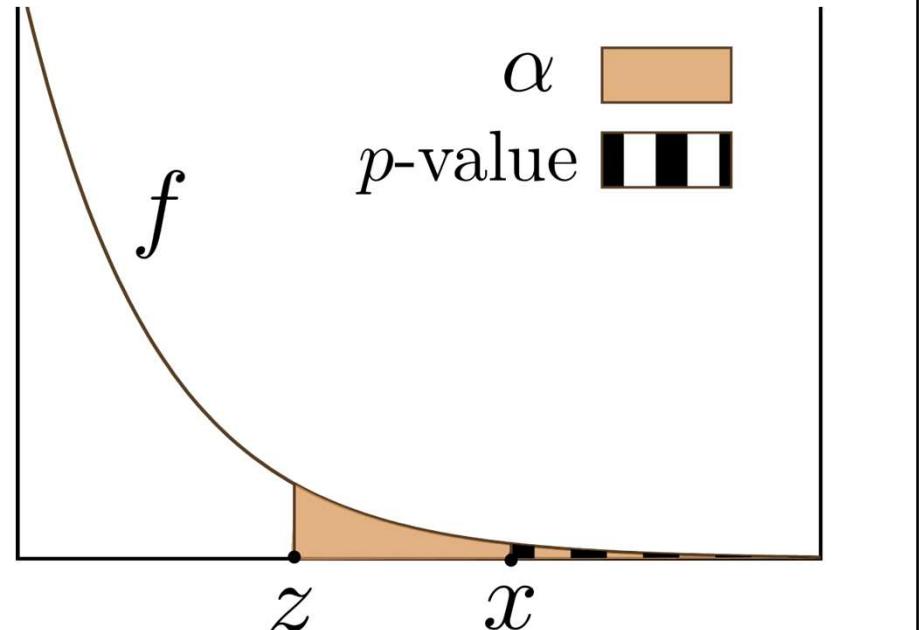
P Value

The probability of getting the result *at least* as extreme as the observed results

Assumes the null hypothesis is correct

Alpha value

- Significance Level
- If $p < \alpha$, accept the null hypothesis



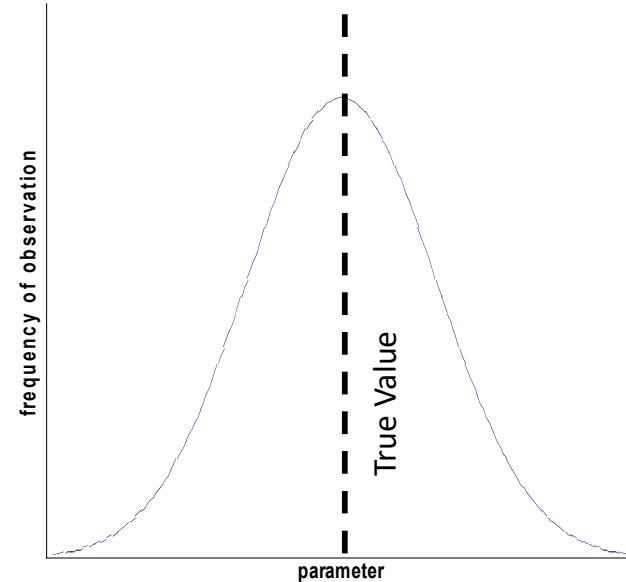
<https://www.investopedia.com/terms/p/p-value.asp>

Summarizing Data

Need to differentiate between population and sample

Population: infinite data, looks at entire population

Sample data: looks at a subsection of the population



Parameters to describe distribution

Gaussian distribution

- Mean (~true value)
- Variance (~variability)



Non-Gaussian distribution

- Median (~true value)
- at least 2 Percentiles (~variability)

Apr. 30, 1777 in Brunswick, Died: Feb 23, 1855 in Gottingen

Interpretation

In a balanced (gaussian) distribution

- Mean == true value
- Variance or standard deviation characterize uncertainty in the individual measurement
- ~68% of measurements are within 1σ
- ~95% of measurements are within 2σ

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right)$$

Likelihood/frequency of getting a value for a measurement

Mean

Population Mean:

$$\lim_{N \rightarrow \infty} \sum_{i=1}^N \frac{x_i}{N} = \mu$$

Sample Mean:

$$\bar{x} = (x_1 + x_2 + x_3 + \dots + x_n) / n$$

Arithmetic Mean:

$$\bar{x} = \frac{1}{n} \cdot \sum_{1}^n x_n$$

AKA: average, centroid, $\langle x \rangle$, etc.

Measurement of Variability

- variance and standard deviation

Population variance:

$$\begin{aligned}\sigma^2 &= \lim_{n \rightarrow \infty} \sum_{i=1}^N \frac{(x_i - \mu)^2}{N} \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^N \frac{x_i^2}{N} - \left[\lim_{n \rightarrow \infty} \sum_{i=1}^N \frac{x_i}{N} \right]^2\end{aligned}$$

Population Standard Deviation:

$$\sigma = \sqrt{\sigma^2}$$

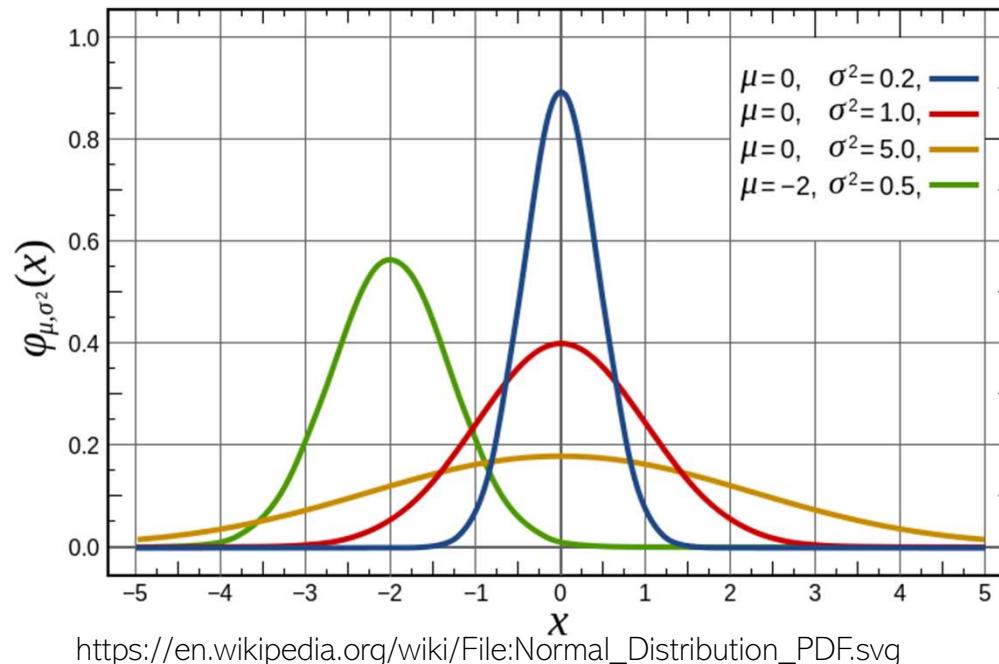
Sample Variance:

$$s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}$$

Sample Standard deviation

$$s = \sqrt{s^2}$$

Shapes of Normal Curves

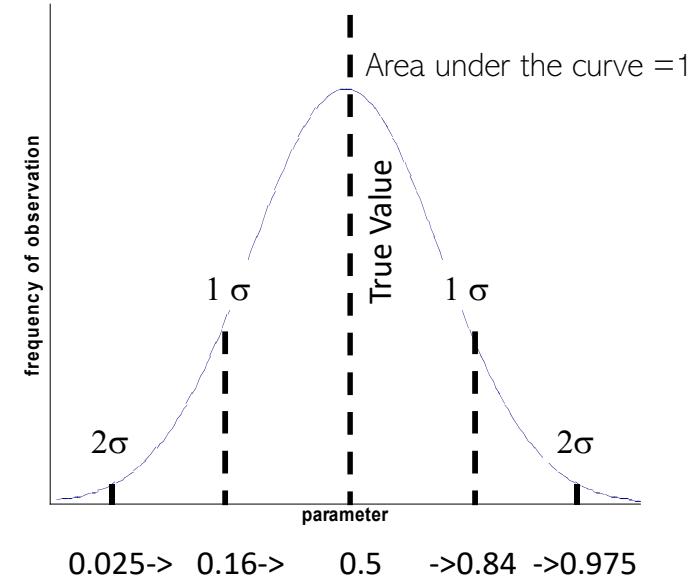
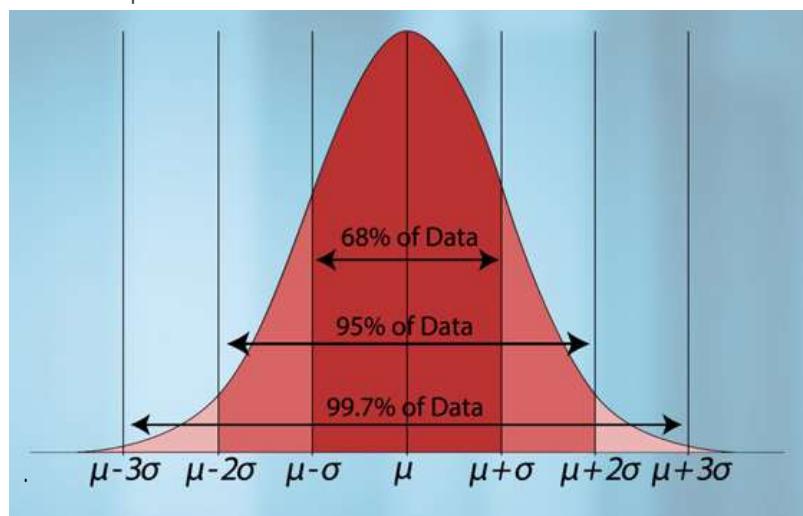


https://en.wikipedia.org/wiki/File:Normal_Distribution_PDF.svg

Z-Distribution vs Normal Distribution

Most test assume/require the data to be normal distributed

Z is a special case with mean=0 and SD = 1



Estimating Population Parameters

e.g if $N < \infty$

Sampling n measurements

estimating variance s^2 and mean \bar{x} from a limited number of samples

$$s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}$$

OR:

$$s^2 = \frac{1}{n-1} \cdot \left[\sum_{i=1}^n x_i^2 - \frac{1}{n} \cdot \left(\sum_{i=1}^n x_i \right)^2 \right]$$

Central Limit Theorem

If we were able to measure repeated \bar{x} → then we would approach μ

(i.e. mean of all experiments is equal to the population mean)

AND:

Therefore, the larger we can make our sample the closer we approach the population mean

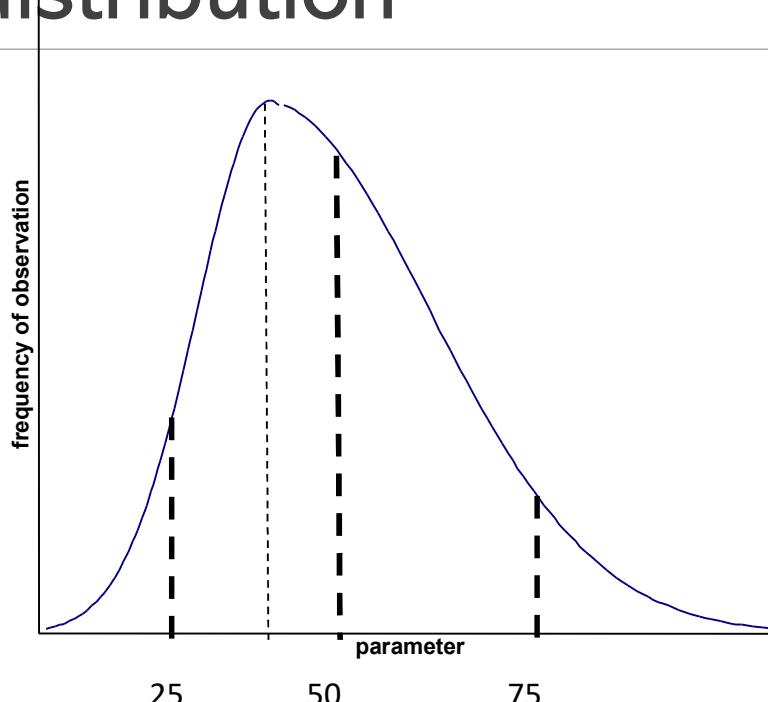
Measure of variability of a sample

Sample variability is described by the standard error or SEM

$$SEM = \frac{\sigma}{\sqrt{n}}$$

- No matter what the initial distribution of x_i are, as n is getting larger the distribution of \bar{x}_i will approach a normal distribution
- The distribution of $\sum_{i=1}^n \bar{x}_i$ will approach μ

Skewed distribution



Skewness

- measure of the asymmetry of data around the sample mean.
- If skewness is -ve, data are spread out more to the left of the mean than to the right.
- If skewness is +ve, data are spread out more to the right.
- The skewness of the normal distribution (or any perfectly symmetric distribution) is zero.
- (AKA 3rd moment about the mean)

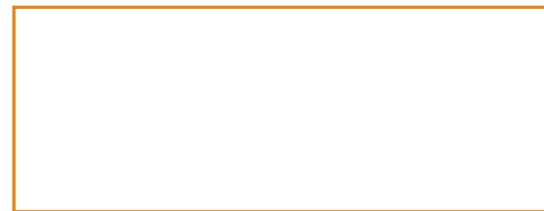
$$m_3 = \frac{\sum(X - \bar{X})^3}{n} \quad m_2 = \frac{\sum(X - \bar{X})^2}{n}$$



Kurtosis

- a measure of how outlier-prone a distribution is.
- The kurtosis of the normal distribution is 3.
- Distributions that are more outlier-prone than the normal distribution have kurtosis greater than 3
- distributions that are less outlier-prone have kurtosis less than 3.
- (AKA 4th order moment about the mean)

$$m_4 = \frac{\sum(X - \bar{X})^4}{n}$$



- sometimes kurtosis-3 is presented so the distribution is around 0.

Example

1.1650

0.6268 $m_4 = 0.3131$

0.0751 $m_3 = -0.688$

0.3516 $m_2 = 0.3802$

-0.6965

Tests for Normality

1. Jarque-Bera

- evaluates the hypothesis that x has a normal distribution with unspecified mean and variance
- vs the alternative that x does not have a normal distribution.
- based on sample skewness (s) and kurtosis (k) of n samples of x .
- tests whether the sample skewness and kurtosis are unusually different than their expected values
- should not be used with small samples.

2. Lilliefors

- useful for smaller samples
- similar to the Kolmogorov-Smirnov test
- Looks at Cumulative Distribution function – the probability that x is $\leq \bar{x}$

$$JB = \frac{n}{6} \left(s^2 + \frac{(k-3)}{4} \right)$$

Other distributions

Not all data is “Normal” (i.e. Gaussian distribution)

Binomial

- head/tail coin flipping

Lorentzian

- resonance in NMR

Poisson

- radioactive decay

Normality should always be assessed.

- histogram analysis
- Kurtosis - distribution has longer tails than normal
- Skewness - data not distributed evenly about a mean

Binomial Distribution

- Parameters n and p that represent a boolean question
- Probability of a number of “yes” in a row
- Is often drawn with a continuous curve but is discrete

The probability $P_B(x;n,p)$ for observing x of n items to be in the state with probability p is given by the binomial distribution:

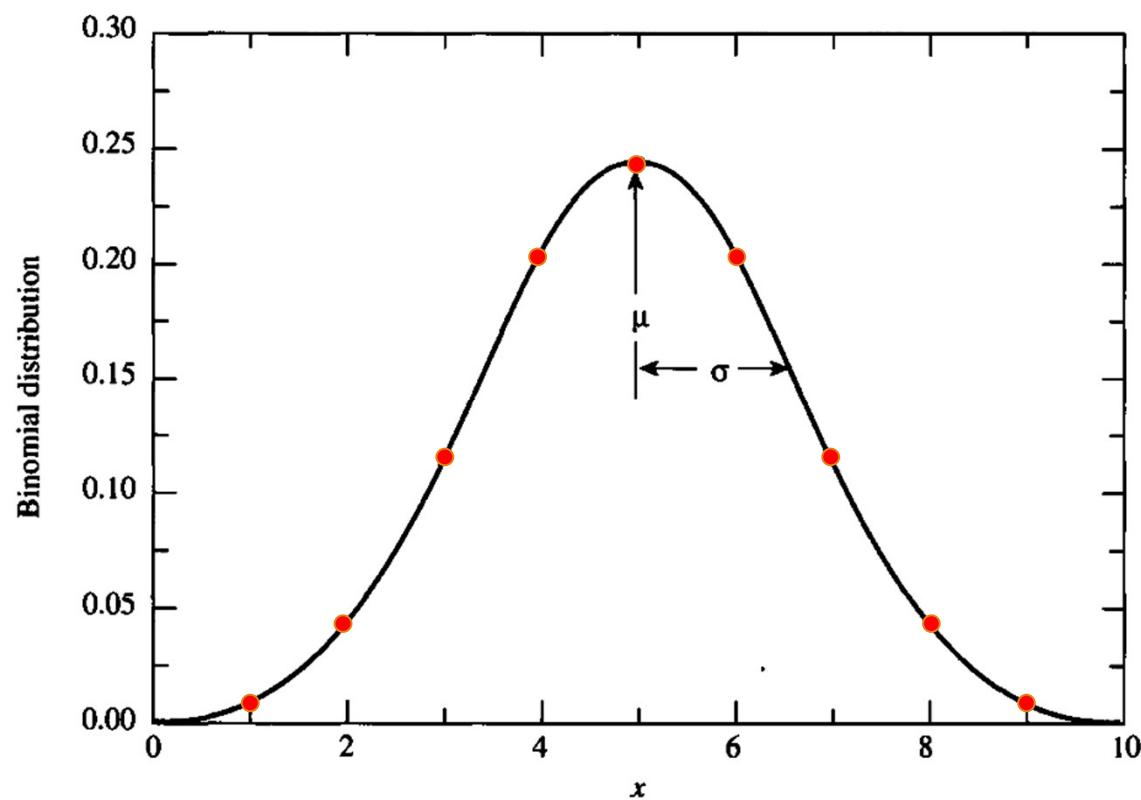
$$P_x = \sum_{k=0}^n \binom{n}{x} p^x q^{n-x}$$

Binomial Distribution

$$P_B(x; n, p) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$\mu = \sum_{x=0}^n \left[x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \right] = np$$

$$\sigma^2 = \sum_{x=0}^n \left[(x - \mu)^2 \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \right] = np(1-p)$$



Binomial Distribution for $\mu=5.0$ and $p=0.5$. The curve is shown as continuous but in reality the function is only defined as the discrete points (red dots)

Example

An engineer working at a particle accelerator makes preliminary measurements of the angular distribution of K mesons scattered from a $H_2(l)$ target. They know there should be equal numbers of particles scattered forwards and backwards in the centre-of-mass system of particles. She measures 1000 interactions and finds 472 scatter forwards and 528 backwards. What uncertainty should be quoted?

For uncertainty use Standard Deviation:

Poisson Distribution

Probability of observing x events in a set period of time t if events occur with a known constant mean and are independent of previous events

Also a discrete distribution but is represented by a continuous curve

$$P_P(x; \mu) = \frac{\mu^x}{x!} e^{-\mu}$$

μ = expected value of x , positive real number, mean number of events over time t

x = number of occurrences

Poisson Distribution

$$P_p(x; \mu) = \frac{\mu^x}{x!} e^{-\mu}$$

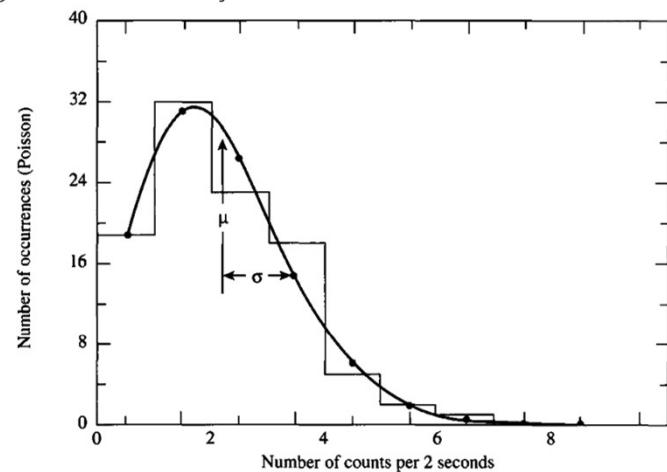
$$\langle x \rangle = \sum_{x=0}^{\infty} \left(x \frac{\mu^x}{x!} e^{-\mu} \right) = \mu e^{-\mu} \sum_{x=1}^{\infty} \frac{\mu^{x-1}}{(x-1)!} = \mu e^{-\mu} \sum_{y=0}^{\infty} \frac{\mu^y}{y!} = \mu$$

$$\sigma^2 = \langle (x - \mu)^2 \rangle = \sum_{x=0}^{\infty} \left[(x - \mu)^2 \frac{\mu^x}{x!} e^{-\mu} \right] = \mu$$

Therefore, the standard deviation, σ is equal to the square root of the mean, m and the Poisson distribution has only a single parameter, m .

Example

In an experiment to determine mean life of radioactive isotopes of silver, a grad student detected background counts from cosmic rays. Values were recorded as counts on their detector for a series of 100, 2-second intervals and the mean number of counts was found to be 1.69 per interval. Using the mean they estimated the standard deviation to be:



Notes

- 1) Poisson is defined as discrete points but here shown as a continuous curve
- 2) As m increases the symmetry of the Poisson distribution increases until it becomes indistinguishable from a Gaussian

Lorentzian Distribution

Also known as Cauchy distribution

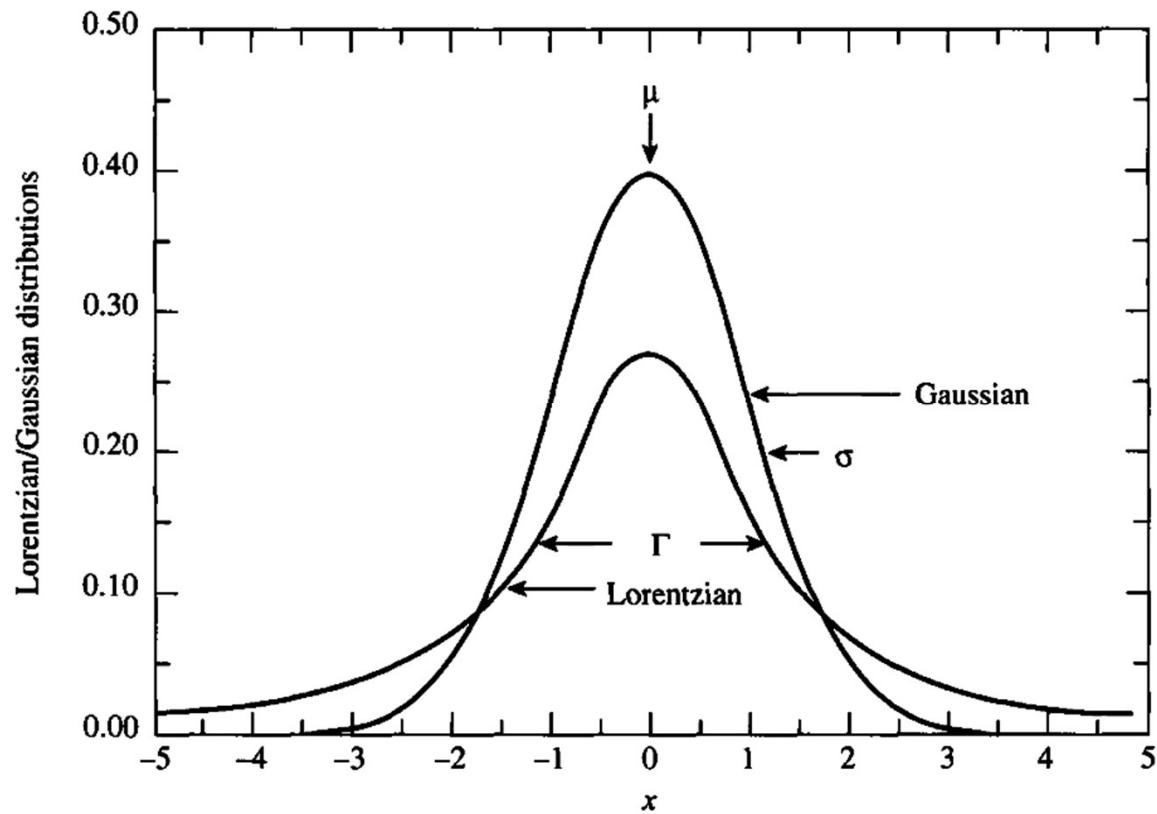
Continuous distribution

Lorentzian Probability Density Function
 $P_L(x; \mu, \Gamma)$

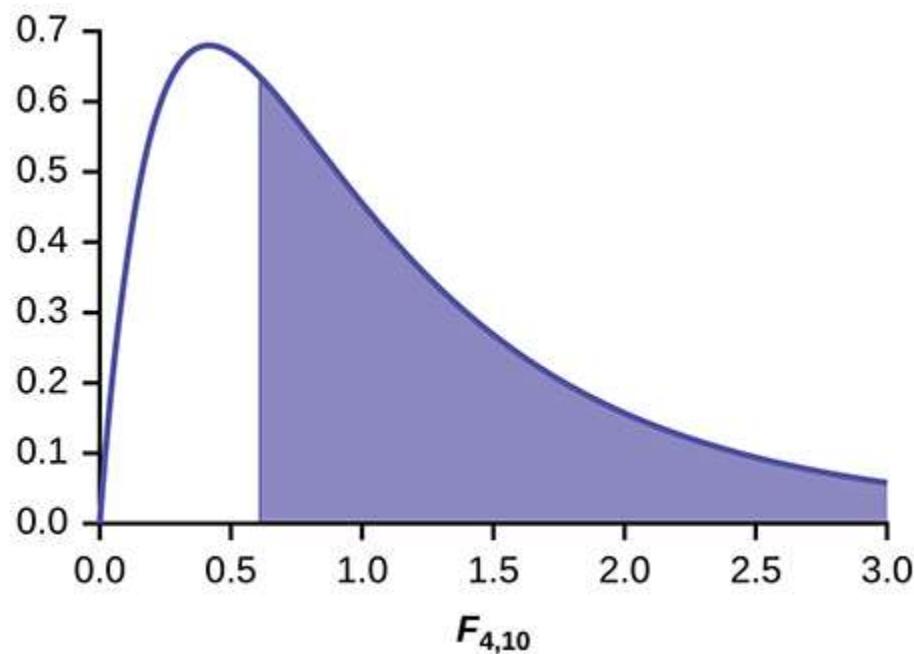
- appropriate for data exhibiting resonant behavior
- mean μ , and full width at half maximum (FWHM), Γ
- similar to Gaussian, but doesn't diminish to zero as fast
- undefined mean and variance

$$p_L(x; \mu, \Gamma) = \frac{1}{\pi} \frac{\Gamma/2}{(x - \mu)^2 + (\Gamma/2)^2}$$

$$\sigma^2 = \langle (x - \mu)^2 \rangle = \frac{1}{\pi} \frac{\Gamma^2}{4} \int_{-\infty}^{\infty} \frac{z^2}{1 + z^2} dz$$



The F -Distribution



The *F*-Distribution

The F distribution is the ratio of two variance estimates:

$$F = \frac{s_1^2}{s_2^2} = \frac{est.\sigma_1^2}{est.\sigma_2^2}$$

Also the ratio of two chi-squares, each divided by its degrees of freedom:

$$F = \frac{\chi_{(v_1)}^2 / v_1}{\chi_{(v_2)}^2 / v_2}$$

- $v_2 > v_1$, and $v_2 > 2$.

Then the mean of the F distribution (expected value) = $v_2 / (v_2 - 2)$

F -Distribution (pt.2)

F depends on v_1 and v_2 (df_1 and df_2).

These dictate the shape of F . Range is 0 to infinity.

F tables show critical values for df in the numerator and df in the denominator.

F tables are 1-tailed (2-tailed are atypical)

A continuous distribution

ANOVA

ANalysis Of VAriance

Statistical model that estimates variance thorough differences in means

Developed by Sir Ronald Fisher

- British Statistician & Geneticist

Source of Variation	df	Sum of Squares (SS)	Mean Sum of Squares (MSS)	F-test	p-value
Treatment	k-1	SSTr	$MSTr=SSTr/(k-1)$	$F=MSTr/MSE$	
Error	N-k	SSE	$MSE=SSE/(N-k)$		
Total	N-1	SSTo			

<https://courses.lumenlearning.com/suny-natural-resources-biometrics/chapter/chapter-5-one-way-analysis-of-variance/>

Summary of Major Distributions

- 1) To understand error you MUST understand the distributions you are dealing with
- 2) But what about multiple measures:
 - Error Propagation

Experimental Design

- Proper application of biostatistics involves experimental design
- Experimental design is done before testing
- Proper application of experimental design and statistical analysis can save a lot of:
 - Time
 - Resources
 - lives (potentially), etc.

Hypothesis Testing

- test a hypothesized value
- need some value based on past experience, claims of other people, dream your thesis advisor had, etc.
- new situation (e.g. new growth hormone) produces results that are no different, on average, from those results previously occurring.
 - Null Hypothesis

e.g. Weight training, with a healthy diet, produces a mean increase in muscle mass of 12.6kg, over 6 weeks. A new growth hormone, yulegosterol, increases this.

$H_0 = \mu = \mu_0 = 12.6$ (Null hypothesis, nothing different)

$H_A = \mu > \mu_0 = 12.6$ (Alternative, there is weight gain)

Hypothesis Testing cont

- this is a one-sided alternative (and one sided test)
- if we have no clue on what new hormone will really do

We can use a 2 sided alternative

$H_0 = \mu = \mu_0 = 12.6$ (Null hypothesis, nothing different)

$H_A = \mu \neq \mu_0 = 12.6$ (Alternative, \pm weight gain)

- Regardless of whichever, we still need some criterion for deciding how far away \bar{x} can be from μ_0 before we reject H_0 .
- choose a level of significance (e.g. $\alpha=0.05$). This is a level of probability that the test will fail.

The appropriate Test in this Example: Student's T-test

8 volunteers (degrees of freedom, df = 8-1=7)

mean weight gain = 20.2kg

standard deviation = 4.3

From Student's t-test table (8-1 = 7df; $\alpha=0.05$):

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

- reject H_0 if $t > 1.895$ (one sided)
- reject H_0 if $t > +2.365$ or $t < -2.365$ (two sided)
(i.e. reject if $|t| > 2.365$)

ASSUMES NORMALLY DISTRIBUTED DATA

- This should be tested

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
Z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Confidence Intervals

- what range of values are we confident that the measurements can take ?
- theory states that 95% of the time for a t value with df = 7:

$$-2.365 \leq t \leq +2.365$$

$$-2.365 \leq \left| \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \right| \leq +2.365$$

A little algebra.....

$$\left(\bar{x} - 2.365 \cdot \frac{s}{\sqrt{n}} \right) \leq \mu \leq \left(\bar{x} + 2.365 \cdot \frac{s}{\sqrt{n}} \right)$$

Test for Two Means

Calculation of Denominator depends on:

1. The 2 populations having common variance, σ^2
2. The 2 σ^2 s, or common σ^2 is known or estimated
3. If both samples are (or are not) the same size
4. Paired vs. Independent

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_{\bar{x}_1 - \bar{x}_2}}$$

The Choice of Rejection depends on:

1. The level of significance chosen (α)
2. The sample size (n)
3. The test required (i.e. 1 or 2 tailed)

Independent vs. Paired

Pairing:

- done prior to experiment on the basis of similar responses in the absence of treatment effects.
- e.g. comparing drug therapy in sets of identical twins
- - if members of a pair tend to be positively correlated an increase in the ability of the experiment to detect a small difference is possible.

Independent

- Compares the means between 2 groups

2 Means, Independent Samples, Equal Variances:

$H_0 : \mu_1 = \mu_2$ (Null Hypothesis)

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_{\bar{x}_1 - \bar{x}_2}}$$

Need weighted average of sample variances:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

Note df = $(n_1 - 1) + (n_2 - 1)$

Situation #1: $n_1 \neq n_2$

$$S_{\bar{x}_1 - \bar{x}_2} = \sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} = \sqrt{s^2 \left(\frac{n_1 + n_2}{n_1 n_2} \right)}$$

Situation #2: $n_1 = n_2$

$$S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{2s^2}{n}}$$

Comparing Paired Sample Means

- compute differences in pairs
- calculate average pair difference

$$s = \sqrt{\frac{\sum_{j=1}^n D_j^2 - \left(\sum_{j=1}^n D_j\right)^2 / n}{n-1}}$$

$$t = \frac{\bar{D}}{s/\sqrt{n}}$$

j = number of pairs

Also note here n = j

Note: df = j - 1

Testing the Hypothesis of Equality of Variances (homoscedasticity):

- up to now it is assumed that variances are equal, based on some pre-decided criterion.

How is this tested ?

Null hypothesis: $\sigma_1^2 = \sigma_2^2$

$$F_{\alpha,m,n} = \frac{s_{BIG}^2}{s_{SMALL}^2}$$

$$m-1 = df \text{ for } s_{big}^2$$

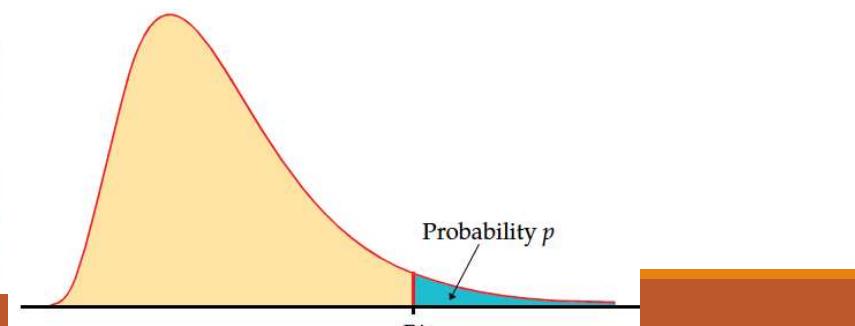
$$n-1 = df \text{ for } s_{small}^2$$

α = level of significance desired (e.g. 0.05 for 95% confidence)

F-distribution

		Degrees of freedom in the numerator										
		p	1	2	3	4	5	6	7	8	9	
			.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
		.050	.100	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
		.025	.050	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28
		.010	.025	4052.2	4999.5	5403.4	5624.6	5763.6	5859.0	5928.4	5981.1	6022.5
		.001	.010	405284	500000	540379	562500	576405	585937	592873	598144	602284
			.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
		.050	.100	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
		.025	.050	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
		.010	.025	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
		.001	.010	998.50	999.00	999.17	999.25	999.30	999.33	999.36	999.37	999.39
			.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
		.050	.100	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
		.025	.050	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
		.010	.025	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
		.001	.010	167.03	148.50	141.11	137.10	134.58	132.85	131.58	130.62	129.86
			.100	4.54	4.32	4.19	4.11					
		.050	.100	7.71	6.94	6.59	6.39					
		.025	.050	12.22	10.65	9.98	9.60					
		.010	.025	21.20	18.00	16.69	15.98					
		.001	.010	74.14	61.25	56.18	53.44					
			.100	4.06	3.78	3.62	3.52					
		.050	.100	6.61	5.79	5.41	5.19					
		.025	.050	10.01	8.43	7.76	7.39					
		.010	.025	16.26	13.27	12.06	11.39					
		.001	.010	47.18	37.12	33.20	31.09					

Degrees of freedom in the denominator



If 2 samples have unequal Variances

$$S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$t' = \frac{\bar{x}_1 - \bar{x}_2}{S_{\bar{x}_1 - \bar{x}_2}}$$

effective $df = \frac{(s_1^2/n_1 + s_2^2/n_2)}{[(s_1^2/n_1)^2/(n_1-1)] + [(s_2^2/n_2)^2/(n_2-1)]}$

X T.TEST function

Returns the probability that is associated with a Student's t-Test. Use T.TEST to determine whether two samples are likely to have come from the same two underlying populations that have the same mean.

Syntax

T.TEST(array1,array2,tails,type)

Argument	Description	Remarks
array1	The first data set.	<ul style="list-style-type: none">• None.
array2	The second data set.	<ul style="list-style-type: none">• None.
tails	Specifies the number of distribution tails.	<ul style="list-style-type: none">• If tails = 1, T.TEST uses the one-tailed distribution. If tails = 2, T.TEST uses the two-tailed distribution.• If tails is any value other than 1 or 2, this function returns the #NUM! error value.• If this argument is nonnumeric, this function returns the #VALUE! error value.• If this argument contains a decimal value, this function ignores the numbers to the right side of the decimal point.
type	The kind of t-Test to perform.	<ul style="list-style-type: none">• If type equals 1, T.TEST performs a paired test.• If type equals 2, T.TEST performs a two-sample equal variance (homoscedastic) test.• If type equals 3, T.TEST performs a two-sample unequal variance (heteroscedastic) test.• If this argument is nonnumeric, this function returns the #VALUE! error value.• If this argument contains a decimal value, this function ignores the numbers to the right side of the decimal point.

MATLAB

ttest

One-sample and paired-sample *t*-test

[expand all in page](#)

Syntax

<code>h = ttest(x)</code>	example
<code>h = ttest(x,y)</code> <code>h = ttest(x,y,Name,Value)</code>	example example
<code>h = ttest(x,m)</code> <code>h = ttest(x,m,Name,Value)</code>	example example
<code>[h,p] = ttest(__)</code> <code>[h,p,ci,stats] = ttest(__)</code>	example example

Description

`h = ttest(x)` returns a test decision for the null hypothesis that the data in `x` comes from a normal distribution with mean equal to zero and unknown variance, using the [one-sample *t*-test](#). The alternative hypothesis is that the population distribution does not have a mean equal to zero. The result `h` is 1 if the test rejects the null hypothesis at the 5% significance level, and 0 otherwise.

[example](#)

What about Multiple Comparisons?

- There are more powerful techniques
- t-test is not appropriate
- get compounding error

e.g.

- Analysis of Variance (ANOVA)
- comparisons of multiple treatments

Power, Sample Size, and the Detection of Differences

The error rate or significance level is chosen = α
(e.g. $\alpha = 0.05$)

TYPE I Error

- α - we make a mistake and falsely reject HO

TYPE II Error

- β - we make a mistake and falsely accept HO

		Decision		Data from a population for which:
		H ₀ is TRUE, H ₁ false	H ₀ is false, H ₁ TRUE	
		Accept H ₀ Reject H ₁		
Non-significant	Accept H ₀ Reject H ₁	<p><i>Correct Decision</i> Probability should be high. Symbol: $1 - \alpha$ = Confidence coefficient</p>		<p><i>Incorrect Decision</i> \rightarrow <u>Type II</u> error made Probability should be low. Symbol: β</p>
	Reject H ₀ Accept H ₁	<p><i>Incorrect Decision</i> \rightarrow <u>Type I</u> error made Probability should be low. Symbol: α (significance level)</p>		<p><i>Correct Decision</i> Probability should be low. Symbol: $1 - \beta$ = power</p>

Errors

- if we use $\alpha = 10\%$
 - there is a high tendency to conclude that H_0 will be false when it is not
- if use $\alpha = 0.1\%$
 - then you are unlikely to erroneously state that H_0 is rejected.
 - Tests such as this are conservative and reliable
 - Also fairly unlikely to state that H_0 is rejected when in truth it should be accepted

	Value of α	
<i>Type of test:</i>	10% (liberal)	0.1% (cautious, conservative)
If H_0 is true	May well reject H_0	Unlikely to reject H_0
If H_0 is false	Good chance of rejecting H_0	Some chance of rejecting H_0
If H_0 is not rejected	Very little reason found to distrust H_0	Support for H_0 may not be impressive
If H_0 is rejected	Possibly over-hasty rejection of H_0	Very convincing evidence against H_0

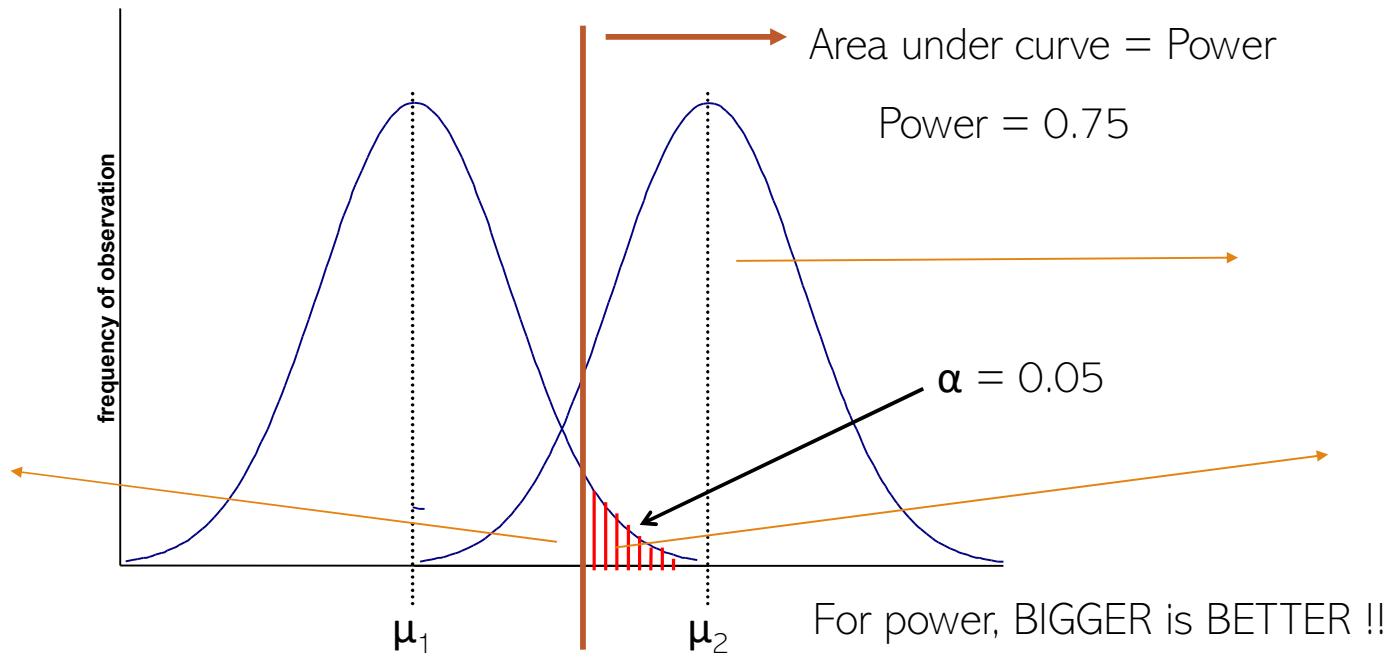
Power

- a test of significance should reject H_0 when it is really false
- the probability a test does this is the Power

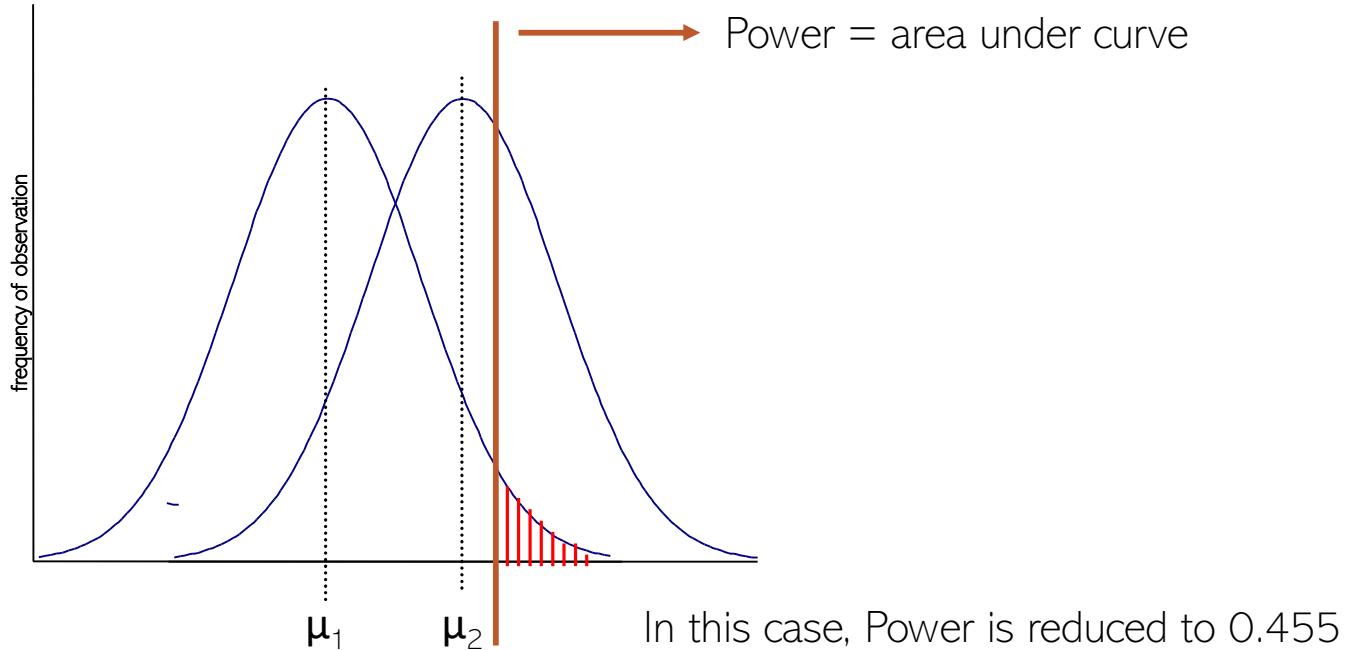
Power is a complex quantity depending on:

- chosen α
- variance σ
- number in sample, n
- difference in means (i.e. $\mu_1 - \mu_2$)

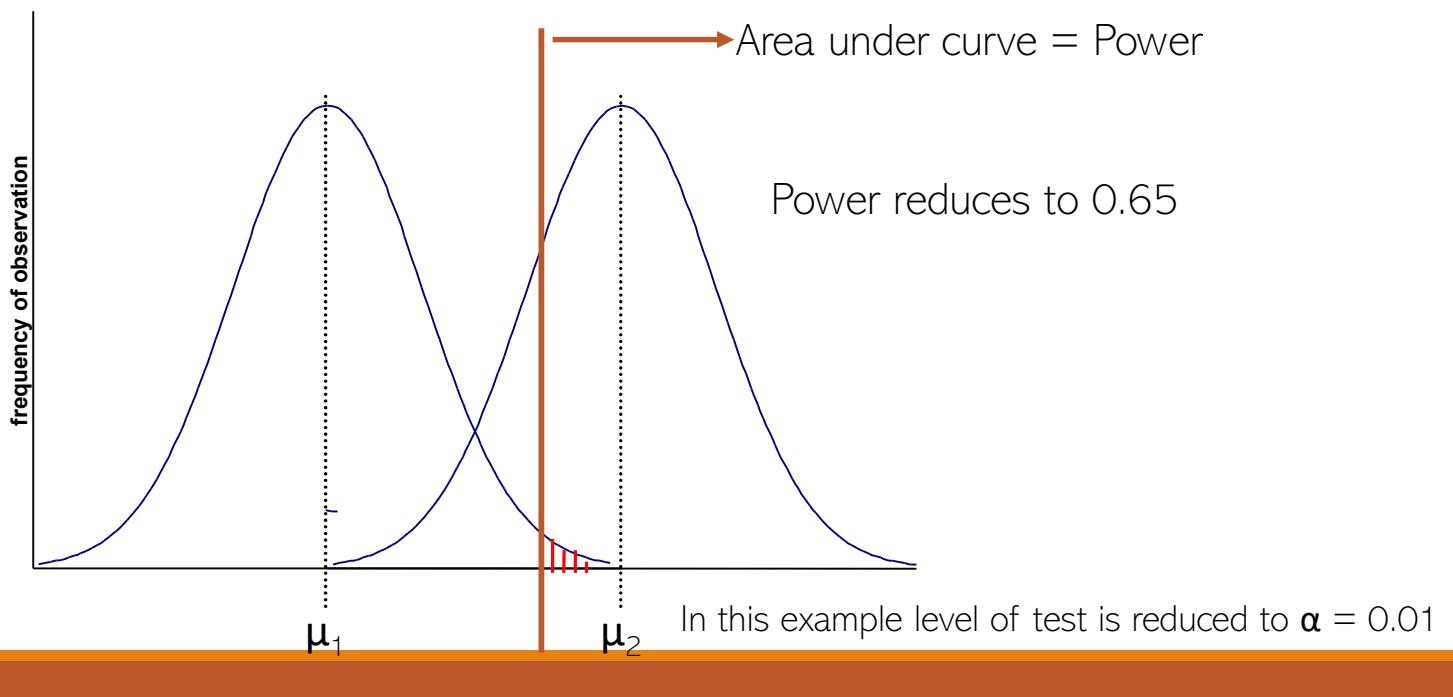
Graphical Representation of Power



Graphical Representation of Power



Graphical Representation of Power



- for a 1-tailed test, at level α , the power is the probability that the normal deviate

$$Z > (-\sqrt{n}/\sigma)(\mu_1 - \mu_2) + Z_{2\alpha}$$

$Z_{2\alpha} = 1.645$ for a 1-tailed test with $\alpha = 0.05$ (i.e. 5% significance level)

$Z_{2\alpha} = 2.326$ for a 1-tailed test with $\alpha = 0.01$ (i.e. 1% significance level)

The key factor in which power depends is:

$$\phi = \sqrt{n}(\mu_1 - \mu_2)/\sigma$$

(Single or paired samples)

Note: Use $\sqrt{n/2}$ if samples are independent !

Sample Power Calculations

$$\phi = \sqrt{n}(\mu_1 - \mu_2)/\sigma =$$

Level of Test	# tails	1.5	2	2.5	3	3.5
		0.44	0.64	0.80	0.91	0.97
0.05	1	0.44	0.64	0.80	0.91	0.97
	2	0.32	0.52	0.71	0.85	0.94
0.01	1	0.20	0.37	0.57	0.75	0.88
	2	0.14	0.22	0.47	0.61	0.82

example: Let's say there are 5 samples already measured ($n=10$) where the difference between means is, $\Delta\mu=18.2-16.6=1.6$, and the standard deviation is $\sigma=2.6$. What is the power ?

So, how many samples (n) are needed anyway?

One Quick Method.....

- 1) decide on the approximate desired power wanted for a specific value of $\mu_1 - \mu_2$
- 2) use table (previous slide) to determine the approximate value needed of ϕ for the intended level of significance and nature(i.e. 1 or 2 tailed) of the test.
- 3) Use formula to solve for n:

$$n = \left[\frac{\phi\sigma}{(\mu_1 - \mu_2)} \right]^2$$

Paired, or single samples

$$n = 2 \cdot \left[\frac{\phi\sigma}{(\mu_1 - \mu_2)} \right]^2$$

Independent samples

Type I Error Rate and Multiple T-tests

- consider no [true] difference between 2 populations
- by random chance alone there is a $100*\alpha\%$ chance of declaring an [incorrect] difference between the two populations.
- error compounded when multiple t-tests are carried out.

if k independent t-tests are performed with α level of significance, then the probability of observing no significant (X) differences is:

Type I Error Rate and Multiple T-tests

- the probability of observing at least one significant difference (when none exist) is:

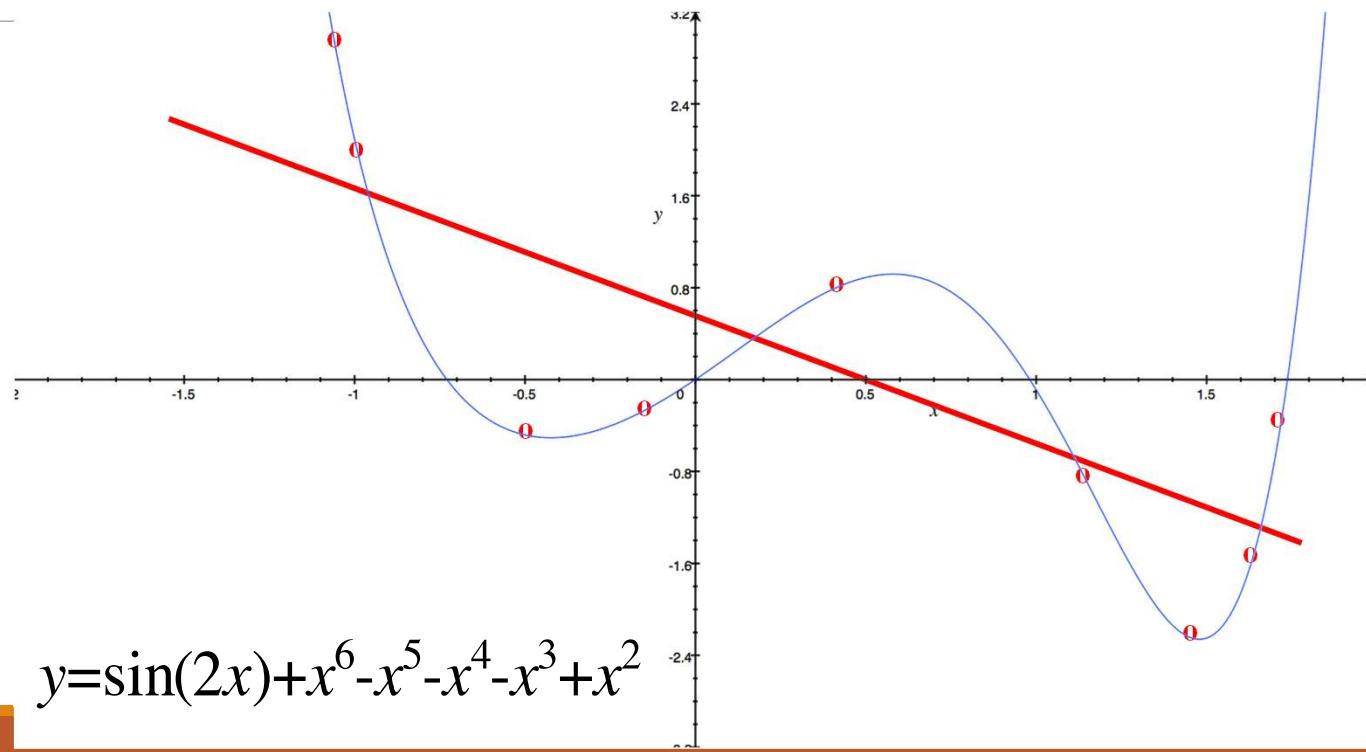
Thus as $k \uparrow$ the probability of a Type I error \uparrow .

e.g. if 10 independent t- tests are carried out, the probability of declaring at least one significant difference (even though there are none) is:

Table 1: Type I error rate for k independent tests with a significance level of Alpha

k	Alpha			
	0.1	0.05	0.01	0.001
1	0.1	0.05	0.01	0.001
2	0.19	0.098	0.02	0.002
3	0.271	0.143	0.03	0.003
4	0.344	0.185	0.039	0.004
5	0.41	0.226	0.049	0.005
6	0.469	0.265	0.059	0.006
7	0.522	0.302	0.068	0.007
8	0.57	0.337	0.077	0.008
9	0.613	0.37	0.086	0.009
10	0.651	0.401	0.096	0.01

What is the Appropriate Mathematical Model: How to choose?



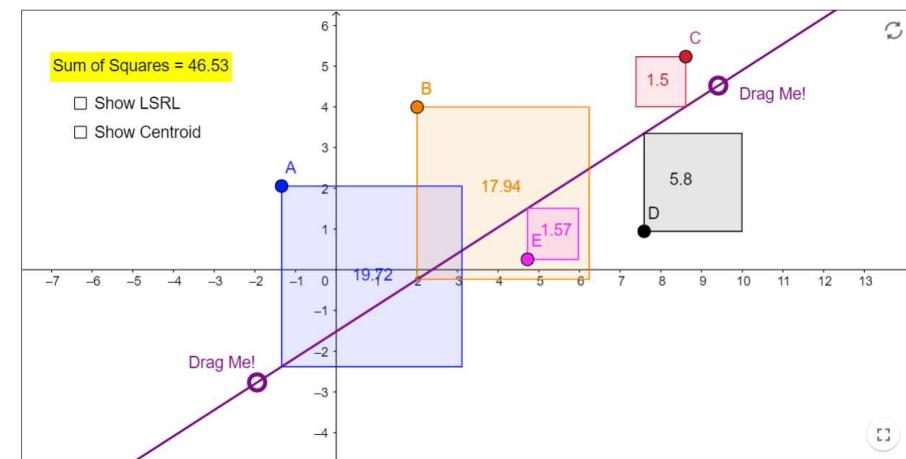
Least Squares Regression

Statistical method to show a relationship between x and y variables

The least squares regression function makes the vertical distance from the data points to the regression line the smallest

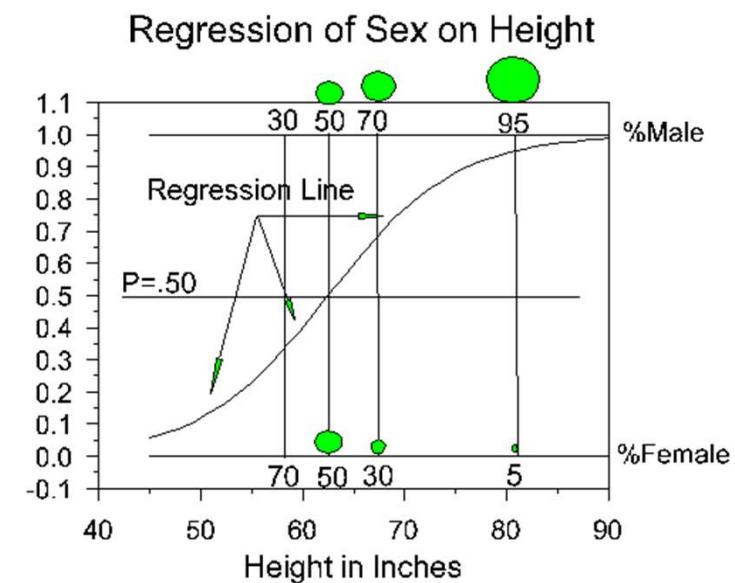
Minimizes the variance (sum of squares error)

“Regression” generally refers to simple linear least squares regression



Logistic Regression

- Independent variable(x) vs Nominal dependent variable (y)
- Regression line is still average but nonlinear
- Data points don't fall on regression line
- Example:
 - T –test look at Null hypothesis that cell reproduction rate is not linked to a tissue being cancerous
 - Logistic regression – predict the probability that tissue with a specific cellular reproduction rate will end up metastasizing in the next 5 months



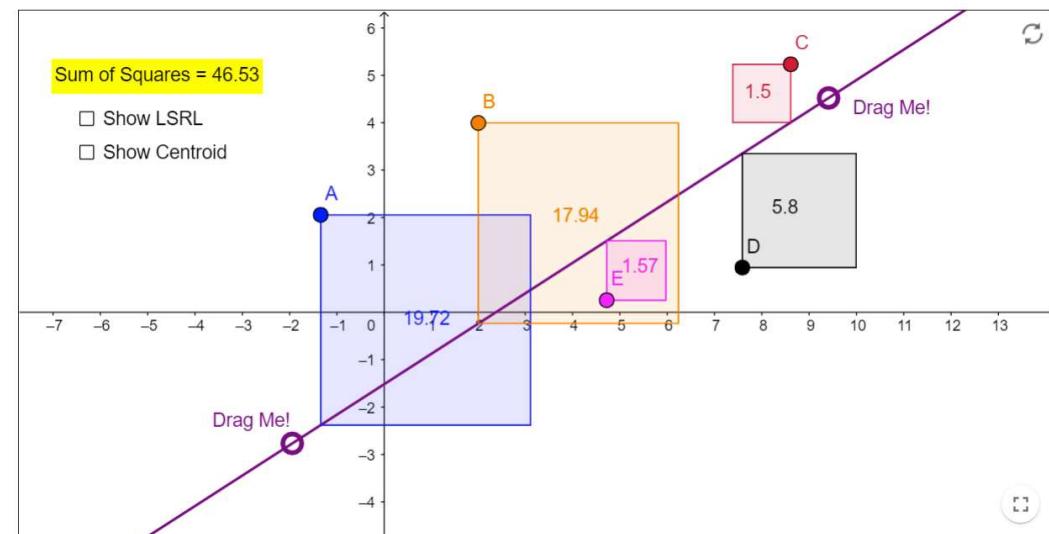
<http://faculty.cas.usf.edu/mbrannick/regression/Logistic.html>

Linear Regression

Statistical method to show a **linear** relationship between x and y variables

Finds the **line** of best fit
(regression line)

The Least Squares Regression Line makes the vertical distance from the data points to the regression line the smallest



Linear Regression

The equation for a straight line:

$$Y = \beta_1 X + \beta_0$$

β_0 = intercept

β_1 = slope

For a data point, $I (x_i, y_i)$, scattered about the line, its position can be represented by:

$$Y_i = \beta_1 X_i + \beta_0 + \epsilon_i$$

where ϵ represents the deviation of the point from the line

Therefore, our set of observations can be represented by a set of equations:

$$Y_1 = \beta_1 X_1 + \beta_0 + \epsilon_1$$

$$Y_2 = \beta_1 X_2 + \beta_0 + \epsilon_2$$

$$Y_3 = \beta_1 X_3 + \beta_0 + \epsilon_3$$

$$Y_4 = \beta_1 X_4 + \beta_0 + \epsilon_4$$

.....

$$Y_n = \beta_1 X_n + \beta_0 + \epsilon_n$$

Linear Regression

This set of equations can then be written in the form of the vectors and matrices:

$$Y = X\beta + \varepsilon$$

$$Y = \begin{vmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{vmatrix} \quad X = \begin{vmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{vmatrix} \quad \beta = \begin{vmatrix} \beta_0 \\ \beta_1 \end{vmatrix} \quad \varepsilon = \begin{vmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{vmatrix}$$

$$\boxed{\begin{vmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{vmatrix} = \begin{vmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{vmatrix} \cdot \begin{vmatrix} \beta_0 \\ \beta_1 \end{vmatrix} + \begin{vmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{vmatrix}}$$

$$Y_{i, \text{ave}} = X_i \beta \quad \boldsymbol{\varepsilon} = Y_i - Y_{i, \text{ave}}$$

Linear Regression

The overall failure of the data to fit the model is the residual, ϵ , sum of squares, $\sum \epsilon_i^2$:

$$|\epsilon_1 \epsilon_2 \dots \epsilon_n| \cdot \begin{vmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{vmatrix} = \sum \epsilon^2$$

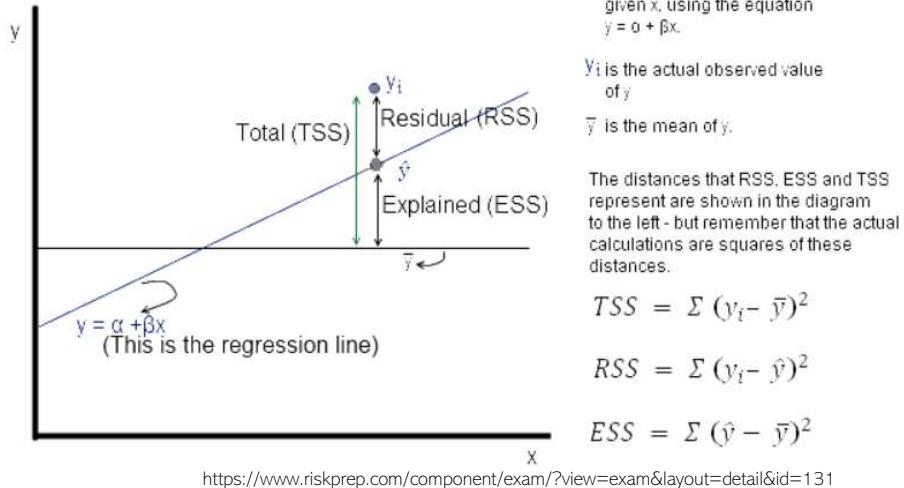
According to the method of least mean squares, we want to minimize $\sum \epsilon_i^2$. The equation that provides this estimates of b_0 and b_1 is:

We are interested in β :

$$\text{where: } (X'X)^{-1} = \frac{1}{n \sum X_i^2 - (\sum X_i)^2} \begin{pmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{pmatrix}$$

From this set of equations you can derive slope and intercept. After calculating sums you just need to perform simple matrix algebra to determine the intercept, β_0 , and slope, β_1 .

ANOVA



Source of Variation	df	Sum of Squares (SS)	Mean Sum of Squares (MSS)	F-test	p-value
Treatment	k-1	SSTr	MStr=SSTr/(k-1)	F=MStr/MSE	
Error	N-k	SSE	MSE=SSE/(N-k)		
Total	N-1	SSTo			

<https://courses.lumenlearning.com/suny-natural-resources-biometrics/chapter/chapter-5-one-way-analysis-of-variance/>

ANOVA

→ Is this the appropriate model?

Source	SS (<i>Sum of Squares, the numerator of the variance</i>)	DF (<i>the denominator</i>)	MS (<i>Mean Square, the variance</i>)	F
Regression (or Model)	$SSR = \sum_{i=1}^n ((\hat{\beta}_0 + \hat{\beta}_1 x_i) - \bar{y})^2$	$2-1=1$	$MSR = \frac{SSR}{1}$	$F = \frac{MSR}{MSE}$
Error	$SSE = \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$	$n-2$	$MSE = \frac{SSE}{n-2}$	
Total	$TSS = \sum_{i=1}^n (y_i - \bar{y})^2$	$n-1$		

df = “degrees of freedom”

F = Calculated F. This is compared to $F_{\alpha, df(R), df(E)}$

Multiple Regression

- Relationship between a dependent variable, Y, and several independent variables which simultaneously influence the dependent variable.
- β 's are called the **partial regression coefficients**
 - β_1 represents the true change in the mean of Y when X_1 changes by 1 unit, and all other variables are held constant
 - similarly for β_2 , β_3 , and β_4 ... ETC.

Multiple Regression Example

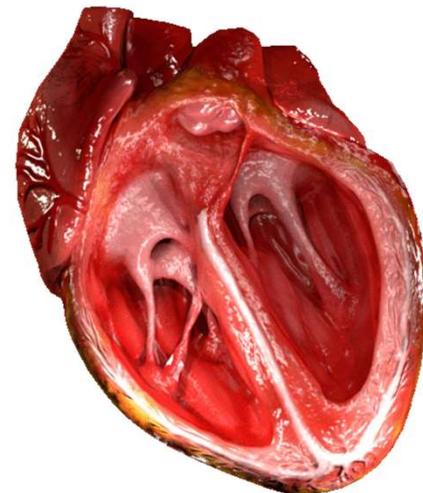
What is the relationship between left myocardial contractile force and serum ionic composition ?

response (dependent variable):

- contractile force (Y)

independent variables

- Chloride (Cl) (X1)
- phosphate (PO₄) (X2)
- Potassium (K) (X3)
- Sodium (Na) (X4)



The Data:

Y = contractile force (N)

X_1 = chloride (units)

X_2 = phosphate (units)

X_3 = potassium (units)

X_4 = sodium (units)

- measured in isolated heart preparations.

	X1 (Cl)	X2 (PO4)	X3 (K)	X4 (Na)	Y (force)
1	2.2	0.417	1.35	1.79	351
2	2.1	0.354	0.9	1.08	249
3	1.52	0.208	0.71	0.47	171
4	2.88	0.335	0.9	1.48	373
5	2.18	0.314	1.26	1.09	321
6	1.87	0.271	1.15	0.99	191
7	1.52	0.164	0.83	0.85	225
8	2.37	0.302	0.89	0.94	291
9	2.06	0.373	0.79	0.8	284
10	1.84	0.265	0.72	0.77	213
11	1.89	0.192	0.46	0.46	138
12	2.45	0.221	0.76	0.95	213
13	1.88	0.186	0.52	0.95	151
14	1.93	0.207	0.6	0.92	130
15	1.8	0.157	0.67	0.6	93
16	1.81	0.195	0.47	0.57	95
17	1.49	0.165	0.66	0.8	147
18	1.53	0.226	0.68	0.66	88
19	1.43	0.224	0.44	0.45	65
20	1.54	0.271	0.51	0.95	120
21	1.13	0.187	0.38	0.63	72
22	1.63	0.2	0.62	1.1	160
23	1.36	0.211	0.71	0.47	72
24	1.76	0.283	0.96	0.96	252
25	2.53	0.284	0.85	1.39	310
26	2.59	0.303	1.02	0.95	336
TOTALS	49.29	6.515	19.81	23.07	5111

Example: Regression Equation

In the absence of any biological understanding of how contractile force may be related to blood ion composition, it's best to first choose a linear model:

for $i = 1, \dots, 26$

Assumptions:

- $\epsilon_1, \epsilon_2, \dots, \epsilon_{26}$ are a random sample from a normal population with a mean = 0 and some constant (unknown) σ^2 . [i.e. $\epsilon_i \sim N(0, \sigma^2)$]
- The relationship between Y and each X (all other X's held constant) is linear.
- The X's do not interact

Example: Estimating the Parameters

- use least squares analysis.
i.e. choose estimates of β as $[b_0, b_1, b_2, b_3, b_4]$ which minimize:

$$\sum_{i=1}^n (Y_i - \hat{\mu}_{Y.1234})^2 = \sum_{i=1}^n (Y_i - b_0 - b_1 X_{i1} - b_2 X_{i2} - b_3 X_{i3} - b_4 X_{i4})^2$$

(equation 1)

- this is a calculus problem. Differentiating the above expression, with respect to each b , results in 5 equations in the 5 variables (b_0, b_1, b_2, b_3, b_4) to be solved.

Example: Normal Equations

$$b_0 n + b_1 X_{\bullet 1} + b_2 X_{\bullet 2} + b_3 X_{\bullet 3} + b_4 X_{\bullet 4} = Y_{\bullet} = \sum Y$$

$$b_0 X_{\bullet 1} + b_1 \sum X_1^2 + b_2 \sum X_1 X_2 + b_3 \sum X_1 X_3 + b_4 \sum X_1 X_4 = \sum X_1 Y$$

$$b_0 X_{\bullet 2} + b_1 \sum X_2 X_1 + b_2 \sum X_2^2 + b_3 \sum X_2 X_3 + b_4 \sum X_2 X_4 = \sum X_2 Y$$

$$b_0 X_{\bullet 3} + b_1 \sum X_3 X_1 + b_2 \sum X_3 X_2 + b_3 \sum X_3^2 + b_4 \sum X_3 X_4 = \sum X_3 Y$$

$$b_0 X_{\bullet 4} + b_1 \sum X_4 X_1 + b_2 \sum X_4 X_2 + b_3 \sum X_4 X_3 + b_4 \sum X_4^2 = \sum X_4 Y$$

(equations 2)

- the solutions to these equations is only simple high school algebra.

Example: Matrix Representation

Equations can be rewritten using matrix notation.

$$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \beta_3 X_{13} + \beta_4 X_{14} + \varepsilon_1$$

Model: $Y_2 = \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \beta_3 X_{23} + \beta_4 X_{24} + \varepsilon_2$

⋮

$$Y_n = \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \beta_3 X_{n3} + \beta_4 X_{n4} + \varepsilon_n$$

Matrix Notation:

$$Y = X \underline{\beta} + \underline{\varepsilon}$$

(equation 3)

Example: Matrix Representation

where:

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} & X_{14} \\ 1 & X_{21} & X_{22} & X_{23} & X_{24} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & X_{n3} & X_{n4} \end{bmatrix} = \begin{bmatrix} 1 & 2.20 & 0.417 & 1.35 & 1.79 \\ 1 & 2.10 & 0.354 & 0.90 & 1.08 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 2.59 & 0.303 & 1.02 & 0.95 \end{bmatrix}_{26 \times 5}$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 351 \\ 249 \\ \vdots \\ 336 \end{bmatrix}_{26 \times 1} \quad \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}_{5 \times 1} \quad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{26} \end{bmatrix}_{26 \times 1}$$

Example: Matrix Representation

Furthermore, the Normal Equations (equations 2) can be rewritten as the Matrix Equation...

$$(X'X) \cdot b = X'Y \quad (\text{equation 4})$$

NOTE: $X'X$ is a symmetric matrix (about the diagonal)

$$X'X = \begin{bmatrix} n & \sum X_1 & \sum X_2 & \sum X_3 & \sum X_4 \\ \sum X_1 & \sum X_1^2 & \sum X_1 X_2 & \sum X_1 X_3 & \sum X_1 X_4 \\ \sum X_2 & \sum X_2 X_1 & \sum X_2^2 & \sum X_2 X_3 & \sum X_2 X_4 \\ \sum X_3 & \sum X_3 X_1 & \sum X_3 X_2 & \sum X_3^2 & \sum X_3 X_4 \\ \sum X_4 & \sum X_4 X_1 & \sum X_4 X_2 & \sum X_4 X_3 & \sum X_4^2 \end{bmatrix} \quad X'Y = \begin{bmatrix} \sum Y \\ \sum X_1 Y \\ \sum X_2 Y \\ \sum X_3 Y \\ \sum X_4 Y \end{bmatrix}$$

$$b_0n + b_1X_{\bullet 1} + b_2X_{\bullet 2} + b_3X_{\bullet 3} + b_4X_{\bullet 4} = Y_{\bullet} = \sum Y$$

$$b_0X_{\bullet 1} + b_1\sum X_1^2 + b_2\sum X_1X_2 + b_3\sum X_1X_3 + b_4\sum X_1X_4 = \sum X_1Y$$

$$b_0X_{\bullet 2} + b_1\sum X_2X_1 + b_2\sum X_2^2 + b_3\sum X_2X_3 + b_4\sum X_2X_4 = \sum X_2Y$$

$$b_0X_{\bullet 3} + b_1\sum X_3X_1 + b_2\sum X_3X_2 + b_3\sum X_3^2 + b_4\sum X_3X_4 = \sum X_3Y$$

$$b_0X_{\bullet 4} + b_1\sum X_4X_1 + b_2\sum X_4X_2 + b_3\sum X_4X_3 + b_4\sum X_4^2 = \sum X_4Y$$

$$X'X = \begin{bmatrix} n & \sum X_1 & \sum X_2 & \sum X_3 & \sum X_4 \\ \sum X_1 & \sum X_1^2 & \sum X_1X_2 & \sum X_1X_3 & \sum X_1X_4 \\ \sum X_2 & \sum X_2X_1 & \sum X_2^2 & \sum X_2X_3 & \sum X_2X_4 \\ \sum X_3 & \sum X_3X_1 & \sum X_3X_2 & \sum X_3^2 & \sum X_3X_4 \\ \sum X_4 & \sum X_4X_1 & \sum X_4X_2 & \sum X_4X_3 & \sum X_4^2 \end{bmatrix} \quad X'Y = \begin{bmatrix} \sum Y \\ \sum X_1Y \\ \sum X_2Y \\ \sum X_3Y \\ \sum X_4Y \end{bmatrix} \quad (\text{equations 2})$$

Example: Solve for b

The solution to Equation 4 can also be rewritten as:

$$\underline{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = (X'X)^{-1} \cdot X'Y$$

where $(X'X)^{-1}$ denotes the inverse of the matrix $X'X$. This now looks simple but it still requires as much computation as the normal equations.

- The advantage to the matrix approach is in book keeping (neater to write down equations and keep track of data).

Example: Sub in real numbers

$$X'X = \begin{bmatrix} 26.0 & 49.29 & 6.515 & 19.81 & 23.07 \\ 49.29 & 97.9781 & 12.7981 & 39.0012 & 45.9843 \\ 6.515 & 12.7981 & 1.7540 & 5.2688 & 6.1600 \\ 19.81 & 39.0012 & 5.2688 & 16.6387 & 18.9306 \\ 23.07 & 45.9843 & 6.1600 & 18.9306 & 23.1015 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 5111.0 \\ 10521.2 \\ 1409.48 \\ 4363.95 \\ 5129.58 \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} 0.9186 & -0.3780 & -0.9663 & -0.1060 & 0.1796 \\ -0.3780 & 0.4209 & -0.6846 & -0.0550 & -0.2328 \\ -0.9663 & -0.6846 & 20.1257 & -2.3701 & -1.0967 \\ -0.1060 & -0.0550 & -2.3701 & 1.5028 & -0.3842 \\ 0.1796 & -0.2328 & -1.0967 & -0.3842 & 0.9346 \end{bmatrix}$$

Example: Solution

$$b = (X' X)^{-1} \cdot X' Y$$

Therefore,

$$b = \begin{bmatrix} 0.9186 \\ -0.3780 \\ -0.9663 \\ -0.1060 \\ 0.1796 \end{bmatrix} \begin{bmatrix} 5111.0 \\ 10521.2 \\ 1409.48 \\ 4363.95 \\ 5129.58 \end{bmatrix} = \begin{bmatrix} -185.33 \\ 97.76 \\ 256.97 \\ 126.57 \\ 40.28 \end{bmatrix}$$

Example: Solution

$$\hat{\mu}_{Y_{1234}} = Y = -185.33 + 97.8X_1 + 257X_2 + 126.6X_3 + 40.3X_4$$

Interpretation:

- it is estimated that hearts from the population sampled with 1 extra unit of blood chloride will beat with 97.8N of force, if all other components of the blood were held constant.

(similarly for PO₄, K, and Na)

- obviously the assumption that linearity is valid will only be true for a certain range (i.e. 50 units of Cl would not result in 50 x 97.8N of extra force as this would not be physiologically possible)

Multiple Regression ANOVA

$$\sum_{\text{(total SS)}} \left(Y_i - \bar{Y} \right)^2 = \sum_{\text{(Model SS)}} \left(\hat{\mu}_{Y.1234} - \bar{Y} \right)^2 + \sum_{\text{(Residual SS)}} \left(Y_i - \hat{\mu}_{Y.1234} \right)^2$$

$$Model(SS) = \sum_{\text{(Model SS)}} \left(\hat{\mu}_{Y.1234} - \bar{Y} \right)^2 = b' \cdot (X' Y) - \frac{1}{n} \left(\sum Y_i \right)^2$$

$$Total(SS) = \sum_{\text{(total SS)}} \left(Y_i - \bar{Y} \right)^2 = \sum Y_i^2 - \frac{1}{n} \left(\sum Y_i \right)^2$$

$$residual(SS) = total(SS) - Model(SS) = \sum Y_i^2 - b' \cdot (X' Y)$$

ANOVA for the Example

Back to the cardiac force example....

$$Total(SS) = 1232659 - \frac{(5111)^2}{26} = 227954.35$$

$$Model(SS) = [-185.33 \quad 97.76 \quad 256.97 \quad 126.57 \quad 40.28] \cdot \begin{bmatrix} 5111.0 \\ 10521.2 \\ 1409.48 \\ 4363.95 \\ 5129.58 \end{bmatrix} - \frac{(5111)^2}{26} = 197832.43$$

Null Hypothesis: None of the independent variables are of any value in explaining the variation in myocardial contraction force .

$$\text{i.e. } H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

Alternative, H_1 : not all β 's are zero.

Source	df	SS	MS	F_c	$F_{ij, \alpha}$
regression	4	197382.43	49458.11	34.48	2.84
residual	21	30121.92	1434.38	-----	-----
TOTAL	25	227954.35	-----	-----	-----

$F_c > F_{ij, \alpha}$ Therefore, reject H_0 ; at least one β is not zero.

Model Validation: Assumptions

Assumptions that were made:

- 1) The true mean of Y has been correctly specified.
- 2) The $\text{var}(\varepsilon_i) = \sigma^2$ are constant
- 3) The ε are independent (i.e. uncorrelated) with one another
- 4) The ε come from a normal distribution.

It is safe to say that not all of these assumptions will be completely met.

All we really require is that they are approximately true.

$$\mu_Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

Assessing Assumptions

There are 2 general ways of assessing whether assumptions are at least approximately true.

1) Overfit the model

- add additional parameters and retest.
- maybe there are non-linear terms, or interactions between some of the X's.
- If it is suspected that ϵ are not independent then additional terms called variance components can be added which allows you to see model correlations.

2) Examine residuals

- these are roughly “estimated” errors and hence reflect the properties of the true errors.
- Graphical analysis is most commonly performed for residual analysis.

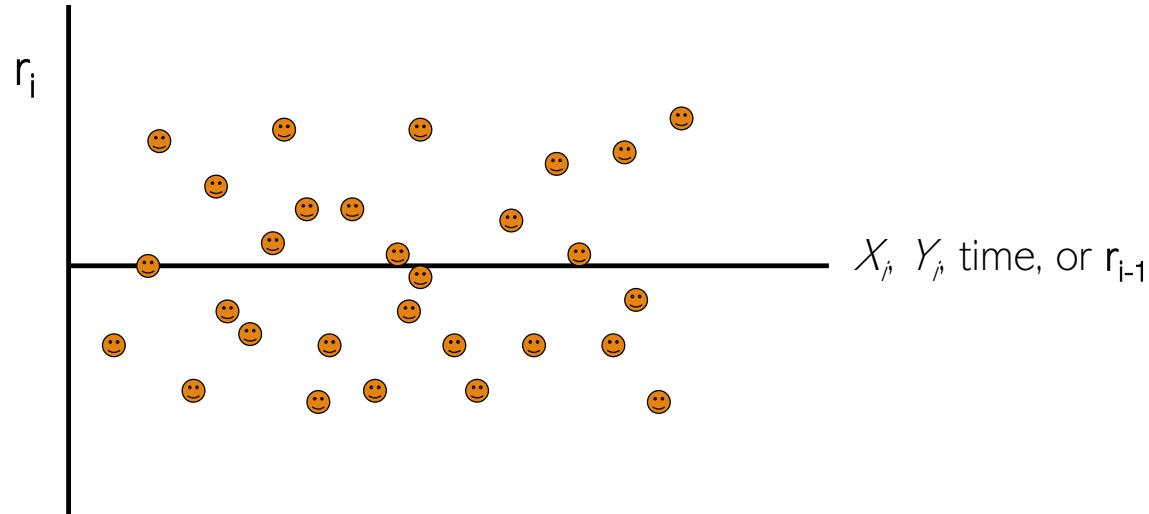
Type of Residual Plots

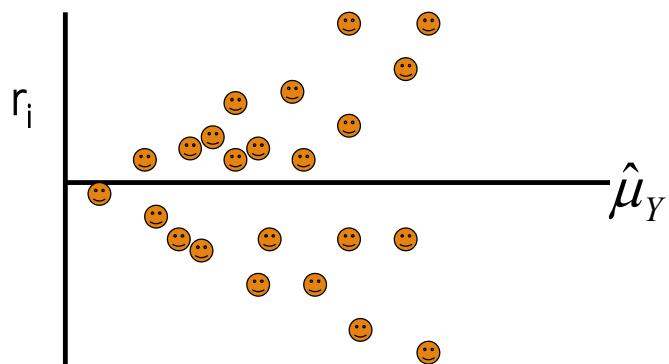
- 1) Plot residuals vs. $\hat{\mu}_Y$
- 2) Plot residuals vs. each independent variable, X_i
- 3) Plot residuals vs. time of observation (if appropriate)
- 4) Plot r_i vs. r_{i-1} (where $i = 2, \dots, n$), to detect serial correlation, assuming the observations are ordered in time or space.

$$r_i = Y_i - \hat{\mu}_Y \text{ (residual = observed - expected)}$$

Residual Plots

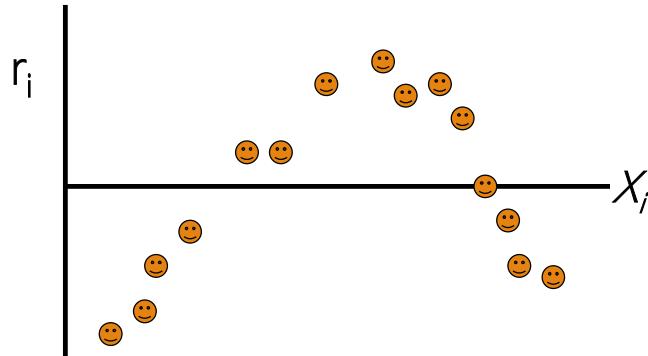
re: residuals should not exhibit any pattern and fall roughly in a band of constant width parallel to the x-axis.





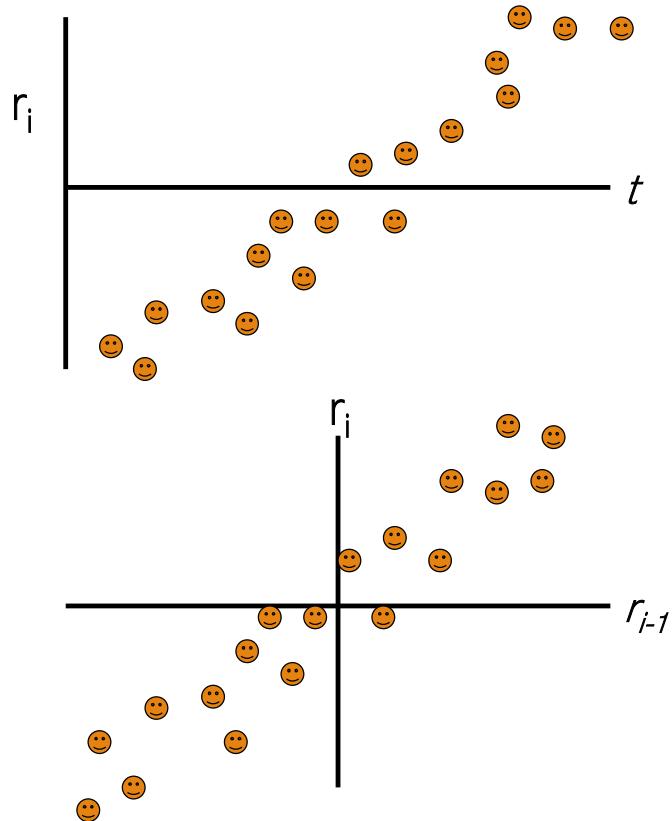
- Assesses assumption of homogeneity of error variance.
- Here error variance \uparrow with mean

FIX: Transform Y



- Detect curvature of relationship between response and X_i

FIX: addition of X_i^2 , X_i^3 , $\log(X_i)$, or even $X_i X_j$ interaction term(s).



- Response may \uparrow or \downarrow with time of collection.

FIX: Add time as an independent variable in the model.

- Detects departure from the assumption of independence of errors.
- When data are collected over time, if the errors are not independent they will have a +ve serial correlation.

FIX: modeled variance components.

Outlier

A data point that differs significantly from others within the data set

I.e. lies an abnormal distance from other values in the population

This could be due to experimental error

Could also be due to variability in the measurement

Are generally removed or excluded from data set

Outlier Detection

- 1) Most algorithms are based on Normal Distributions
- 2) Typically work one point at a time. But, if more than one point is suspected use multiple outlier test.
- 3) Approach doesn't work when <6 points to assess
- 4) Can use mathematical approaches or graphical (e.g. box plot, histogram)

Masking vs Swamping

- a difficult problem!

Masking = too few outliers suggested in the test.

Swamping = specify too many outliers in the test.

One should always complement formal outlier tests with graphical methods.

Swamping and masking are why many tests require that the exact number of outliers being tested is specified

Z-Score and Modified Z-Score

$$Z = \frac{Y_i - \bar{Y}}{s}$$

\bar{Y} = sample mean
 s = sample standard deviation

$$M_i = \frac{0.6745(Y_i - \tilde{Y})}{\text{median}(|Y_i - \tilde{Y}|)}$$

\tilde{Y} = sample median
denominator = MAD (median absolute deviation)

If $M_i > 3.5$ then there's good chance that value is an outlier

Outlier tests

Sample formal outlier tests are grouped by the following characteristics:

- 1) How are the data distributed? Most tests assume approximately normal distribution.
- 2) Is the test designed for a single outlier or multiple outliers?
- 3) If designed for multiple outliers, does the number need to be known exactly or can a range be given?

Outlier Tests

- 1) [Grubbs' Test](#). Recommended test when testing for a single outlier.
- 2) [Tietjen-Moore Test](#). This is a generalization of Grubbs' test to account for more than one outlier. It has the limitation that the number of outliers must be specified exactly.
- 3) [Generalized Extreme Studentized Deviate \(ESD\) Test](#). Only an upper bound on the suspected number of outliers is needed. Recommended test when the exact number of outliers is not known.

What if want to find the outlier?

Maybe you want to find the outlier (i.e. it is the needle in the haystack you are looking for)

Things get way more complicated!

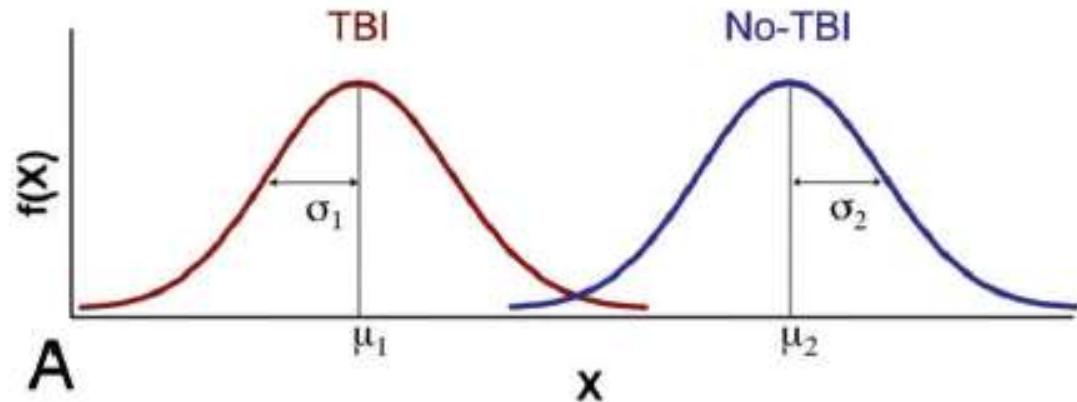
- Anomaly detection
- Data mining

e.g. Use the “Mahalanobis distance” to find outliers.

- Looks at difference between point and a distribution

However this in itself is being effected by the outliers

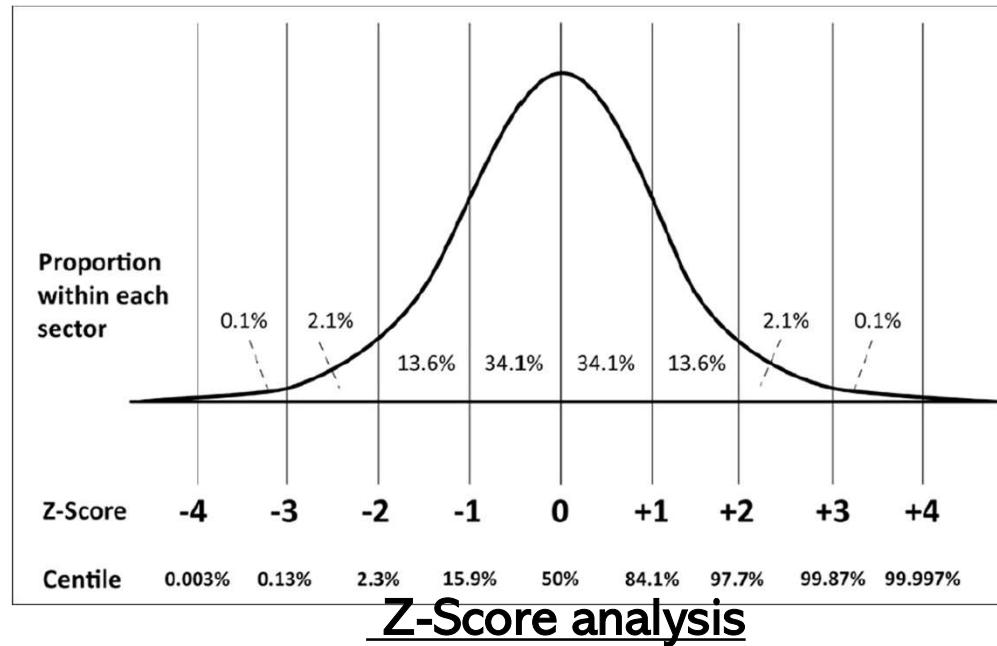
Example

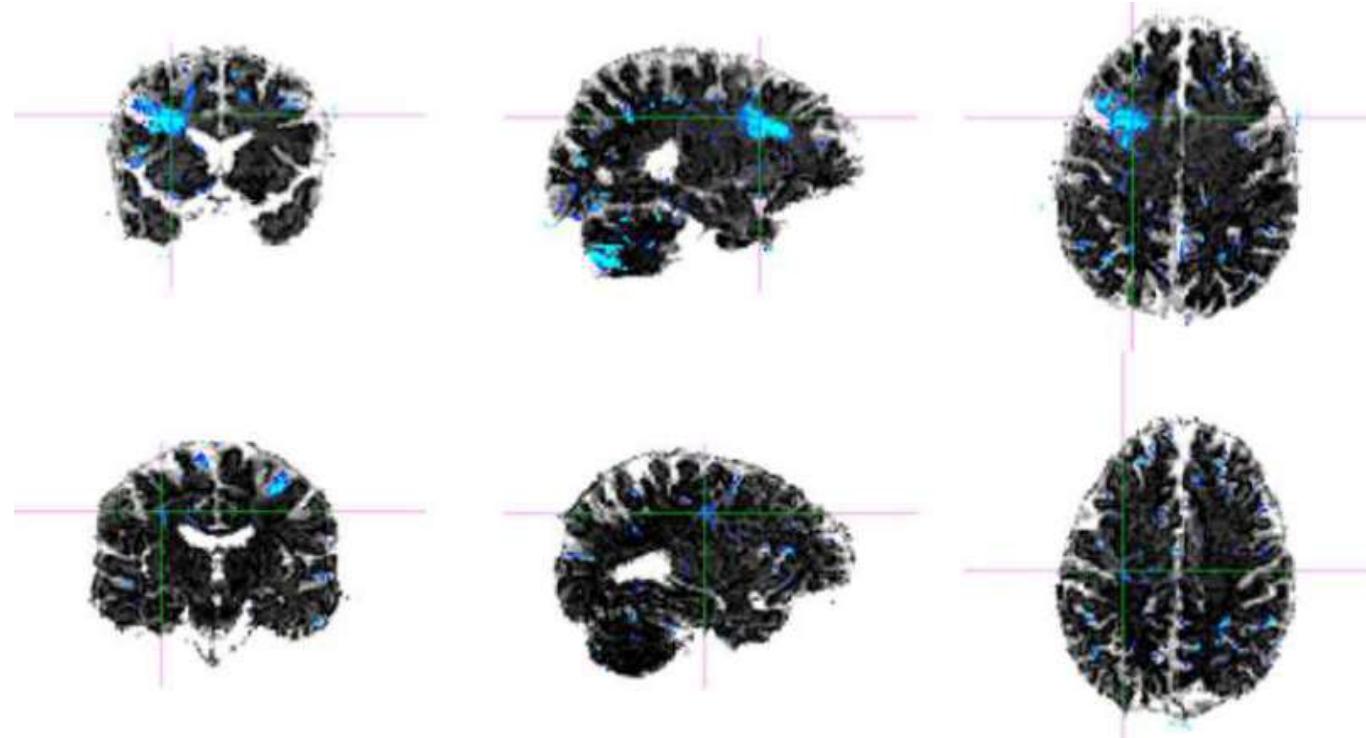


HYPOTHESIS: If the brain has been mechanically injured there will be increase in water diffusivity.

- Needs to be assessed voxel-by-voxel
- Requires normative data (~50 controls)
- All brains need to be spatially warped to a standard template
- Verification of normality (Skewness and Kurtosis) is critical

Group Analysis





Z-score maps for two patients with chronic mTBI (blue indicates statistically significant areas of free-water compared with the normative atlas)

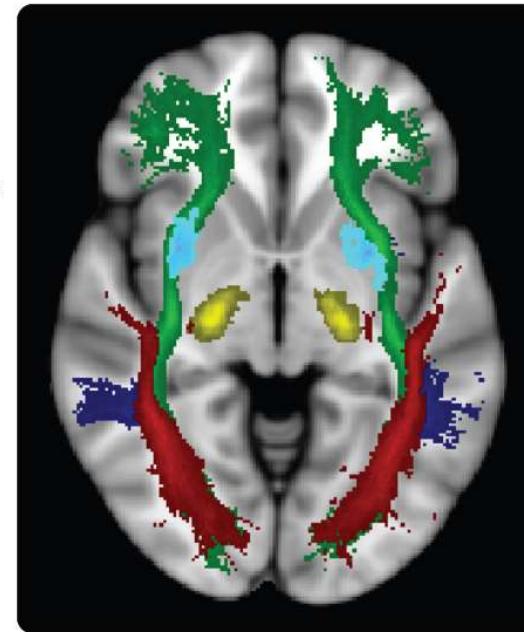
Brain Imaging and Behavior (2012) 6:137–192

Case-based z-scoring

- Identified abnormalities using z-scoring for 24 unique brain ROI within each concussed subject:
- $$z = \frac{x - \mu}{\sigma}$$
- ROI is considered abnormal if it is greater than $\pm 2\sigma$ relative to the control-group mean

Metric:	Abnormal Region:
FA	Superior longitudinal fasciculus
FA	Inferior longitudinal fasciculus
FA	Inferior fronto-occipital fasciculus
AD, RD	Corticospinal tract
RD	Uncinate fasciculus

*Regions found to be abnormal in >40% of the concussed participants



Outliers, following removal of non-normative control voxels (2-3%), suggestive of damage

ROI	Control Mean	Control SD	No. Outliers ($\pm 2\sigma$)	No. Outliers ($\pm 3\sigma$)
Acoustic Radiation Left	0.2844	0.0257	5	0
Acoustic Radiation Right	0.2724	0.0257	4	0
Cingulate Gyrus Left	0.3609	0.0379	2	0
Cingulate Gyrus Right	0.3095	0.0485	1	0
Cingulum Left	0.3455	0.0407	3	0
Cingulum Right	0.3580	0.0513	0	0
Corpus Callosum	0.4076	0.0319	0	0
Corticospinal Tract Left	0.4713	0.0229	3	2
Corticospinal Tract Right	0.4652	0.0219	5	1
Forceps Major	0.3871	0.0512	0	0
Forceps Minor	0.3770	0.0265	3	1
Fornix	0.2999	0.0415	2	0
Hippocampus Left	0.2652	0.0405	2	0
Hippocampus Right	0.2764	0.0399	1	0
Inferior Fronto-occipital Fasciculus Left	0.3978	0.0207	12	2
Inferior Fronto-occipital Fasciculus Right	0.3945	0.0255	5	0
Inferior Longitudinal Fasciculus Left	0.3517	0.0230	12	1
Inferior Longitudinal Fasciculus Right	0.3600	0.0282	6	1
Optic Radiation Left	0.2975	0.0157	1	0
Optic Radiation Right	0.3092	0.0192	2	0
Superior Longitudinal Fasciculus Left	0.3337	0.0178	11	4
Superior Longitudinal Fasciculus Right	0.3482	0.0185	17	8
Uncinate Fasciculus Left	0.3786	0.0322	7	1
Uncinate Fasciculus Right	0.3534	0.0401	4	0



Integrated Biomedical
Engineering & Health
Sciences Program

IBEHS - 4QZ3
Modelling of Biological Systems

Lecture 3

TAYLOR DEVET MASC.

PHD. CANDIDATE BIOLOGICAL AND BIOMEDICAL ENGINEERING

MCGILL UNIVERSITY

SHRINERS HOSPITAL FOR CHILDREN

Todays Aims...



Correlation



Complete Randomized
Design



Post Hoc Testing

Let's go back to the cardiac data.....

Cardiac Example

Null Hypothesis: None of the independent variables are of any value in explaining the variation in cardiac force output.
 i.e. $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

Alternative, H_1 : not all β 's are zero.

Source	df	SS	MS	F_c	$F_{i,j,\alpha}$
regression	4	197382.43	49458.11	34.48	2.84
residual	21	30121.92	1434.38	-----	-----
TOTAL	25	227954.35	-----	-----	-----

$F_c > F_{i,j,\alpha}$. Therefore, reject H_0 ; at least one β is not zero.

SSCP Matrix

Sum of Squares and Cross Product Matrix

Main diagonal are sum of squares for each column

Off main diagonals are sums of cross products

Inverse of this times Stdev gives us a matrix with variances and covariances

26.0000	49.2900	6.5150	19.8100	23.0700	0.9186	-0.3780	-0.9663	-0.1060	0.1796
49.2900	97.9781	12.7980	39.0012	45.9843	-0.3780	0.4209	-0.6846	-0.0550	-0.2328
6.5150	12.7980	1.7540	5.2688	6.1600	-0.9663	-0.6846	20.1257	-2.3701	-1.0967
19.8100	39.0012	5.2688	16.6387	18.9306	-0.1060	-0.0550	-2.3701	1.5028	-0.3842
23.0700	45.9843	6.1600	18.9306	23.1015	0.1796	-0.2328	-1.0967	-0.3842	0.9346

$$(X'X)$$

$$(X'X)^{-1}$$

Variance Covariance Matrix

$$\begin{vmatrix}
 Var(b_0) & Cov(b_0, b_1) & Cov(b_0, b_2) & Cov(b_0, b_3) & Cov(b_0, b_4) \\
 Cov(b_1, b_0) & Var(b_1) & Cov(b_1, b_2) & Cov(b_1, b_3) & Cov(b_1, b_4) \\
 Cov(b_2, b_0) & Cov(b_2, b_1) & Var(b_2) & Cov(b_2, b_3) & Cov(b_2, b_4) \\
 Cov(b_3, b_0) & Cov(b_3, b_1) & Cov(b_3, b_2) & Var(b_3) & Cov(b_3, b_4) \\
 Cov(b_4, b_0) & Cov(b_4, b_1) & Cov(b_4, b_2) & Cov(b_4, b_3) & Var(b_4)
 \end{vmatrix} = s^2 \times (X' X)^{-1}$$

$$= \begin{vmatrix}
 1317.55 & -542.151 & -1386.08 & -152.066 & 257.6247 \\
 -542.151 & 603.774 & -981.925 & -78.8711 & -333.958 \\
 -1386.08 & -981.925 & 28867.82 & -3399.59 & -1573.08 \\
 -152.066 & -78.8711 & -3399.59 & 2155.64 & -551.089 \\
 257.6247 & -333.958 & -1573.08 & -551.089 & 1340.625
 \end{vmatrix}$$

Cardiac Example cont

With rejection of H_0 , one can test for specific factors.

e.g. $H_0 : \beta_4 = c$

where c is any number specified (often zero).

$$(X'X)^{-1}$$

0.9186	-0.3780	-0.9663	-0.1060	0.1796
-0.3780	0.4209	-0.6846	-0.0550	-0.2328
-0.9663	-0.6846	20.1257	-2.3701	-1.0967
-0.1060	-0.0550	-2.3701	1.5028	-0.3842
0.1796	-0.2328	-1.0967	-0.3842	0.9346

1317.55	-542.151	-1386.08	-152.066	257.6247
-542.151	603.774	-981.925	-78.8711	-333.958
-1386.08	-981.925	28867.82	-3399.59	-1573.08
-152.066	-78.8711	-3399.59	2155.64	-551.089
257.6247	-333.958	-1573.08	-551.089	1340.625

$$\beta = \begin{bmatrix} -185.33 \\ 97.76 \\ 256.97 \\ 126.57 \\ 40.28 \end{bmatrix}$$

- s^2 is the residual(MS) from the full model (ANOVA) table.

Therefore, there is little evidence to reject $H_0 : \beta_4 = 0$

Cardiac Example cont

A 95% confidence interval may be calculated for β_4 :

$$b_4 \pm t_{\alpha, df} \times stderr(b_4) = 40.3 \pm (2.080 \times 36.61)$$

$$stderr(b_4) = \sqrt{s^2 \times (X'X)^{-1}_{5,5}}$$

Where: $t_{\alpha, df} = t_{0.05, 26-5} = 2.080$

Therefore, $-35.8 \leq b_4 \leq 116.4$

Reject $H_0: \beta_4 = 0$ Based on the confidence interval it could very well be zero.

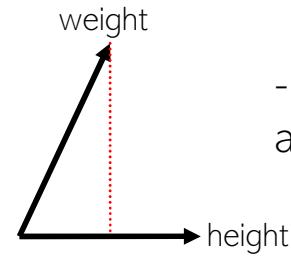
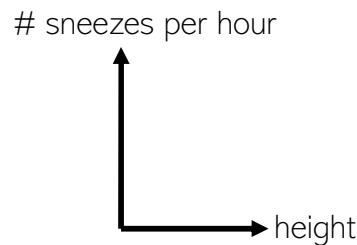
Correlated Variables

Need to understand whether independent variables are correlated.

If independent variables are not correlated among themselves they are said to be orthogonal.

When correlated they are non-orthogonal.

Non-correlated independent variables = **Orthogonal**



- weight and height
are correlated

Multicollinearity

A lot of these calculations pend on the $X'X$ matrix being invertible.

This problem is referred to as the multicollinearity problem

Multicollinearity occurs when there are high correlations between two or more predictor variables.

i.e. one predictor variable can be used to predict the other giving redundancy and skewing the model.

Multicollinearity

Examples of multicollinear predictors:

- 1) a person's height and weight
- 2) years of education and annual income.
- 3) Height in metres and height in feet

An easy way to detect:

Calculate correlation coefficients (r) for all pairs of predictor variables. If r is exactly +1 or -1, this is called perfect multicollinearity and one of the variables should be removed from the model

Main Causes of Multicollinearity:

- 1) Data-based
 - poorly designed experiments
 - data that is 100% observational, or data collection methods that cannot be manipulated.
 - In some cases, variables may be highly correlated (usually due to collecting data from purely observational studies) and there is no error on the researcher's part.
Can test this in advance!!
- 2) Structural multicollinearity
 - caused by the researcher, poorly creating new predictor variables

There is a fix!! Plan the experiment!

Observational Example:

Estimate 4 slopes and 1 intercept of zero (i.e. not estimating)

X =

1	0	0	0
0	1	0	1
0	0	0	1
1	0	0	0
0	1	0	1
0	0	0	1
1	0	0	0
0	1	0	1
0	0	0	1
1	0	0	0
0	1	0	1
0	0	0	1

>> X'*X

ans =

4	0	0	0
0	4	0	4
0	0	0	0
0	4	0	8

>> (X'*X)^-1

Warning: Matrix is singular to working precision.

ans =

Inf	Inf	Inf	Inf
Inf	Inf	Inf	Inf
Inf	Inf	Inf	Inf
Inf	Inf	Inf	Inf

Observational Example:

Estimate 4 slopes and 1 intercept as b

X =

1	1	0	0	0
1	0	1	0	0
1	0	0	1	0
1	0	0	0	1
1	1	0	0	0
1	0	1	0	0
1	0	0	1	0
1	0	0	0	1
1	1	0	0	0
1	0	1	0	0
1	0	0	1	0
1	0	0	0	1
1	1	0	0	0
1	0	1	0	0
1	0	0	1	0
1	0	0	0	1

>> X'*X

ans =

16	4	4	4	4
4	4	0	0	0
4	0	4	0	0
4	0	0	4	0
4	0	0	0	4

$$(X'X)^{-1} = 1e+15 \cdot \begin{bmatrix} 4.50 & -4.50 & -4.50 & -4.50 & -4.50 \\ -4.50 & 4.50 & 4.50 & 4.50 & 4.50 \\ -4.50 & 4.50 & 4.50 & 4.50 & 4.50 \\ -4.50 & 4.50 & 4.50 & 4.50 & 4.50 \\ -4.50 & 4.50 & 4.50 & 4.50 & 4.50 \end{bmatrix}$$

Essentially this is infinite!!

Correlated Variables

In the heart example the independent variables are correlated.

i.e. force is related to a combination of ions!

- What about sodium, β_4 ? We tested and accepted $H_0: \beta_4 = 0$

The general approach to the analysis is the comparison of two competing models:

1) The full model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i$$

2) The reduced model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

Model Reduction

Fit the full model (done already) and reduced model to the data and for each obtain the model SS and residual SS.

- The appropriate SS for testing $H_0: \beta_4 = 0$ is:

$$\begin{aligned} R(\beta_4 | \beta_1, \beta_2, \beta_3) &= \text{SS model (full)} - \text{SS model (reduced)} \\ &= \text{SS residual (reduced)} - \text{SS residual (full)} \end{aligned}$$

- this 'R' notation helps to differentiate the different SS.
- 'R' means 'Reduction in residual SS'

Model Reduction

The full model (already calculated) is:

$$R(\beta_1, \beta_2, \beta_3, \beta_4) = \text{SS (full model)} = 197832.43$$

The reduced model is:

$$R(\beta_1, \beta_2, \beta_3) = \text{SS (reduced model)} = 196096.77$$

Therefore,

$$\begin{aligned} R(\beta_4 | \beta_1, \beta_2, \beta_3) &= R(\beta_1, \beta_2, \beta_3, \beta_4) - R(\beta_1, \beta_2, \beta_3) \\ &= 197832.43 - 196096.77 \\ &= 31857.58 - 30121.92 \\ &= 1735.66 \end{aligned}$$

NOTE: red #'s are from reduced model - you should verify these !!

Model Reduction

The reduced model is fitted in the same way as the full model !!

$$\hat{\mu}_{Y.123} = -193.07 + 107.80X_1 + 304.24X_2 + 143.13X_3$$

Source	df	SS	MS
Reduced model	3	196096.77	65365.59
residual	22	31857.58	1448.07
TOTAL	25	227954.35	-----

NOTE: you should verify the ANOVA and reduced b results !!

Model Reduction

- the df for $\beta_4 = 1$ (i.e. 4 - 3)

Calculation of the F statistic to test $H_0: \beta_4 = 0$

For the denominator MS of the F-test the residual MS from the FULL MODEL is used!

$$F = \frac{R(\beta_4 | \beta_1, \beta_2, \beta_3) / df_{\beta_4}}{\text{Error}(MS)_{\text{FULL}}} =$$

$$F_{1,21,\alpha=0.05} = 4.32 \text{ (table)}$$

Since $F_c < F_{\text{table}}$ there is little evidence to suggest $\beta_4 \neq 0$

Which model do we want?

Use F statistic

$$F = \frac{\frac{SSE_{reduced} - SSE_{Full}}{df_{reduced} - df_{full}}}{\frac{SSE_{full}}{df_{full}}}$$

$$F = \frac{\frac{31857.58 - 30121.92}{22 - 21}}{\frac{30121.9}{21}}$$

$$F = 1735.66/1434.37 = 1.21$$

$$\frac{df_{Reduced} - df_{Full}}{df_{Full}} =$$
$$F(\text{alpha } 1, 21) = 4.32$$

Model Reduction

Therefore, it can be concluded that Na is not related to force after allowing for the influence of Cl⁻, PO₄ and K⁺.

Note: If we had only done simple linear regression of Y on X4, ignoring X1, X2, and X3:

- Would have been fitting the model $Y_i = \beta_0 + \beta_4 X_4 + \varepsilon_i$)
- Testing the significance of the regression (i.e. $H_0: \beta_4 = 0$) would have resulted in $F_c = 34.4$
- Testing two VERY different hypotheses
→ Therefore careful experimental planning are essential !

Estimating Y for Given Values of X's

Example. Estimate the contractile force with $\text{Cl}^- = 2$, $\text{PO}_4 = 0.3$, and $\text{K}^+ = 0.8$ in the buffer. (note X_4 removed, based on previous test)

Therefore the estimate of force is:

In other words, we estimate the force output will be 228.3 N.

Standard Error of Estimate:

Recall, the b's are not independent.

Consequently the standard error will involve the covariances among b's as well as their variances.

$$\hat{\mu}_{Y.123} = x_0' \times b$$

Note:

- Where $x_0' = [1, 2, 0.3, 0.8]$ and

$$var(\hat{\mu}_{Y.123}) = x_0' \cdot (X'X)^{-1} \cdot x_0 \cdot \sigma^2$$

Standard Error of Estimate

$$x_0' (X'X)^{-1} x_0 = \begin{vmatrix} 1 & 2 & 0.3 & 0.8 \end{vmatrix} \times \begin{vmatrix} 0.8840 & -0.3332 & -0.7556 & -0.0322 \\ -0.3332 & 0.3629 & -0.9578 & -0.1507 \\ -0.7556 & -0.9578 & 18.8388 & -2.8209 \\ -0.0322 & -0.1507 & -2.8209 & 1.3449 \end{vmatrix} \times \begin{vmatrix} 1 \\ 2 \\ 0.3 \\ 0.8 \end{vmatrix} = 0.06869$$

Therefore:

Hence, a 95% confidence interval can be constructed for the true contractility with Cl⁻=2, PO₄=0.3, and K⁺=0.8 as :

$$228.3 \pm (2.074)(9.973) = (207.6, 249.0) \text{ (Student's-t with df = 22)}$$

An Introduction To Statistical Design: Terminology

ANOVA = analysis of variance

rmANOVA = repeated measures ANOVA

CRD = completely randomized design

RCBD = randomized complete block design

ANCOVA = analysis of covariance

SPD = split plot design

FACTORIAL designs

MANOVA = multivariate analysis of variance
etc. etc. etc.....

ANOVA Part 1: One way

testing to see whether many means come from the same population

Goal

- Determine likely values of measure if samples in each group are from the same Population
- Develop a measure for the difference between experimental groups based on the means and using the estimate of variability for scaling of the difference

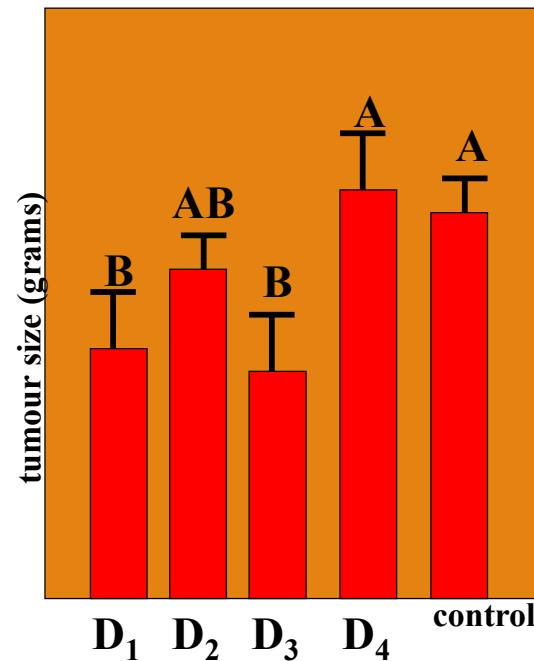
Example:

- 4 new anticancer drugs compared against a control

Null Hypothesis:

Alternative:

- letters denote treatments that are significantly different from one another.



Use F test for Multiple Groups

Is there a difference between m groups of n samples each? (also sometimes i and j, or t and r).

To verify if all groups are from the same population one guesses that they are actually identical and validates or invalidates the statement.

If the hypothesis is true:

- The average variance of the individual groups should be smaller or equal to the variance of a given population.

Completely Randomized Design (CRD)

- treatments are randomly assigned, completely at random to the experimental units, which are assumed to be homogeneous.
- Model 1: Fixed Effects Case (examine effects of treatments)
- Model 2: Random Effects Case (identify sources of variation)

	Treatments				
	T_1	T_2	T_3	T_4	T_5
Replicates	1.1	3.2	5.6	8.2	8.3
	1.7	3.4	4.8	7.6	8.0
	1.5	3.0	5.1	7.9	8.4

matrix notation goes from larger to smaller groups e.g. block, treatment, replicate, etc.

Model I: Fixed Model

Example: Want to examine the effect of 4 diets on the growth of rainbow trout fry.

		<i>Dietary Treatment</i>						
		1	2	3	4			
Replicates (mass in grams)		59	70	93	124	$i = 4$ treatments		
		47	59	85	135	$j = 5$ replicates		
		40	52	79	167			
		32	87	72	83			
		39	61	88	152			
		$\sum_j Y_{ij} = Y_{i\bullet} =$	217	329	417	661	<i>totals</i>	
		$\sum Y_{ij}^2 =$	9835	22375	35043	91483	$1624 = Y_{\bullet\bullet}$	
		$Y_{i\bullet} =$	43.4	65.8	83.4	132.2	$158736 = \bar{Y}_{\bullet\bullet}$	
							81.2	

Model I: Fixed Model

Compute Sums
of Squares:

$$\text{Total}(SS) = \sum Y_{ij}^2 - \frac{(Y_{\bullet\bullet})^2}{i \cdot j} =$$

$$\text{SS}_{(\text{Treatments})} = SS(T) = \frac{\sum (Y_{i\bullet})^2}{j} - \frac{(Y_{\bullet\bullet})^2}{i \cdot j} =$$

$$\text{SS(error)} = \text{SS}(E) = \text{Total}(SS) - \text{SS}(T)$$

Model I: Fixed Model

Source	df	SS	MS	F_c	$F_{i,j,\alpha}$
treatment	3	21359.2	7119.73	20.68	3.24
error	16	5508.0	344.25	-----	-----
TOTAL	19	26867.2	-----	-----	-----

$$df_{\text{treatment}} = (\# \text{ treatments} - 1) = (i - 1)$$

$$df_{\text{TOTAL}} = (i \times j) - 1$$

$$df_{\text{error}} = df_{\text{TOTAL}} - df_{\text{treatment}}$$

$$MS = SS / df$$

$$F_c = T(MS) / E(MS)$$

$F_{i,j,\alpha}$ = from table
e.g. $F_{3,16,0.05} = 3.24$

Model I: Fixed Model

The Bottom line.....

Since $F_{\text{C}} > F_{\text{table}}$, reject H_0 .

Therefore:

- there is a difference between one or more treatments
- Following ANOVA we would need to explore the differences (more on this soon...)

Model II: Random Effects Model

- not interested in specific treatments, but rather on sources of variation.

e.g. In order to study the sources of variation in synthesis of protein gumbycin, a sample of 5 cell cultures was selected at random from an incubator by a chemical engineer. A total of 4 western blots for the protein were made on each of the 5 randomly selected cultures.

	1	2	3	4	5	= culture number
protein content (ng/ 10^6 cells)	85	62	46	67	54	
	81	67	52	57	72	$i = 5$
	83	61	55	65	68	$j = 4$
	76	58	41	54	45	$Y_{..} = 1249$
						$\sum Y_{ij}^2 = 81023$

Model II: Random Effects Model

Null Hypothesis:

- there is no variation between cultures i.e. $H_0: \sigma^2_t = 0$

Calculation of ANOVA table is exactly the same:

Source	df	SS	MS	F_c	$F_{ij, \alpha}$
Between	4	2233.7	558.425	10.61	3.06
within	15	789.25	52.617	-----	-----
TOTAL	19	3022.95	-----	-----	-----

Model I vs Model II?

What are we estimating?

- the MS(T) is an estimate of :
- Model 1: Fixed Effects Case (examine effects of treatments)
- Model 2: Random Effects Case (identify sources of variation)

	Model I	Model II
treatment (between)	$\sigma^2 + \frac{j}{i-1} \sum \tau_i^2$	$\sigma^2 + j\sigma_\tau^2$
error (within)	σ^2	σ^2

Model I vs Model II?

- in the protein case $F_C > F_{table}$.
- therefore σ^2 is significantly different from zero, and hence there is significant variability between cell cultures with respect to protein gumbycin synthesis.
- can σ^2 be estimated (i.e. variance between cultures) ? YES :

$$s^2 = \sigma^2 = \frac{MS(T) - MS(E)}{j} =$$

Model I vs Model II?

- in this protein experiment the ratio of σ_t^2 to σ^2 = 126.452 : 52.616 or about 2.4 to 1

What percentage of the total variation does the variation between (σ_t^2) cultures account for ?

Real life experiments

- what can be done when there is unequal replication ?
- for example, consider an experiment to assess anti-carcinogens.
 - Rats were pre-medicated with one of 5 anticarcinogens prior to being given a single dose of benzo[a]pyrene.
 - The next day you want to assess mass of feed eaten.
 - However, for whatever reason many of the rats have died !!

	1	2	3	4	5	= treatment
mass of feed eaten (g) for each rat	8.4	6.5	7.2	7.2	7.9	
8.2	7.6	8.1	7.4	7.5	9.6	
8.2	7.7	6.2	–	–	9.9	
7.4	–	6.6	–	–	–	
8.2	–	–	–	–	–	

Real Life Experiments

Source	df	SS	MS	F_c	$F_{i,j,\alpha}$
treatment	4	9.8208	2.4552	5.435	3.26
error	12	5.4203	0.4517	-----	-----
TOTAL	16	15.2412	-----	-----	-----

$$df_{TOTAL} = n - 1 \text{ (i.e. total samples - 1)}$$

except...

$$df_{error} = n - i \text{ (i.e. total samples - # treatments)}$$

$$df_{treatment} = i - 1 \text{ (i.e. total treatments - 1)}$$

	Model I	Model II
treatment	$\sigma^2 + \frac{j}{i-1} \sum \tau_i^2$	$\sigma^2 + j\sigma_\tau^2$
error	σ^2	σ^2
treatment	$\sigma^2 + \frac{\sum j_i(\tau_i)^2}{i-1}$	$\sigma^2 + \frac{n - (\sum j^2)}{i-1} \sigma_\tau^2$
error	σ^2	σ^2

Multiple Comparisons: Post hoc Procedures

What happens if you reject H₀ ?

- need to explore where the differences lie and their magnitudes.

Methods will vary in conservatism

Repeated t-tests can result in errors so we need other methods

Statistical Testing of Means

- 1) Student's t-test (2 means)
- 2) Least Significant Difference (lsd)
- 3) Duncan's new multiple range test
- 4) Contrast analysis
- 5) Scheffé Test

There are many others you can investigate on your own time.

e.g. SNK (Student-Neumann Keuls) test

Model I: Fixed Model

Example: Want to examine the effect of 4 diets on the growth of rainbow trout fry.

		Dietary Treatment				
		1	2	3	4	
Replicates (mass in grams)		59	70	93	124	$i = 4$ treatments
		47	59	85	135	$j = 5$ replicates
		40	52	79	167	
		32	87	72	83	
		39	61	88	152	
		$\sum_j Y_{ij} = Y_{i\bullet} =$	217	329	417	661
		$\sum Y_{ij}^2 =$	9835	22375	35043	91483
		$Y_{i\bullet} =$	43.4	65.8	83.4	132.2
		<i>totals</i>				
		1624				$= Y_{\bullet\bullet}$
		158736				$= \bar{Y}_{\bullet\bullet}$
		81.2				

Model I: Fixed Model

Compute Sums
of Squares:

$$\text{Total}(SS) = \sum Y_{ij}^2 - \frac{(Y_{\bullet\bullet})^2}{i \cdot j} = 158736 - \frac{1624^2}{4 \times 5} = 26867.2$$

$$\text{SS}_{(\text{Treatments})} = SS(T) = \frac{\sum (Y_{i\bullet})^2}{j} - \frac{(Y_{\bullet\bullet})^2}{i \cdot j} = \frac{217^2 + 329^2 + 417^2 + 661^2}{5} - \frac{1624^2}{4 \times 5} = 21359.2$$

$$\text{SS(error)} = \text{SS}(E) = \text{Total}(SS) - \text{SS}(T)$$

Model I: Fixed Model

Source	df	SS	MS	F_c	$F_{i,j,\alpha}$
treatment	3	21359.2	7119.73	20.68	3.24
error	16	5508.0	344.25	-----	-----
TOTAL	19	26867.2	-----	-----	-----

$$df_{\text{treatment}} = (\# \text{ treatments} - 1) = (i - 1)$$

$$df_{\text{TOTAL}} = (i \times j) - 1$$

$$df_{\text{error}} = df_{\text{TOTAL}} - df_{\text{treatment}}$$

$$MS = SS / df$$

$$F_c = T(MS) / E(MS)$$

$F_{i,j,\alpha}$ = from table
e.g. $F_{3,16,0.05} = 3.24$

Least Significant Difference (LSD)

- examines differences between means
- ideally this is used for planned comparisons (i.e. specify in advance of getting the data.)

The equation for the standard error of the difference between 2 means is:

$$SE = \sqrt{\left(\bar{Y}_{i_a} - \bar{Y}_{i_b} \right)^2} = \sqrt{E(MS) \cdot \frac{2}{j}} =$$

$$lsd = t_{v, \alpha/2} \cdot \sqrt{E(MS) \cdot \frac{2}{j}} =$$

- where $v = df_{E(MS)} = i(j-1) = 4(5-1) = 16$

Least Significant Difference (LSD)

If LSD < (difference between 2 means), then reject H_0

- i.e. the means are significantly different (* = sig. different)

NOTE: 4/6 possible combinations were declared as significant different (i.e. 66%).

Difference Table: (24.877)

Diets	2(65.8)	3(83.4)	4(132.2)
1(43.4)	22.4	40.0*	88.8*
2(65.8)	-----	17.6	66.4*
3(83.4)	-----	-----	48.8*

Least Significant Difference (LSD)

Underscore Representation

- Underline pairs of means that are NOT significantly different.

1(43.4) 2(65.8) 3(83.4) 4(132.2)

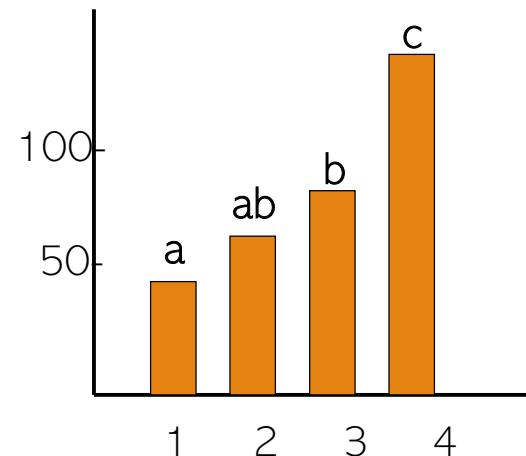
LSD results :

Diets 1 & 2, and diets 2&3 are not significantly different from each other.

However, diets 1&3, 1&4, 2&4, and 3&4 are significantly different.

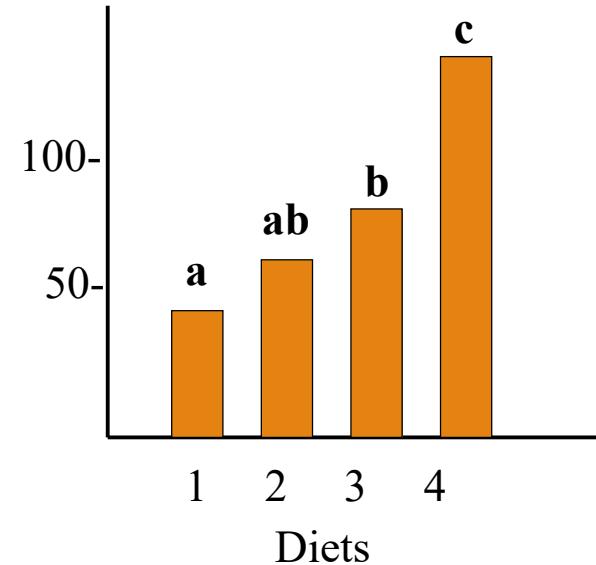
Graphical Representation

same letter = *not* significantly different



Some Notes About the Results:

- Which diet would you use if you wished to raise the largest fish in the shortest period of time ?
- If diet 3 cost much more than diet 2, which yields a greater weight gain per unit dollar of expenditure ?
- Suppose Diets 1 and 2 are equally priced. If they are the only 2 available diets which should be used ?



Duncan's Multiple Range Method

- differs from LSD method which has a single 'least significant difference' with which to compare treatment effects.
- the Duncan method employs test criteria which vary in magnitude, depending on the number of means involved in the test.

First compute the standard error of a sample mean:

$$SE(\bar{Y}) = \sqrt{\frac{MS(E)}{j}} =$$

Duncan's Multiple Range Method

- consult table to determine the values of Studentized range: $q_\alpha(p, f_e)$
 - q_α = significance level (e.g. $\alpha = 0.05$)
 - p = number of means (i.e. treatments) being tested
 - f_e = the number of degrees of freedom of experimental error
 - (i.e. $i(j-1) = 16$, in this example)

$$R_p = q_\alpha(p, f_e) \cdot \sqrt{\frac{MS(E)}{j}}$$

Decision Rule:

- if the difference in means is greater than the calculated significant ranges parameter, R_p , then it is declared significant.

Duncan's Multiple Range Method

$$\text{SE}(Y) = 8.29759 \quad R_p = q_\alpha(p, f_e) \cdot \sqrt{\frac{MS(E)}{j}}$$

p	2	3	4
$q_\alpha(p, f_e)$	$q_{0.05}(2, 16) = 3.00$	$q_{0.05}(3, 16) = 3.15$	$q_{0.05}(4, 16) = 3.23$
R_p	$3.00(8.29759) = 24.893$	$3.15(8.29759) = 26.137$	$3.23(8.29759) = 26.801$

Difference Table

Diets	2(65.8)	3(83.4)	4(132.2)
1(43.4)	22.4	40.0*	88.8*
2(65.8)	-----	17.6	66.4*
3(83.4)	-----	-----	48.8*

Duncan's Multiple Range Method

Notice, in this particular circumstance, Duncan's and lsd result in the same conclusions for the data.

- This can differ !!

If the difference in means is particularly close then the lsd method will result in significance more often (i.e. is less conservative than Duncan's).

Linear Combinations of Means (i.e. contrasts)

- suppose there are a number of ways to improve the length of time of musculoskeletal repair after injury:

X1 = blueberry and kale smoothies

X2 = supplements of branched chain AAs + proline

X3 = static magnetic field (B0)

X4 = pulsed magnetic field (B0)

X5 = Therapeutic Ultrasound (US)

Example of a linear combination (2 sample case):

$$L_1 : \bar{X}_1 - \bar{X}_2$$

- estimates $\mu_1 - \mu_2$ or the difference between dietary treatments

Linear Combinations of Means

Another possible linear combination:

$$L_2 : \bar{X}_3 - \bar{X}_4$$

- measures mean difference in time of healing for magnetic field (B_0) methods.

While:

$$L_3 : \frac{1}{2} \bar{X}_3 + \frac{1}{2} \bar{X}_4$$

- i.e. an estimate of $\frac{1}{2}(\mu_3 + \mu_4)$
- measures the average time to achieve full repair due to magnets.

Linear Combinations of Means

The Question:

- does US take less time, on average, than magnetic fields (ignoring any difference between pulsed vs. static BO)?

Can be written mathematically:

$$\mu_5 < \frac{1}{2}(\mu_3 + \mu_4) ?$$

Alternatively:

$$\mu_5 - \frac{1}{2}(\mu_3 + \mu_4) < 0 ?$$

To estimate this difference among mean values we would use:

$$L_4 : \bar{X}_5 - \frac{1}{2}(\bar{X}_3 + \bar{X}_4)$$

Then test whether L_4 was significantly different from 0.

Linear Combinations of Means

More specifically in this case we would be doing a 1-sided test, since we are only looking at whether there is evidence that μ_5 is less than the average of μ_3 and μ_4 .

To compare diet with the average of 'engineering' methods we would ask whether the average repair times due to diet, estimated by:

$$\frac{1}{2} \bar{X}_1 + \frac{1}{2} \bar{X}_2$$

differs from the average times taken when using tech, estimated by:

$$\frac{1}{3} \bar{X}_3 + \frac{1}{3} \bar{X}_4 + \frac{1}{3} \bar{X}_5$$

Linear Combinations of Means

The difference between these 2 estimates:

$$L_5 : \frac{1}{2} \bar{X}_1 + \frac{1}{2} \bar{X}_2 - \frac{1}{3} \bar{X}_3 - \frac{1}{3} \bar{X}_4 - \frac{1}{3} \bar{X}_5$$

estimates:

$$\frac{1}{2} \mu_1 + \frac{1}{2} \mu_2 - \frac{1}{3} \mu_3 - \frac{1}{3} \mu_4 - \frac{1}{3} \mu_5$$

Linear Combinations of Means

To reiterate, here are the linear combinations again:

$$L_1 : \bar{X}_1 - \bar{X}_2$$

$$L_5 : \frac{1}{2}\bar{X}_1 + \frac{1}{2}\bar{X}_2 - \frac{1}{3}\bar{X}_3 - \frac{1}{3}\bar{X}_4 - \frac{1}{3}\bar{X}_5$$

$$L_2 : \bar{X}_3 - \bar{X}_4$$

$$L_3 : \frac{1}{2}\bar{X}_3 + \frac{1}{2}\bar{X}_4$$

$$L_4 : \bar{X}_5 - \frac{1}{2}(\bar{X}_3 + \bar{X}_4)$$

Linear Combinations of Means

Now here are the same linear contrasts in tabular form:

Means	\bar{X}_1	\bar{X}_2	\bar{X}_3	\bar{X}_4	\bar{X}_5	
Coefficients:	λ_1	λ_2	λ_3	λ_4	λ_5	$\Sigma\lambda$
Combination						
L ₁	+1	-1	0	0	0	0
L ₂	0	0	+1	-1	0	0
L ₃	0	0	+1/2	+1/2	0	1
L ₄	0	0	-1/2	-1/2	+1	0
L ₅	+1/2	+1/2	-1/3	-1/3	-1/3	0

Linear Combinations of Means

Definition:

- Linear combinations with $\sum \lambda = 0$ are called contrasts.
- A sample contrast, denoted L, is an estimator of the population contrast.
- The Standard Error of this estimate is:

$$SE(L) = SE\left(\sum \lambda_i \bar{X}_i\right) = \sqrt{s^2 \cdot \frac{\left(\sum \lambda_i^2\right)}{n_i}}$$

Where s^2 is the MS(E) from the ANOVA table, $\sum \lambda_i^2$ is the coefficient sum of squares, and n_i is the number of samples in the i^{th} group.

Linear Combinations of Means

If, for example, each $n_i = 5$, then the standard errors for the different linear combinations are:

$$SE(L) = \sqrt{s^2 \cdot \frac{(\sum \lambda_i^2)}{n_i}}$$

Combination	$\Sigma \lambda^2$	SE(L)
L1	2.000	$s \cdot \sqrt{2.000/5}$
L2	2.000	$s \cdot \sqrt{2.000/5}$
L3	0.500	$s \cdot \sqrt{0.500/5}$
L4	1.500	$s \cdot \sqrt{1.500/5}$
L5	0.833	$s \cdot \sqrt{0.833/5}$

Hypothesis Testing Using Contrasts:

Consider L4, a comparison of ultrasound with the average of magnetic field induced repair times:

$$L_4 : \bar{X}_5 - \frac{1}{2}(\bar{X}_3 + \bar{X}_4)$$

The $\Sigma \lambda^2 = 1.500$. If all 3 means are based on samples of 5 times each, then:

$$SE(L) = \sqrt{s^2 \cdot \frac{(\sum \lambda_i^2)}{n_i}} =$$

Hypothesis Testing Using Contrasts:

If s = pooled estimate of the population standard deviation (σ), based on 5 pooled variances $s_1^2, s_2^2, s_3^2, s_4^2, s_5^2$, each with 4 df

- then s will have 5×4 (i.e. 20) degrees of freedom.

Using a Student's t-table: $t_{\alpha=0.05,20} = 1.725$; $t_{\alpha=0.025,20} = 2.086$

The null hypothesis to be tested:

$$H_0 : \mu_5 - \frac{1}{2}(\mu_3 + \mu_4) = 0$$

Which can also be re-written as:

$$H_0 : \mu_5 = \frac{1}{2}(\mu_3 + \mu_4)$$

Hypothesis Testing Using Contrasts:

Test H_0 against one of the alternative hypothesis:

$$H_A : \mu_5 \neq \frac{1}{2}(\mu_3 + \mu_4) \quad (\text{2-sided alternative})$$

$$H_A : \mu_5 > \frac{1}{2}(\mu_3 + \mu_4) \quad (\text{1-sided alternative})$$

$$H_A : \mu_5 < \frac{1}{2}(\mu_3 + \mu_4) \quad (\text{another 1-sided alternative})$$

Hypothesis Testing Using Contrasts:

A t-test can be used here:

$$t = \frac{\text{Estimated Value} - \text{Hypothesized True Value}}{\text{Standard Error of Estimated Value}} \quad t = \frac{(\bar{X}_5 - \frac{1}{2}(\bar{X}_3 + \bar{X}_4)) - 0}{0.5477 \cdot s}$$

Depending on whether we have chosen a 1 or 2 sided test we would reject H₀ in favour of the selected alternative hypothesis:

- if $|t| > 2.086$ (i.e. $t>2.086$ or $t<-2.086$) for the 2 sided test.
- if $t > +1.725$ for the 1 sided test for the alternative which predicted $\mu_5 > 1/2(\mu_3 + \mu_4)$.
- if $t < -1.725$ for the 1 sided test for the alternative which predicted $\mu_5 < 1/2(\mu_3 + \mu_4)$.

Hypothesis Testing Using Contrasts:

The 95% confidence interval for a linear combination:

(Estimated value – $t_{\alpha/2, df} \times \text{stderr}$ of estimated value) < Linear combination of true values < (Estimated value + $t_{\alpha/2, df} \times \text{stderr}$ of estimated value)

In this case:

$$\left(\bar{X}_5 - \frac{1}{2}(\bar{X}_3 + \bar{X}_4) \right) - (2.086 \cdot 0.5477s) \leq \left(\mu_5 - \frac{1}{2}(\mu_3 + \mu_4) \right) \leq \left(\bar{X}_5 - \frac{1}{2}(\bar{X}_3 + \bar{X}_4) \right) + (2.086 \cdot 0.5477s)$$

Orthogonal Contrasts:

Means	\bar{X}_1	\bar{X}_2	\bar{X}_3	\bar{X}_4	\bar{X}_5	
Coefficients:	λ_1	λ_2	λ_3	λ_4	λ_5	$\Sigma\lambda$
Combination						
L ₁	+1	-1	0	0	0	0
L ₂	0	0	+1	-1	0	0
L ₄	0	0	-1/2	-1/2	+1	0
L ₅	+1/2	+1/2	-1/3	-1/3	-1/3	0

- Notice each contrast looks at a different characteristic of the data
- Not all contrasts look at genuinely different characteristics !!

Hypothesis Testing Using Contrasts:

For example, the contrasts:

$$\bar{X}_1 - \bar{X}_2 \quad \bar{X}_1 - \bar{X}_3 \quad \bar{X}_2 - \bar{X}_3$$

compare 1 with 2, 1 with 3, and 2 with 3. The third contrast, however, really tells us nothing we couldn't have figured out with the other two, since:

$$\bar{X}_2 - \bar{X}_3 = \bar{X}_1 - \bar{X}_3 - (\bar{X}_1 - \bar{X}_2)$$

A way to ensure that contrasts are looking at completely different aspects of the data is to require that all contrasts be orthogonal.

Hypothesis Testing Using Contrasts:

The numerical verification that 2 contrasts are orthogonal is that the sum of the products of their corresponding coefficients is zero.

For example, L_1 and L_2 are orthogonal:

L1	+1	-1	0	0	0	
L2	0	0	+1	-1	0	
Products	0	0	0	0	0	Sum = 0

Also, L_4 and L_5 are orthogonal:

L4	0	0	-1/2	-1/2	1	
L5	+1/2	+1/2	-1/3	-1/3	-1/3	
Products	0	0	1/6	1/6	-1/3	Sum = 0

Hypothesis Testing Using Contrasts:

However, $X_1 - X_2$ and $X_1 - X_3$ are not orthogonal:

X1-X2	+1	-1	0	0	0	
X1-X3	+1	0	-1	0	0	
Products	+1	0	0	0	0	Sum = +1

i.e. if 2 contrasts, with q number of coefficients:

$$\sum_{i=1}^q \lambda_{A_i} \lambda_{B_i} = 0$$

Then contrasts A and B are orthogonal.

Non-orthogonal contrasts do not provide any extra information !!

Hypothesis Testing Using Contrasts:

It's a bit tricky at first to come up with orthogonal contrasts. The best thing to do is think up contrasts which address specific and distinct questions- Then check for orthogonality.

e.g. How do 1 and 2 compare ?

How do 3 and 4 compare ?

How do 3 and 4 compare with 5 ?

How do 1 and 2 together compare with 3, 4, and 5 together ?

→ These questions led to the orthogonal contrasts L1, L2, L3, and L5

Hypothesis Testing Using Contrasts:

NOTES:

- if we have i treatments then there exactly $i-1$ possible orthogonal contrasts.
- the $i-1$ is exactly equal to the df for measuring variability among the treatment means.
- These orthogonal contrasts correspond to a decomposing of this variability !!
- each contrast has 1 df associated with it.

Sum of Squares of Contrasts:

- a measure of a size of a contrast is it's sum of squares i.e. $SS(L)$

$$SS(L) = \frac{n \times (\text{estimated value of } L)^2}{\sum \lambda_i^2}$$

- the estimated value of L is calculated usually using mean values.
- Totals can also be used (and are equivalent)

$$L_{_1}^{Totalbased} : \bar{X}_{1\bullet} + \bar{X}_{2\bullet}$$

Where $\bar{X}_{1\bullet}$ is the total of all observations taken for the first treatment

$$SS(L) = \frac{(\text{estimated value of } L^{Totalbased})^2}{n \times \sum \lambda_i^2}$$

Back to the fish example

We'd like to test the following:

$$\text{for } L_1 \quad H_0 : \mu_4 - \mu_2$$

$$\text{for } L_2 \quad H_0 : \mu_4 - 1/2(\mu_2 + \mu_3)$$

Note- try it yourself !! Check that L_1 and L_2 are orthogonal !!

Fish Example Contrasts

Contrast L_1 : $\hat{L}_1 = \bar{X}_4 - \bar{X}_2 = 132.2 - 65.8 = 66.4$

The coefficients are: $\lambda_1 = 0$; $\lambda_2 = -1$; $\lambda_3 = 0$; $\lambda_4 = +1$

The sum of squares of $\Sigma\lambda$

$$SE(L_1) = \sqrt{s^2 \cdot \frac{(\sum \lambda^2)}{n_i}} = t = \frac{\hat{L}_1 - 0}{11.73} =$$

From t-table, with df = 16, t = 2.120 (2 tailed)

Since $t_{\text{calc}} > t_{\text{table}}$ reject H_0 i.e. diets 4 and 2 are significantly different

Fish Example Contrasts

Similarly, Contrast L_2 :

$$\hat{L}_2 = \bar{X}_4 - \frac{1}{2}(\bar{X}_2 - \bar{X}_3) = 132.2 - \frac{1}{2}(65.8 - 83.4)$$

The coefficients are: $\lambda_1 = 0$; $\lambda_2 = -1/2$; $\lambda_3 = -1/2$; $\lambda_4 = +1$

The sum of squares of $\Sigma\lambda$ (i.e. $\Sigma\lambda^2$) = 1.5

$$SE(L_1) = \sqrt{s^2 \cdot \frac{(\sum \lambda^2)}{n_i}} = \sqrt{344.25 \cdot \frac{1.5}{5}} = 10.16$$

$t_{\text{calc}} = 6.535$

More Experimental Designs

- CRD with subsampling
- Randomized Complete Block Design (RCBD)
- Analysis of Covariance (ANCOVA)

Subsampling

- the term used to describe the situation in which more than one observation is taken per experimental unit.
- such observations are made on sampling units.
- when subsampling is performed, the linear model and the ANOVA must be expanded to take into account the variation among samples (the source of sampling error)

CRD with Subsampling: Model I (Treatment Effects)

Consider the following experiment:

- A new drug phenphodyne-HCl, was thought to enhance liver cyt-P450 in people with late-stage cirrhosis
- when given with alcohol the effect was thought to be diminished.
- a total of 6 randomly assigned patients were used to test this drug:
 - 2 controls (no drug)
 - 2 phenphodyne-HCl
 - 2 phenphodyne-HCl + ethanol
- After an appropriate time and dose, 4 liver biopsies were taken under ultrasound/MRI (co-registered) guidance from each patient and cyt-P450 was measured in piece.



82



(Source: <http://www.skills.uct.ac.za/activities.htm>)

Experimental Objectives:

- 1) To determine if there was a significant difference among the three treatments
- 2) To estimate the 2 variance components:
 - variation among measurements within patients
(i.e. sampling error).
 - variation among patients within a given treatment
(i.e. experimental error).

CRD with Subsampling: Model I

The data ($Y_{i,j,k}$) from 6 randomly chosen/assigned patients and 4 randomly selected pieces of liver from each.

Patients cyt-P450 readings	Control		Phelphodyne		Phelphodyne +EtOH	
	1	2	1	2	1	2
131	148	157	152	124	140	
130	143	153	155	125	138	
125	150	154	162	136	138	
131	150	149	161	130	139	
517	591	613	630	515	555	
	1108		1243		1070	

$$Y_{ij\cdot}$$

$$\sum Y_{ij\cdot}^2 = 1962489$$

$$Y_{i..}$$

$$\sum Y_{i..}^2 = 3917613$$

$$Y_{...} = 3421$$

CRD with Subsampling: Model I

Notes:

- an experimental unit here is a patient
- a sampling unit is a piece of patient liver

Degrees of Freedom

- $i = 3$ (treatments)
- $j = 2$ (replicates)
- $k = 4$ (subsamples)
- Total df = $(3 \times 2 \times 4) - 1 = 24 - 1 = 23$

There are 2 hypothesis that can be tested:

- 1) $H_0: \tau_i = 0$, for all i vs. $H_A: \tau_i \neq 0$ (i.e. treatment effect)
- 2) $H_0: \sigma_e^2 = 0$ vs. $H_A: \sigma_e^2 \neq 0$ (i.e. error variance)

Compute Sums of Squares:

$$Total(SS) = \sum Y_{ijk}^2 - \frac{(Y_{\dots\dots})^2}{i \cdot j \cdot k} =$$

$$SS(\text{Treatments}) = SS(T) = \frac{\sum (Y_{i\dots\dots})^2}{j \cdot k} - \frac{(Y_{\dots\dots})^2}{i \cdot j \cdot k} =$$

$$SS(\text{Subsamples}) = SS(SS) = \frac{\sum Y_{ij\dots}^2}{k} - \frac{(Y_{\dots\dots})^2}{i \cdot j \cdot k} =$$

$$SS(\text{sampling error}) = SS(SE) = Total(SS) - SS(SS) = 252.75$$

$$SS(\text{experimental error}) = SS(EE) = SS(SS) - SS(T) = 920.625$$

ANOVA Table

Source	df	SS	MS	Fc	F _{i,j,α}
subsamples (AMONG PATIENTS)	5	2987.2083	-----	-----	-----
Treatment	2	2066.5833	1033.2417	3.367	9.55
Exp. Error	3	920.625	306.875	21.85	3.16
Samp. Error	18	252.75	14.042	-----	-----
TOTAL	23	3239.9583	-----	-----	-----

*Note treatment df
+ EE df =
subsample df

$$df_{TOTAL} = (i \cdot j \cdot k) - 1 = (3 \times 2 \times 4) - 1 = 23$$

$$df_{SS} = (i \cdot j) - 1 = (3 \times 2) - 1 = 5$$

$$df_{SE} = i \cdot j \cdot (k-1) = 3 \times 2 \times (4-1) = 18$$

$$df_T = i - 1 = 3 - 1 = 2$$

$$df_{EE} = i(j-1) = 3(2-1) = 3$$

ANOVA Analysis

$$F_C \text{ (among treatments)} = 1033.2417 / 306.875 = 3.367$$

Since $F_{0.05,2,3} = 9.55$

we $H_0: \tau_i = 0$. Therefore there is effect of treatment.

$$F_C \text{ (experimental error)} = 306.875 / 14.042 = 21.854$$

Since $F_{0.05,3,18} = 3.16$

We $H_0: \sigma_e^2 = 0$. Therefore there is significant source of error between patients.

REMEMBER CRD what MS is estimating !!

	Model I	Model II
treatment	$\sigma^2 + \frac{j}{i-1} \sum \tau_i^2$	$\sigma^2 + j\sigma_\tau^2$
error	σ^2	σ^2
treatment	$\sigma^2 + \frac{\sum j_i (\tau_i)^2}{i-1}$	$\sigma^2 + \frac{n - (\sum j^2)/n}{i-1} \sigma_\tau^2$
error	σ^2	σ^2

CRD with Subampling

For Fixed (Type I) Models, what does the mean (MS) estimate?

ANOVA Table

$$E[MS(T)] = \sigma^2 + k\sigma_\varepsilon^2 + \frac{j \cdot k}{i-1} \sum \tau_i^2$$

$$E[MS(EE)] = \sigma^2 + k\sigma_\varepsilon^2$$

$$E[MS(SE)] = \sigma^2$$

Estimation of variance components:

$$\hat{\sigma}_\varepsilon^2 = s_\varepsilon^2 = (306.875 - 14.042)/4 = 73.208$$

$$\hat{\sigma}^2 = s^2 = 14.042 \quad s^2 + s_\varepsilon^2 = 87.250$$

Where is variance?

σ^2 = variation among cyt-P₄₅₀ within patients

s_e^2 = variation in patients within treatments

$\sigma^2 + s_e^2$ = total variance (within and among patients within a treatment)

The variation within patients (i.e. among liver samples) accounts for $100 \times (14.042 / 87.250) = 16.1\%$ of the estimated total variance.

The variation among patients of a given treatment accounts for $100 \times (73.208 / 87.250) = 83.9\%$ of the estimated total variance.

i.e. There is approximately 5.1x as much variation among patients as there is within patients.

CRB SS NOTES:

- The critical F Value (FC) value for treatments was noted to be quite small. This is due to the large denominator
- The large MS(EE) indicates high variability among the patients for any one preparation.
- Although no differences in treatments were detected, any differences may actually have been hidden by the variation among experimental units.
- Future experiments should take into account heterogeneity of experimental units.
Also could performing a block design in which 2 blocks of 3 patients are examined too.

CRD with Subsampling (Model II- Random Effects)

Consider this experiment:

- an experiment was conducted to assess the precision to which EEG could be measured for application in a brain-computer interface.
 - 4 people were randomly selected
 - 3 brain regions were randomly chosen from each person
 - 2 samples taken per region for signal power analysis

Primary Objectives:

- 1) Estimate EEG signal power
- 2) Find the EEG signal power standard error
- 3) Isolate and estimate the sources of variation.

CRD with SS (Model II-Random Effects)

From this analysis:

- recommendation can be made with regard to optimizing future brain EEG sampling strategies.
- This would allow the researcher to reduce the standard error of their estimate in future studies.

Person	1			2			3			4		
Brian Region	1	2	3	1	2	3	1	2	3	1	2	3
Subsample	3.48	3.72	3.03	2.66	2.07	2.39	2.97	3.94	2.75	3.98	4.27	3.51
	3.29	3.68	3	2.64	2.12	2.39	2.86	3.64	2.75	4.07	4.32	3.51
	6.77	7.4	6.03	5.3	4.19	4.78	5.83	7.58	5.5	8.05	8.59	7.02
	20.2			14.27			18.91			23.66		
$Y_{ij\cdot}$												
$Y_{i\cdot\cdot}$												
$Y_{\cdot\cdot\cdot} = 77.09$												

$i = 4$
 $j = 3$
 $k = 2$
 $\text{total df} = (ijk) - 1 = 23$

CRD II SS Model Equation:

$$Y_{ijk} = \mu + \tau_i + \varepsilon_{ij} + \delta_{ijk}$$

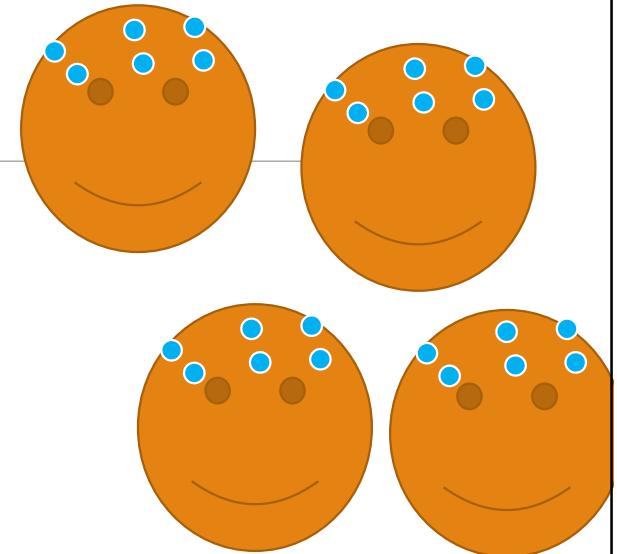
Treatment Effect

Experimental Error

Sampling Error

Necessary assumptions:

- τ_i are $N(0, \sigma_\tau^2)$, where σ_τ^2 is the variation between people
- ε_{ij} are $N(0, \sigma_\varepsilon^2)$, where σ_ε^2 is the variation between brain areas
- δ_{ijk} are $N(0, \sigma_\delta^2)$, and σ_δ^2 is variation among samples between brain areas.



CRD II SS ANOVA

Source	df	SS	MS	Fc	$F_{ij,\alpha}$
subsamples (AMONG AREAS)	11	10.19055	-----	-----	-----
Treatment (AMONG PEOPLE)	3	7.56035	2.5201167	7.665	4.07
Exp. Error	8	2.6302	0.320775	49.41	2.85
Samp. Error	12	0.07985	0.0066542	-----	-----
TOTAL	23	10.2704	-----	-----	-----

*Note treatment df +
EE df = subsamp df

Testable hypothesis

$H_0: \sigma_\tau^2 = 0$ (i.e. no significant difference between people)

$H_0: \sigma_\varepsilon^2 = 0$ (i.e. no significant difference between brain areas)

CRD II SS ANOVA Results:

- variance component due to brain areas within people is significantly different from 0?
- variance component from person to person is significantly different from 0?

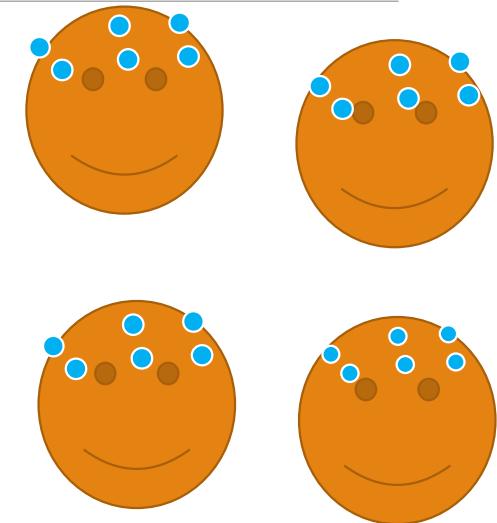
In CRD II with Subsampling

Random Effects (Type II) Models: what does the mean (MS) estimate?

$$E[MS(T)] = \sigma_\delta^2 + k\sigma_\varepsilon^2 + jk\sigma_\tau^2$$

$$E[MS(EE)] = \sigma_\delta^2 + k\sigma_\varepsilon^2$$

$$E[MS(SE)] = \sigma_\delta^2$$



Estimates of the 3 variance components:

$$\hat{\sigma}_{\delta}^2 = s^2 = MS(SE) = 0.0066$$

$$\hat{\sigma}_{\varepsilon}^2 = s_{\varepsilon}^2 = \frac{MS(EE) - MS(SE)}{k} = \frac{0.328775 - 0.0066542}{2} = 0.1611$$

$$\hat{\sigma}_{\tau}^2 = s_{\tau}^2 = \frac{MS(T) - MS(EE)}{jk} = \frac{2.5201167 - 0.320775}{6} = 0.365232$$

$$Total = s^2 + s_{\varepsilon}^2 + s_{\tau}^2 = 0.0066 + 0.1611 + 0.3652 = 0.5329$$

CRD II with Subsampling Conclusions

Variation within brain areas represents = $100 * (0.0066 / 0.5329) = 1.24\%$

Variation among brain areas represents = $100 * (0.1611 / 0.5329) = 30.22\%$

Variation among people represents = $100 * (0.3652 / 0.5329) = 68.54\%$

Estimate the total mean EEG power = 3.212

$$\text{Standard Error} = SE = \sqrt{\frac{MS(EE)}{n}} = \sqrt{\frac{0.0066542}{24}} = 0.01665$$

Randomized Complete Block Design

- removes source of variation
- if it is known in advance that the experimental units are NOT homogeneous then the CRD is no longer appropriate.
- the RCBD is used to remove sources of heterogeneity among experimental units.
- here experimental units are allocated to blocks such that those assigned to the same block should be similar in response to their treatment (i.e. homogeneous as possible).

RCBD

- treatments are then allocated to the experimental units of each block, by a separate randomization that is carried out within each block.

Some blocking factors could include:

- DATE of experiment
- cage battery for animal housing
- plot of land
- incubator or oxygen chamber
- individual hospital

RCBD Fruit Fly Example

Consider the following genetics experiment.

- 5 people in the lab were all assigned to a project where they were to assess protein levels of CuZnSOD (superoxide dismutase) in *Drosophila melanogaster* (fruit fly) that had been transfected with human CuZnSOD. The 'boss' wanted to know which cross (i.e. homozygote (hom), heterozygote (het), or wild type (wld)) had higher CuZnSOD.

RCBD Fruit Fly Example

For this RCBD design:

A block is an individual person

A treatment is a genotype

An experimental unit is a fruit fly

A model equation for the RCBD design:

$$Y_{ij} = \mu + \tau_i + B_j + \varepsilon_{ij}$$

Null hypothesis: H_0 : all $\tau_i = 0$

Alternative: H_A : all $\tau_i \neq 0$

Null hypothesis: H_0 : all $B_j = 0$

Alternative: H_A : all $B_j \neq 0$

RCBD Fruit Fly Example

$df_{block} = j - 1 = 5 - 1 = 4$
 $df_{treat} = i - 1 = 3 - 1 = 2$
 $df_{total} = (i \cdot j) - 1 = 3 \times 5 - 1 = 14$
 $df_{error} = total - (block + treat) = (i-1)(j-1) = 8$
 i = 3 (treatments); j = 5 (blocks)

$$Total(SS) = \sum Y_{\bullet j}^2 - \frac{Y_{\bullet\bullet}^2}{ij} =$$

$$SS(Blocks) = \frac{\sum Y_{\bullet j}^2}{i} - \frac{Y_{\bullet\bullet}^2}{ij} =$$

RCBD Fruit Fly Example

$df_{block} = j - 1 = 5 - 1 = 4$
 $df_{treat} = i - 1 = 3 - 1 = 2$
 $df_{total} = (i \cdot j) - 1 = 3 \cdot 5 - 1 = 14$
 $df_{error} = total - (block + treat) = (i-1)(j-1) = 8$
i = 3 (treatments); j = 5 (blocks)

$$SS(Treat) = \frac{\sum Y_{i\bullet}^2}{j} - \frac{Y_{\bullet\bullet}^2}{ij} =$$

$$SS(\text{Error}) = \text{Total}(SS) - [SS(\text{blocks}) + SS(\text{Treat})]$$

RCBD Fruit Fly Example

Source	df	SS	MS	F_c	$F_{i,j,\alpha}$
Block _{person}	4	18.947706	4.73926	15.76	3.84
treatment _{type}	2	65.81397	32.906987	109.4	4.46
Exp. Error	8	2.406194	0.3007743	-----	-----
TOTAL	14	87.167873	-----	-----	-----

Here $F_{i,j,\alpha} = F_{2,8,0.05} = 4.46$

Conclusion: Reject H_0 , (i.e. $F_C > F_{\text{table}}$)
i.e. treatments are different !

Here $F_{i,j,\alpha} = F_{4,8,0.05} = 3.84$

Conclusion: Reject H_0 , (i.e. $F_C > F_{\text{table}}$)
i.e. Blocks are NOT homogeneous !

RCBD Fruit Fly Example

If did not block on person, we could potentially contaminate the real source of differences in the data with differences between the people's lab techniques.

So, where's the source of differences ?

- can use lsd, Scheffe, etc.

e.g. critical value for lsd:

$$lsd = t_{v, \alpha/2} \times \sqrt{E(MS) \times \frac{2}{j}} = 2.306 \times \sqrt{0.3007743 \times \frac{2}{5}} = 0.799852$$

Where $t_{v, \alpha/2} = t_{8, 0.025} = 2.306$

v (nu) = df for $E(MS) = (i-1)(j-1) = 8$

RCBD Fruit Fly Example

	HET (42.79)	WLD (45.206)
HOM (40.078)	2.712*	5.128*
HET (42.79)	-----	2.416 *

* = significantly different

Final Statements

- the ‘boss’ had hoped CuZnSOD transfection would work. Obviously it didn’t !!
- If anything the resultant flies had less.
- the lab has 5 people with significantly differing technical skills.

Differences between CRD and RCBD

- if we didn't block on person E(MS) would have been equal to $21.3539/12 = 1.7795$
- therefore the $F_c = 18.49$ (re: $F_{table} = F_{2,12,0.05} = 3.89$).
- the CRD doesn't partition the Error.

Let's revisit those trout fry:

Before, we had 4 diets, 5 fish/diet. Now let's suppose we obtained 4 fish (1/diet) from each of 5 hatcheries.

- this time, 4 fish had been randomized to 4 diets at each fish hatchery.
- fish are homogeneous in their response to treatment, but the hatchery may be a source of heterogeneity we wish to remove.

Fish Example: CRD

Example: Want to examine the effect of 4 diets on the growth of rainbow trout fry.

Dietary Treatment				
1	2	3	4	
59	70	93	124	
47	59	85	135	i = 4 treatments
40	52	79	167	j = 5 replicates
32	87	72	83	$\sum^j Y_{ij} = Y_{i\bullet} = 217 \quad 329 \quad 417 \quad 661$
39	61	88	152	$\sum^1 Y_{ij}^2 = 9835 \quad 22375 \quad 35043 \quad 91483$
217	329	417	661	$\bar{Y}_{i\bullet} = 43.4 \quad 65.8 \quad 83.4 \quad 132.2$
Replicates (mass in grams)				
totals				
$\rightarrow 1624 = Y_{\bullet\bullet}$				
$\rightarrow 158736 = \bar{Y}_{\bullet\bullet}$				
$\rightarrow 81.2 = \bar{Y}_{\bullet\bullet}$				

Fish Example: RCBD

Example: Want to examine the effect of 4 diets on the growth of rainbow trout fry in n 4 different hatcheries

i = 4 treatments

j = 5 blocks

Hatchery	Dietary Treatment				$\sum_i Y_{ij} = Y_{\bullet j}$
Replicates (mass in grams)	1	2	3	4	
A	59	70	93	124	346
B	47	59	85	135	326
C	40	52	79	167	338
D	32	87	72	83	274
E	39	61	88	152	340
	217	329	417	661	totals
					$\sum_j Y_{ij} = Y_{i\bullet} = 217 \quad 329 \quad 417 \quad 661$
					$\sum_j Y_{ij}^2 = 9835 \quad 22375 \quad 35043 \quad 91483$
					$\bar{Y}_{i\bullet} = 43.4 \quad 65.8 \quad 83.4 \quad 132.2$
					$1624 \quad 158736 \quad 81.2 = \bar{Y}_{\bullet\bullet}$

Compute Sums of Squares CRD:

$$Total(SS) = \sum Y_{ij}^2 - \frac{(Y_{..})^2}{i \times j} = 158736 - \frac{1624^2}{4 \times 5} = 26867.2$$

SS(Treatments)=

$$SS(T) = \frac{\sum (Y_{i.})^2}{j} - \frac{(Y_{..})^2}{i \times j} = \frac{217^2 + 329^2 + 417^2 + 661^2}{5} - \frac{1624^2}{4 \times 5} = 21359.2$$

SS(error) = SS(E) = Total(SS) - SS(T)

ANOVA: CRD

Source	df	SS	MS	F_c	$F_{i,j,\alpha}$
treatment	3	21359.2	7119.73	20.68	3.24
error	16	5508.0	344.25	-----	-----
Total	19	26867.2	-----	-----	-----

$$df_{\text{treatment}} = (\# \text{ treatments} - 1) = (i - 1)$$

$$df_{\text{TOTAL}} = (i \times j) - 1$$

$$df_{\text{error}} = df_{\text{TOTAL}} - df_{\text{treatment}}$$

$$\begin{aligned} MS &= SS / df \\ F_c &= T(MS) / E(MS) \end{aligned}$$

$F_{i,j,\alpha}$ = from table

e.g. $F_{3,16,0.05} = 3.24$

ANOVA: RCBD

Source	df	SS	MS	F_c	$F_{i,j,\alpha}$
Block _{hatchery}	4	859.2	214.8	0.554	3.26
treatment _{diet}	3	21359.20	7119.733	18.38	3.49
Exp. Error	12	4648.8	387.4		
Total	19	26867.2	-----	-----	-----

Here $F_{i,j,\alpha} = F_{3,12,0.05} = 3.49$ Conclusion: Reject H_0 , (i.e. $F_c > F_{\text{table}}$)

The difference?

- Error SS from CRD gets divided up into Error(SS) & Block(SS) in the RCBD design.
- The RCBD design removes some of the experimental error as error due to block effect.

CRD

Source	df	SS	MS	F_c	$F_{i,j,\alpha}$
treatment	3	21359.2	7119.73	20.68	3.24
error	16	5508.0	344.25	----	----
Total	19	26867.2	----	----	----

RCBD

Source	df	SS	MS	F_c	$F_{i,j,\alpha}$
Block _{hatchery}	4	859.2	214.8	0.554	3.26
treatment _{diet}	3	21359.20	7119.733	18.38	3.49
Exp. Error	12	4648.8	387.4	----	----
Total	19	26867.2	----	----	----

Assessing the Efficiency of Blocking

$$\hat{\sigma}_{RCBD}^2 = MS(E) = 387.4$$

$$\hat{\sigma}_{CRD}^2 = \frac{(j-1)s_{block}^2 + j(i-1)s^2}{(i-1)(j-1)} = \frac{(5-1)214.8 + 5(4-1)387.4}{(4-1)(5-1)} = 344.25$$

$$\frac{\hat{\sigma}_{RCBD}^2}{\hat{\sigma}_{CRD}^2} = \frac{387.4}{344.25} = 1.125$$

- if this ratio > 1.0 then the RCBD is not any more efficient.

Fruit Fly Example

$$\hat{\sigma}_{RCBD}^2 = MS(E) = 0.3007743$$

$$\hat{\sigma}_{CRD}^2 = \frac{(j-1)s_{block}^2 + j(i-1)s^2}{(i-1)(j-1)} = 1.7795$$

$$\frac{\hat{\sigma}_{RCBD}^2}{\hat{\sigma}_{CRD}^2} = \frac{0.3007743}{1.7795} = 0.169 = 16.9\%$$

Interpretation:

- a CRD design with, say, 100 experimental units not assembled into blocks will give answers that are about as precise as those for a RCBD with about 17 experimental units !!

RCBD with Subsampling

weanling rats fed 5 diets for 2 weeks

A = ZnDF

B = PEM

C = ZnPf

D = ZnAL

E = +ZnAL

- measured final weight (all started at exactly the same weight)

BLOCK	DIET					Totals	
	A	B	C	D	E		
1	72	82	110	117	138	519	56741
	61	87	105	103	116	472	46380
	58	79	99	110	113	459	44275
subtotal	191	248	314	330	367	1450	
2	54	82	106	117	127	486	50694
	55	76	97	108	119	455	44035
	61	80	102	114	131	488	50682
subtotal	170	238	305	339	377	1429	
3	65	83	110	122	139	519	57419
	53	80	99	104	117	453	43515
	50	75	98	125	125	473	48979
subtotal	168	238	307	351	381	1445	442720
TOTALS	529	724	926	1020	1125	4374	

Y...

$\sum Y_{ij}^2$

121

RCBD with Subsampling Model Equation

Assumptions:

1) ϵ_{ij} are $N(0, s\epsilon^2)$

$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} + \delta_{ijk}$$

2) $\sum \beta_j = 0$

3) $\sum \tau_i = 0$

4) δ_{ijk} are $N(0, s\delta^2)$

How many parameters does the model try to fit ?

$$= (\# \text{ treatments}) + (\# \text{ blocks}) + 3 = 11$$

(i.e. $\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \beta_1, \beta_2, \beta_3, \mu, \sigma\delta^2, \sigma\epsilon^2$)

RCBD with Subsampling :Calculations

$$Total(SS) = \sum Y_{ijk}^2 - \frac{Y_{\dots\dots}^2}{ijk} = \frac{4324^2}{(5)(3)(3)} = 442720 - 415488.36 = 27231.65$$

i = 5 (treatments) = diet
j = 3 (blocks) = cage battery
k = 3 (subsamples) = rat

$$SS(AmongRats) = \frac{\sum Y_{ij\dots}^2}{k} - \frac{Y_{\dots\dots}^2}{ijk} = \frac{191^2 + 248^2 + \dots + 381^2}{3} - \frac{4324^2}{(5)(3)(3)} = 25600.98$$

$$SS(Error) = Total(SS) - SS(AmongRats) = 27231.65 - 25600.98 = 1630.67$$

$$SS(Blocks) = \frac{\sum Y_{\dots j\dots}^2}{ik} - \frac{Y_{\dots\dots}^2}{ijk} = \frac{1450^2 + 1429^2 + 1445^2}{(5)(3)} - \frac{4324^2}{(5)(3)(3)} = 16.04$$

$$SS(T) = \frac{\sum Y_{i\dots\dots}^2}{jk} - \frac{Y_{\dots\dots}^2}{ijk} = \frac{529^2 + 724^2 + \dots + 1175^2}{(3)(3)} - \frac{4324^2}{(5)(3)(3)} = 25346.98$$

$$SS(ExpError) = SS(AmongRats) - [SS(Blocks) + SS(T)] = 25600.98 - [16.04 + 25346.98] = 237.96$$

RCBD with SS ANOVA Table:

Source	df	SS	MS	F_c	$F_{i,j,\alpha}$
Among ExpUnits _{Rats}	14	25600.98	-----	-----	-----
Block _{battery}	2	16.04	4.01	-----	-----
treatment _{diet}	4	25346.98	6336.74	213.07	3.84
Exp. Error	8	237.96	29.745	0.5479	2.27
ERROR	30	1630.67	54.36	-----	-----
TOTAL	44	27231.65	-----	-----	-----

$$df_{\text{expUnits}} = ij-1 = (5)(3)-1 = 14 \quad df_{\text{ExpError}} = (i-1)(j-1) = 4 \times 2 = 8$$

$$df_{\text{Block}} = j-1 = 2$$

$$df_{\text{treatment}} = i-1 = 4$$

$$df_{\text{Error}} = ij(k-1) = (5)(3)(3-1) = 30$$

$$df_{\text{Total}} = ijk-1 = (5)(3)(3)-1 = 44$$

RCBD with Subsampling: ANOVA

$$F_c(\text{treatment}) = \frac{MS(T)}{MS(EE)} = \frac{6336.74}{29.745} = 213.07$$

$$F_c(\text{ExpError}) = \frac{MS(EE)}{MS(E)} = \frac{29.745}{54.36} = 0.5427$$

- To test the null hypothesis of no differences between treatments one should use $MS(T)/MS(EE)$ as above.
- However, if $MS(EE) \leq MS(E)$ it is recommended that you should use the pooled error:

$$MS(\text{PooledError}) = \frac{SS(EE) + SS(E)}{df_{EE} + df_E} = \frac{237.96 + 1630.67}{8 + 40} = 49.2$$

RCBD with Subsampling: ANOVA

$$F_C(\text{treatment}_{\text{pooled}}) = \frac{MS(T)}{MS(PE)} = \frac{6336.74}{49.2} = 128.86$$

This parameter would be used to test the null hypothesis concerning treatment effects.

- this situation could happen when the variation among experimental units is insignificant, and error is only within experimental units, as measured by the sampling error (i.e. $MS(E)$).
- in other words $MS(E)$ and $MS(EE)$ are essentially measuring the same thing (i.e. the quantity σ^2)

RCBD with Subsampling: MS

$$MS(B) = \sigma^2 + k\sigma_\varepsilon^2 + \frac{ik}{(j-1)} \sum \beta_j^2 = \sigma^2 + 3\sigma_\varepsilon^2 + 7.5 \sum \beta_j^2$$

$$MS(T) = \sigma^2 + k\sigma_\varepsilon^2 + \frac{jk}{(i-1)} \sum \beta_j^2 = \sigma^2 + 3\sigma_\varepsilon^2 + 2.25 \sum \beta_j^2$$

$$MS(EE) = \sigma^2 + k\sigma_\varepsilon^2$$

$$MS(E) = \sigma^2$$

RCBD with Subsampling: Results

The best estimate of $\sigma\epsilon^2 = 0$

→ This is because we accepted the null hypothesis that experimental error was not significant.

(i.e. $F_c < F_{table} = 0.5479 < 2.27$)

The best estimate of σ^2 is the pooled error 49.2

Is there a significant difference among diets ?

→ YES reject H_0 since $F_C > F_{table}$ (i.e. $213.1 > 3.84$)

RCBD with Subsampling: Results

So, there is a significant diet effect. Where is(are) the differences?

To evaluate the nature of these differences one can use the lsd method:

$$lsd = t_{v,\alpha/2} \times \sqrt{MS(EE) \times \frac{2}{jk}} = 2.306 \times \sqrt{29.74 \times \frac{2}{(3)(3)}} = 5.9282$$

Note that k is added, only in subsampling

Note: other uses of the lsd method uses E(MS), the error mean SS. For subsampling the MS(EE) is used instead.

RCBD with Subsampling: Post Hoc

Diet	B _{PEM}	C _{ZnPF}	D _{ZnAL}	E _{+ZnAL}
A _{ZnDF}	21.666*	44.111*	54.555*	66.222*
B _{PEM}		22.445*	32.889*	44.556*
C _{ZnDF}			10.444*	22.111*
D _{ZnAL}				11.667*

* = significant

Where,

$$A_{ZnDF} = 58.778 \text{ g}$$

$$B_{PEM} = 80.444 \text{ g}$$

$$C_{ZnPF} = 102.889 \text{ g}$$

$$D_{ZnAL} = 113.333 \text{ g}$$

$$E_{+ZnAL} = 125.0 \text{ g}$$

Other Important Designs

1). Factorial Design

- to this point only one factor has been investigated (e.g. diet on weight gain)
- what about treatment combinations that are somehow jointly responsible for the response

2). Analysis of Covariance (ANCOVA)

- any of the other models. However, include a covariate term (e.g. initial age, initial weight, scalp/skin impedance, SNR, etc.)

Lecture 3

TAYLOR DEVET MASC.

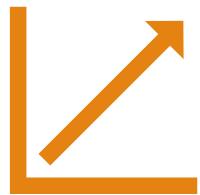
PHD. CANDIDATE BIOLOGICAL AND BIOMEDICAL ENGINEERING

MCGILL UNIVERSITY

SHRINERS HOSPITAL FOR CHILDREN

...

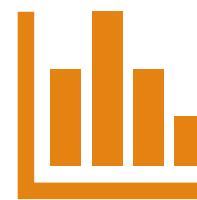
Todays Aims...



Correlation



Complete Randomized
Design



Post Hoc Testing

Let's go back to the cardiac data.....

Cardiac Example

Null Hypothesis: None of the independent variables are of any value in explaining the variation in cardiac force output.

$$\text{i.e. } H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

Alternative, H_1 : not all β 's are zero.

Source	df	SS	MS	F_c	$F_{i,j,\alpha}$
regression	4	197382.43	49458.11	34.48	2.84
residual	21	30121.92	1434.38	-----	-----
TOTAL	25	227954.35	-----	-----	-----

$F_c > F_{i,j,\alpha}$. Therefore, reject H_0 ; at least one β is not zero.

SSCP Matrix

Sum of Squares and Cross Product Matrix

Main diagonal are sum of squares for each column

Off main diagonals are sums of cross products

Inverse of this times Stdev gives us a matrix with variances and covariances

26.0000	49.2900	6.5150	19.8100	23.0700	0.9186	-0.3780	-0.9663	-0.1060	0.1796
49.2900	97.9781	12.7980	39.0012	45.9843	-0.3780	0.4209	-0.6846	-0.0550	-0.2328
6.5150	12.7980	1.7540	5.2688	6.1600	-0.9663	-0.6846	20.1257	-2.3701	-1.0967
19.8100	39.0012	5.2688	16.6387	18.9306	-0.1060	-0.0550	-2.3701	1.5028	-0.3842
23.0700	45.9843	6.1600	18.9306	23.1015	0.1796	-0.2328	-1.0967	-0.3842	0.9346

$(X'X)$

$(X'X)^{-1}$

Variance Covariance Matrix

$$\begin{vmatrix} Var(b_0) & Cov(b_0, b_1) & Cov(b_0, b_2) & Cov(b_0, b_3) & Cov(b_0, b_4) \\ Cov(b_1, b_0) & Var(b_1) & Cov(b_1, b_2) & Cov(b_1, b_3) & Cov(b_1, b_4) \\ Cov(b_2, b_0) & Cov(b_2, b_1) & Var(b_2) & Cov(b_2, b_3) & Cov(b_2, b_4) \\ Cov(b_3, b_0) & Cov(b_3, b_1) & Cov(b_3, b_2) & Var(b_3) & Cov(b_3, b_4) \\ Cov(b_4, b_0) & Cov(b_4, b_1) & Cov(b_4, b_2) & Cov(b_4, b_3) & Var(b_4) \end{vmatrix} = s^2 \times (X' X)^{-1}$$

$$= \begin{vmatrix} \mathbf{1317.55} & -542.151 & -1386.08 & -152.066 & 257.6247 \\ -542.151 & \mathbf{603.774} & -981.925 & -78.8711 & -333.958 \\ -1386.08 & -981.925 & \mathbf{28867.82} & -3399.59 & -1573.08 \\ -152.066 & -78.8711 & -3399.59 & \mathbf{2155.64} & -551.089 \\ 257.6247 & -333.958 & -1573.08 & -551.089 & \mathbf{1340.625} \end{vmatrix}$$

Cardiac Example cont

$(X'X)^{-1}$	0.9186	-0.3780	-0.9663	-0.1060	0.1796
	-0.3780	0.4209	-0.6846	-0.0550	-0.2328
	-0.9663	-0.6846	20.1257	-2.3701	-1.0967
	-0.1060	-0.0550	-2.3701	1.5028	-0.3842
	0.1796	-0.2328	-1.0967	-0.3842	0.9346

With rejection of H_0 , one can test for specific factors.

$$\text{e.g. } H_0 : \beta_4 = c$$

where c is any number specified (often zero).

$$t = \frac{\hat{b}_4 - c}{\text{Stderr}(\hat{b}_4)} = \frac{\hat{b}_4 - c}{\sqrt{s^2 \cdot (X'X)^{-1}}} = \frac{40.31 - 0}{\sqrt{1434.38 \cdot 0.9346}} = 1.1$$

1317.55	-542.151	-1386.08	-152.066	257.6247
-542.151	603.774	-981.925	-78.8711	-333.958
-1386.08	-981.925	28867.82	-3399.59	-1573.08
-152.066	-78.8711	-3399.59	2155.64	-551.089
257.6247	-333.958	-1573.08	-551.089	1340.625

$$\begin{aligned} \hat{e} &= 185.33 \\ \hat{e} &= 97.76 \\ \hat{e} &= 256.97 \\ \hat{e} &= 126.57 \\ \hat{e} &= 40.28 \end{aligned}$$

$$\beta = \hat{e}$$

- s^2 is the residual(MS) from the full model (ANOVA) table.

Therefore, there is little evidence to reject $H_0 : \beta_4 = 0$

Cardiac Example cont

A 95% confidence interval may be calculated for β_4 :

$$b_4 \pm t_{\alpha, df} \times stderr(b_4) = 40.3 \pm (2.080 \times 36.61)$$

$$stderr(b_4) = \sqrt{s^2 \times (X' X)^{-1}_{5,5}}$$

Where: $t_{\alpha, df} = t_{0.05, 26-5} = 2.080$ $t = 1.1$

Therefore, $-35.8 \leq b_4 \leq 116.4$

Reject $H_0? \beta_4 = 0$ Based on the confidence interval it could very well be zero.

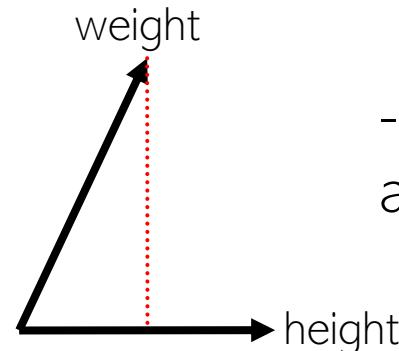
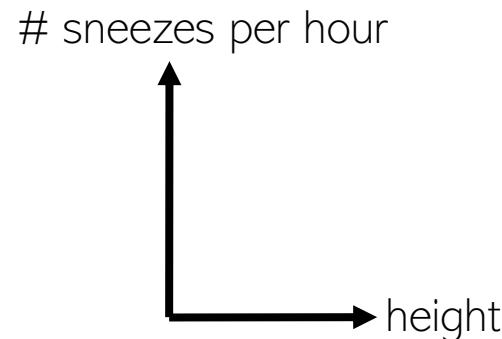
Correlated Variables

Need to understand whether independent variables are correlated.

If independent variables are not correlated among themselves they are said to be orthogonal.

When correlated they are non-orthogonal.

Non-correlated independent variables = **Orthogonal**



- weight and height
are correlated

Multicollinearity

A lot of these calculations pend on the $X'X$ matrix being invertible.

This problem is referred to as the multicollinearity problem

Multicollinearity occurs when there are high correlations between two or more predictor variables.

i.e. one predictor variable can be used to predict the other giving redundancy and skewing the model.

Multicollinearity

Examples of multicollinear predictors:

- 1) a person's height and weight
- 2) years of education and annual income.
- 3) Height in metres and height in feet

An easy way to detect:

Calculate correlation coefficients (r) for all pairs of predictor variables. If r is exactly +1 or -1, this is called perfect multicollinearity and one of the variables should be removed from the model

Main Causes of Multicollinearity:

1) Data-based

- poorly designed experiments
- data that is 100% observational, or data collection methods that cannot be manipulated.
- In some cases, variables may be highly correlated (usually due to collecting data from purely observational studies) and there is no error on the researcher's part.
Can test this in advance!!

2) Structural multicollinearity

- caused by the researcher, poorly creating new predictor variables

There is a fix!! Plan the experiment!

Observational Example:

Estimate 4 slopes and 1 intercept of zero (i.e. not estimating)

X =

1	0	0	0
0	1	0	1
0	0	0	1
1	0	0	0
0	1	0	1
0	0	0	1
1	0	0	0
0	1	0	1
0	0	0	1
1	0	0	0
0	1	0	1
0	0	0	1

>> X'*X

ans =

4	0	0	0
0	4	0	4
0	0	0	0
0	4	0	8

>> (X'*X)^-1

Warning: Matrix is singular to working precision.

ans =

Inf	Inf	Inf	Inf
Inf	Inf	Inf	Inf
Inf	Inf	Inf	Inf
Inf	Inf	Inf	Inf

Observational Example:

Estimate 4 slopes and 1 intercept as b

X =

1	1	0	0	0
1	0	1	0	0
1	0	0	1	0
1	0	0	0	1
1	1	0	0	0
1	0	1	0	0
1	0	0	1	0
1	0	0	0	1
1	1	0	0	0
1	0	1	0	0
1	0	0	1	0
1	0	0	0	1
1	1	0	0	0
1	0	1	0	0
1	0	0	1	0
1	0	0	0	1
1	1	0	0	0
1	0	1	0	0
1	0	0	1	0
1	0	0	0	1

>> X'*X

ans =

16	4	4	4	4
4	4	0	0	0
4	0	4	0	0
4	0	0	4	0
4	0	0	0	4

$$(X'X)^{-1} = 1e+15 \cdot \begin{bmatrix} 4.50 & -4.50 & -4.50 & -4.50 & -4.50 \\ -4.50 & 4.50 & 4.50 & 4.50 & 4.50 \\ -4.50 & 4.50 & 4.50 & 4.50 & 4.50 \\ -4.50 & 4.50 & 4.50 & 4.50 & 4.50 \\ -4.50 & 4.50 & 4.50 & 4.50 & 4.50 \end{bmatrix}$$

Essentially this is infinite!!

Correlated Variables

In the heart example the independent variables are correlated.

i.e. force is related to a combination of ions!

- What about sodium, β_4 ? We tested and accepted $H_0: \beta_4 = 0$

The general approach to the analysis is the comparison of two competing models:

1) The full model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \cancel{\beta_4 X_{i4}} + \varepsilon_i$$

2) The reduced model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

Model Reduction

Fit the full model (done already) and reduced model to the data and for each obtain the model SS and residual SS.

- The appropriate SS for testing $H_0: \beta_4 = 0$ is:

$$\begin{aligned} R(\beta_4 | \beta_1, \beta_2, \beta_3) &= \text{SS model (full)} - \text{SS model (reduced)} \\ &= \text{SS residual (reduced)} - \text{SS residual (full)} \end{aligned}$$

- this 'R' notation helps to differentiate the different SS.
- 'R' means 'Reduction in residual SS'

Model Reduction

The full model (already calculated) is:

$$R(\beta_1, \beta_2, \beta_3, \beta_4) = \text{SS (full model)} = 197832.43$$

The reduced model is:

$$R(\beta_1, \beta_2, \beta_3) = \text{SS (reduced model)} = 196096.77$$

Therefore,

$$\begin{aligned} R(\beta_4 | \beta_1, \beta_2, \beta_3) &= R(\beta_1, \beta_2, \beta_3, \beta_4) - R(\beta_1, \beta_2, \beta_3) \\ &= 197832.43 - 196096.77 \\ &= 31857.58 - 30121.92 \\ &= 1735.66 \end{aligned}$$

NOTE: red #s are from reduced model - you should verify these !!

Model Reduction

The reduced model is fitted in the same way as the full model !!

$$\hat{\mu}_{Y_{.123}} = -193.07 + 107.80X_1 + 304.24X_2 + 143.13X_3$$

Source	df	SS	MS
Reduced model	3 ← 4	196096.77 197k...	65365.59
residual	22 21	31857.58 30121	1448.07
TOTAL	25 25	227954.35	-----

NOTE: you should verify the ANOVA and reduced b results !!

Model Reduction

- the df for $\beta_4 = 1$ (i.e. 4 - 3)

Calculation of the F statistic to test $H_0: \beta_4 = 0$

For the denominator MS of the F-test the residual MS from the FULL MODEL is used!

$$F = \frac{R(\beta_4 | \beta_1, \beta_2, \beta_3) / df_{\beta_4}}{\text{Error}(MS)_{\text{FULL}}} = \frac{1735.66 / 1}{1434.38} = 1.21$$

$$F_{1,21,\alpha=0.05} = 4.32 \text{ (table)}$$

Since $F_c < F_{\text{table}}$ there is little evidence to suggest $\beta_4 \neq 0$

Which model do we want?

Use F statistic

$$F = \frac{\frac{SSE \text{ reduced} - SSE \text{ Full}}{df \text{ reduced} - df \text{ full}}}{\frac{SSE \text{ full}}{df \text{ full}}}$$

$$F = \frac{\frac{31857.58 - 30121.92}{22 - 21}}{\frac{30121.9}{21}}$$

$$F = 1735.66 / 1434.37 = 1.21$$

$$\frac{df \text{ Reduced} - df \text{ Full}}{df \text{ Full}} =$$
$$F(\text{alpha } 1, 21) = 4.32$$

Model Reduction

Therefore, it can be concluded that Na is not related to force after allowing for the influence of Cl⁻, PO₄ and K⁺.

Note: If we had only done simple linear regression of Y on X4, ignoring X1, X2, and X3:

- Would have been fitting the model $Y_i = \beta_0 + \beta_4 X_4 + \epsilon_i$)
 - Testing the significance of the regression (i.e. $H_0: \beta_4 = 0$) would have resulted in $F_c = 34.4$
 - Testing two VERY different hypotheses
- Therefore careful experimental planning are essential !

Estimating Y for Given Values of X's

Example. Estimate the contractile force with Cl-=2, PO4=0.3, and K+=0.8 in the buffer. (note X4 removed, based on previous test)

Therefore the estimate of force is:

$$\hat{Y}_{1,23} = \beta_1 + \beta_2(2) + \beta_3(0.3) + 143.13(0.8) = 228.3 \text{ N}$$

In other words, we estimate the force output will be 228.3 N.

Standard Error of Estimate:

Recall, the b's are not independent.

Consequently the standard error will involve the covariances among b's as well as their variances.

$$\hat{\mu}_{Y.123} = x_0' \times b$$

Note:

- Where $x_0' = [1, 2, 0.3, 0.8]$ and

$$var(\hat{\mu}_{Y.123}) = x_0' \cdot (X'X)^{-1} \cdot x_0 \cdot \sigma^2 \rightarrow \begin{matrix} \text{MS(E)} \\ \text{FULL MODEL} \end{matrix}$$

$\uparrow \cancel{5 \times 5} \quad 4 \times 4$

Standard Error of Estimate

$$x_0' (X'X)^{-1} x_0 = \begin{vmatrix} 1 & 2 & 0.3 & 0.8 \end{vmatrix} \times \begin{vmatrix} 0.8840 & -0.3332 & -0.7556 & -0.0322 \\ -0.3332 & 0.3629 & -0.9578 & -0.1507 \\ -0.7556 & -0.9578 & 18.8388 & -2.8209 \\ -0.0322 & -0.1507 & -2.8209 & 1.3449 \end{vmatrix} \times \begin{vmatrix} 1 \\ 2 \\ 0.3 \\ 0.8 \end{vmatrix} = 0.06869$$

Therefore: $SE = \sqrt{\text{Var}(\mu_{123}) \cdot \sigma^2} = \sqrt{0.06869 \cdot 1448.07} = 9.973$

Hence, a 95% confidence interval can be constructed for the true contractility with Cl=2, PO₄=0.3, and K⁺=0.8 as :

$$228.3 \pm (2.074)(9.973) = (207.6, 249.0) \quad (\text{Student's-t with } df = 22)$$

An Introduction To Statistical Design: Terminology

ANOVA = analysis of variance

rmANOVA = repeated measures ANOVA

CRD = completely randomized design →

RCBD = randomized complete block design →

~~ANCOVA~~ = analysis of covariance

~~SPD~~ = split plot design

· FACTORIAL designs

~~MANOVA~~ = multivariate analysis of variance
etc. etc. etc.....

ANOVA Part 1: One way

testing to see whether many means come from the same population

Goal

- Determine likely values of measure if samples in each group are from the same Population
- Develop a measure for the difference between experimental groups based on the means and using the estimate of variability for scaling of the difference

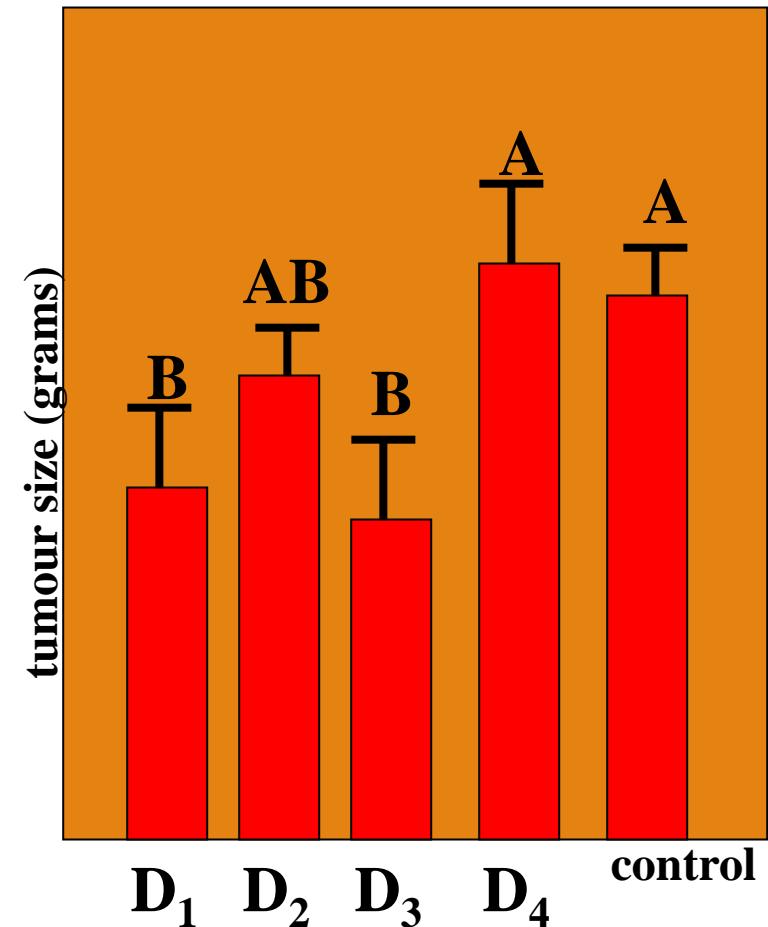
Example:

- 4 new anticancer drugs compared against a control

Null Hypothesis: all treatments
the same

Alternative: at least one is
different.

- letters denote treatments that are significantly different from one another.



Use F test for Multiple Groups

Is there a difference between m groups of n samples each? (also sometimes i and j, or t and r).

To verify if all groups are from the same population one guesses that they are actually identical and validates or invalidates the statement.

If the hypothesis is true:

- The average variance of the individual groups should be smaller or equal to the variance of a given population.

Completely Randomized Design (CRD)

- treatments are randomly assigned, completely at random to the experimental units, which are assumed to be homogeneous.
- Model 1: Fixed Effects Case (examine effects of treatments)
- Model 2: Random Effects Case (identify sources of variation)

		Treatments				
		T ₁	T ₂	T ₃	T ₄	T ₅
Replicates	1	1.1	3.2	5.6	8.2	8.3
	2	1.7	3.4	4.8	7.6	8.0
	3	1.5	3.0	5.1	7.9	8.4

matrix notation goes from larger to smaller groups e.g. block, treatment, replicate, etc.

Model I: Fixed Model

Example: Want to examine the effect of 4 diets on the growth of rainbow trout fry.

Dietary Treatment				
	1	2	3	4
Replicates (mass in grams)	59	70	93	124 → $\sum_{j=1}^4 Y_{ij}$
	47	59	85	135 → $\sum_{j=1}^4 Y_{ij}$
	40	52	79	167 → $\sum_{j=1}^4 Y_{ij}$
	32	87	72	83
	39	61	88	152
$\sum_{j=1}^4 Y_{ij} = Y_{i\cdot} =$				<i>totals</i>
	217	329	417	661 → $1624 = Y_{\cdot\cdot}$
$\sum_{j=1}^4 Y_{ij}^2 =$				$158736 = \bar{Y}_{\cdot\cdot}$
	9835	22375	35043	91483 → 81.2
$\bar{Y}_{i\cdot} =$				
	43.4	65.8	83.4	132.2 → 81.2

Model I: Fixed Model

Compute Sums
of Squares:

$$Total(SS) = \sum Y_{ij}^2 - \frac{(\bar{Y}_{..})^2}{i \times j}$$

$$158736 - \left[\frac{1624^2}{4 \cdot 5} \right] = 26867.2$$

SS

(Treatments) =

$$SS(T) = \frac{\sum (\bar{Y}_i)^2}{j} - \frac{(\bar{Y}_{..})^2}{i \times j}$$

$$\frac{217^2 + 329^2 + 417^2 + 661^2}{5} - = 21359.2$$

$$SS(\text{error}) = SS(E) = Total(SS) - SS(T) \approx 5000$$

Model I: Fixed Model

Source	df	SS	MS	F_c	$F_{i,j,\alpha}$
treatment	3	21359.2	7119.73	20.68	3.24
error	16	5508.0	344.25	-----	-----
TOTAL	19	26867.2	-----	-----	-----

$$df_{\text{treatment}} = (\# \text{ treatments} - 1) = (i - 1)$$
$$df_{\text{TOTAL}} = (i \times j) - 1$$
$$df_{\text{error}} = df_{\text{TOTAL}} - df_{\text{treatment}}$$

$$MS = SS / df$$
$$F_c = T(MS) / E(MS)$$

$F_{i,j,\alpha}$ = from table
e.g. $F_{3,16,0.05} = 3.24$

Model I: Fixed Model

The Bottom line.....

AB ___ -

Since FC > Ftable, reject H0.

Therefore:

- there is a difference between one or more treatments
- Following ANOVA we would need to explore the differences (more on this soon...)

Model II: Random Effects Model

- not interested in specific treatments, but rather on sources of variation.

e.g. In order to study the sources of variation in synthesis of protein gumbycin, a sample of 5 cell cultures was selected at random from an incubator by a chemical engineer. A total of 4 western blots for the protein were made on each of the 5 randomly selected cultures.

	1	2	3	4	5	= culture number
protein content (ng/10 ⁶ cells)	85	62	46	67	54	
	81	67	52	57	72	$i = 5$
	83	61	55	65	68	$j = 4$
	76	58	41	54	45	$\bar{Y}_{..} = 1249$
						$\sum Y_{ij}^2 = 81023$

Model II: Random Effects Model

Null Hypothesis:

- there is no variation between cultures i.e. $H_0: \sigma^2 = 0$

Calculation of ANOVA table is exactly the same:

Source	df	SS	MS	F_c	$F_{i,j, \alpha}$
Between	4	2233.7	558.425	10.61	3.06
within	15	789.25	52.617	-----	-----
TOTAL	19	3022.95	-----	-----	-----

Model I vs Model II?

What are we estimating?

- the MS(T) is an estimate of :
- Model 1: Fixed Effects Case (examine effects of treatments)
- Model 2: Random Effects Case (identify sources of variation)

	Model I	Model II
treatment (between)	$S^2 + \frac{j}{i-1} \bar{t}_i^2$	$S^2 + jS_t^2$
error (within)	S^2	S^2

Model I vs Model II?

- in the protein case $F_C > F_{table}$.
- therefore σ^2 is significantly different from zero, and hence there is significant variability between cell cultures with respect to protein gumbycin synthesis.
- can σ^2 be estimated (i.e. variance between cultures) ? YES :

$$S_t^2 = S_e^2 = \frac{MS(T) - MS(E)}{j} = \frac{558.425 - 52.617}{4} = 126.45$$

Model I vs Model II?

- in this protein experiment the ratio of σ_t^2 to $\sigma^2 = 126.452 : 52.616$ or about 2.4 to 1

What percentage of the total variation does the variation between (σ_t^2) cultures account for ?

$$\text{total var} = \sigma_t^2 + \sigma^2 = 126.452 + 52.616 = 179.066$$
$$\frac{100 \times 126.452}{179.066} = 70.6 = 71\%$$

Real life experiments

- what can be done when there is unequal replication ?
- for example, consider an experiment to assess anti-carcinogens.
 - Rats were pre-medicated with one of 5 anticarcinogens prior to being given a single dose of benzo[a]pyrene.
 - The next day you want to assess mass of feed eaten.
 - However, for whatever reason many of the rats have died !!

	1	2	3	4	5	= treatment
mass of feed eaten (g) for each rat	8.4	6.5	7.2	7.2	7.9	
	7.6	8.1	7.4	7.5	9.6	
	8.2	7.7	6.2	-	9.9	
	7.4	-	6.6	-	-	
	8.2	-	-	-	-	

Real Life Experiments

Source	df	SS	MS	F _c	F _{i,j, α}
treatment	4	9.8208	2.4552	5.435	3.26
error	12	5.4203	0.4517	-----	-----
TOTAL	16	15.2412	-----	-----	-----

$$df_{TOTAL} = n - 1 \text{ (i.e. total samples - 1)}$$

except....

$$df_{error} = n - i \text{ (i.e. total samples - # treatments)}$$

$$df_{treatment} = i - 1 \text{ (i.e. total treatments - 1)}$$

	Model I	Model II
treatment	$S^2 + \frac{j}{i-1} \bar{t}_i^2$	$S^2 + jS_t^2$
error	S^2	S^2
treatment	$S^2 + \frac{\bar{j}_i(t_i)^2}{i-1}$	$S^2 + \frac{n - (\bar{j}^2)/n}{i-1} S_t^2$
error	S^2	S^2

Multiple Comparisons: Post hoc Procedures

What happens if you reject H₀ ?

- need to explore where the differences lie and their magnitudes.

Methods will vary in conservatism

Repeated t-tests can result in errors so we need other methods

Statistical Testing of Means

- 1) Student's t-test (2 means)
- 2) Least Significant Difference (LSD)
- 3) Duncan's new multiple range test
- 4) Contrast analysis
- 5) Scheffé Test



There are many others you can investigate on your own time.

e.g. SNK (Student-Neumann Keuls) test

Model I: Fixed Model

Example: Want to examine the effect of 4 diets on the growth of rainbow trout fry.

		Dietary Treatment					
		1	2	3	4		
Replicates (mass in grams)							
	59	70	93	124	$j = 4$ treatments		
	47	59	85	135	$j = 5$ replicates		
	40	52	79	167			
	32	87	72	83			
	39	61	88	152	<i>totals</i>		
$\sum_j Y_{ij} = Y_{i\cdot} =$		217	329	417	661	1624	$= Y_{\cdot\cdot}$
$\sum_j Y_{ij}^2 =$		9835	22375	35043	91483	158736	$= \bar{Y}_{\cdot\cdot}$
$\bar{Y}_{i\cdot} =$		43.4	65.8	83.4	132.2	81.2	

Model I: Fixed Model

Compute Sums
of Squares:

$$Total(SS) = \bar{Y}_{ij}^2 - \frac{(Y_{..})^2}{i \times j} = 158736 - \frac{1624^2}{4 \times 5} = 26867.2$$

SS
(Treatments) = $SS(T) = \frac{\bar{a}(Y_i)^2}{j} - \frac{(Y_{..})^2}{i \times j} = \frac{217^2 + 329^2 + 417^2 + 661^2}{5} - \frac{1624^2}{4 \times 5} = 21359.2$

$$SS(\text{error}) = SS(E) = \text{Total}(SS) - SS(T)$$

Model I: Fixed Model

Source	df	SS	MS	F_c	$F_{i,j,\alpha}$
treatment	3	21359.2	7119.73	20.68 >	3.24
error	16	5508.0	344.25	-----	-----
TOTAL	19	26867.2	-----	-----	-----

$$df_{\text{treatment}} = (\# \text{ treatments} - 1) = (i - 1)$$
$$df_{\text{TOTAL}} = (i \times j) - 1$$
$$df_{\text{error}} = df_{\text{TOTAL}} - df_{\text{treatment}}$$

$$MS = SS / df$$
$$F_c = T(MS) / E(MS)$$

$F_{i,j,\alpha}$ = from table
e.g. $F_{3,16,0.05} = 3.24$

Least Significant Difference (LSD)

- examines differences between means
- ideally this is used for planned comparisons (i.e. specify in advance of getting the data.)

The equation for the standard error of the difference between 2 means is:

$$SE = \left(\bar{Y}_{i_a} - \bar{Y}_{i_b} \right) = \sqrt{E(MS) \times \frac{2}{j}} = \sqrt{344.25 \times \frac{2}{5}} = 11.73$$

$$lsd = t_{n,a/2} \times \sqrt{E(MS) \times \frac{2}{j}} = 2.120 \times 11.73 = 24.8776$$

- where $v = df_{E(MS)} = i(j-1) = 4(5-1) = 16$

Least Significant Difference (LSD)

If $lsd < (\text{difference between 2 means})$, then reject H_0

- i.e. the means are significantly different ($*$ = sig. different)

NOTE: 4/6 possible combinations were declared as significant different (i.e. 66%).

Difference Table: (24.877)

Diets	2(65.8)	3(83.4)	4(132.2)
1(43.4)	22.4	40.0*	88.8*
2(65.8)	-----	-17.6	-66.4*
3(83.4)	-----	-----	48.8*

Least Significant Difference (LSD)

Underscore Representation

- Underline pairs of means that are NOT significantly different.

1(43.4) 2(65.8) 3(83.4) 4(132.2)

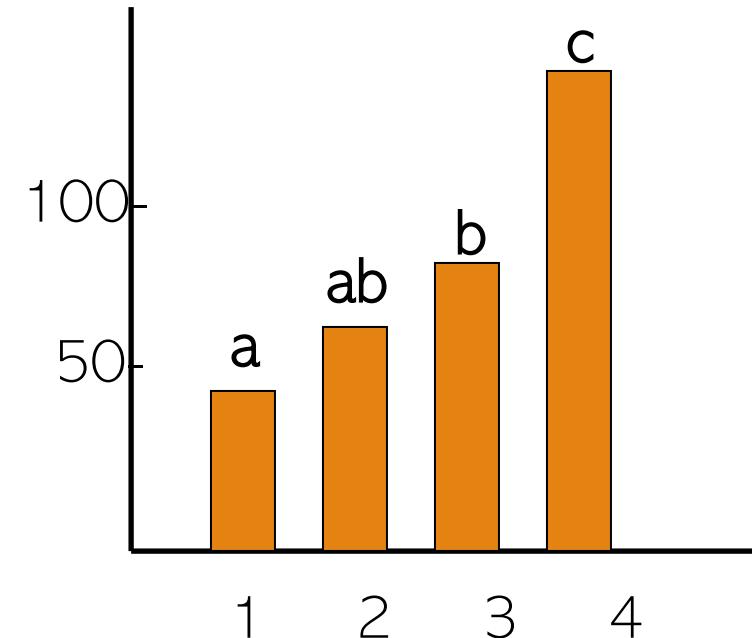
LSD results :

Diets 1 & 2, and diets 2&3 are not significantly different from each other.

However, diets 1&3, 1&4, 2&4, and 3&4 are significantly different.

Graphical Representation

same letter = *not* significantly different



Some Notes About the Results:

1 thumb

2 ↗ - Which diet would you use if you wished to raise the largest fish in the shortest period of time ?

3 clap

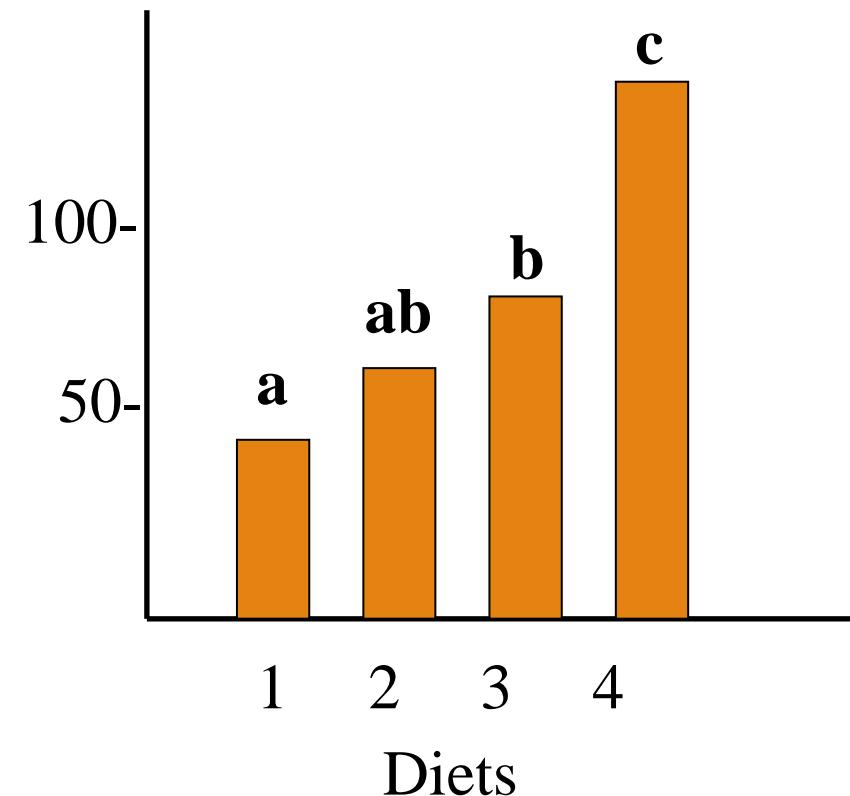
4

4 smile - If diet 3 cost much more than diet 2, which yields a greater weight gain per unit dollar of expenditure ?

2

- Suppose Diets 1 and 2 are equally priced. If they are the only 2 available diets which should be used ?

either .



Duncan's Multiple Range Method

- differs from LSD method which has a single 'least significant difference' with which to compare treatment effects.
- the Duncan method employs test criteria which vary in magnitude, depending on the number of means involved in the test.

First compute the standard error of a sample mean:

$$SE(\bar{Y}) = \sqrt{\frac{MS(E)}{j}}$$

$$= \sqrt{\frac{344.25}{5}} = 8.29759$$

Duncan's Multiple Range Method

- consult table to determine the values of
Studentized range: $q_a(p, f_e)$

- q_a = significance level (e.g. $\alpha = 0.05$)
- p = number of means (i.e. treatments) being tested
- f_e = the number of degrees of freedom of experimental error
 - (i.e. $i(j-1) = 16$, in this example)

Decision Rule:

- if the difference in means is greater than the calculated significant ranges parameter, R_p , then it is declared significant.

$$R_p = q_a(p, f_e) \times \sqrt{\frac{MS(E)}{j}}$$

Duncan's Multiple Range Method

$$SE(Y) = 8.29759$$

$$R_p = q_a(p, f_e) \times \sqrt{\frac{MS(E)}{j}}$$

p	2	3	4
$q_a(pf_e)$	$q_{0.05}(2,16) = 3.00$	$q_{0.05}(3,16) = 3.15$	$q_{0.05}(4,16) = 3.23$
R_p	$3.00(8.29759) = 24.893$	$3.15(8.29759) = 26.137$	$3.23(8.29759) = 26.801$

Difference Table

Diets	2(65.8)	3(83.4)	4(132.2)
1(43.4)	22.4	40.0*	88.8*
2(65.8)	-----	17.6	66.4*
3(83.4)	-----	-----	48.8*

Duncan's Multiple Range Method

Notice, in this particular circumstance, Duncan's and lsd result in the same conclusions for the data.

- This can differ !!

If the difference in means is particularly close then the lsd method will result in significance more often (i.e. is less conservative than Duncan's).

Linear Combinations of Means (i.e. contrasts)

- suppose there are a number of ways to improve the length of time of musculoskeletal repair after injury:

X1 = blueberry and kale smoothies .

X2 = supplements of branched chain AAs + proline .

X3 = static magnetic field (BO)

X4 = pulsed magnetic field (BO)

X5 = Therapeutic Ultrasound (US)

Example of a linear combination (2 sample case):

$$L_1 : \bar{X}_1 - \bar{X}_2$$

- estimates $\mu_1 - \mu_2$ or the difference between dietary treatments

Linear Combinations of Means

Another possible linear combination:

$$L_2 : \bar{X}_3 - \bar{X}_4$$

- measures mean difference in time of healing for magnetic field (B_0) methods.

While:

$$L_3 : \frac{1}{2} \bar{X}_3 + \frac{1}{2} \bar{X}_4$$

- i.e. an estimate of $\frac{1}{2}(\mu_3 + \mu_4)$
- measures the average time to achieve full repair due to magnets.

Linear Combinations of Means

The Question:

- does US take less time, on average, than magnetic fields (ignoring any difference between pulsed vs. static BO)?

Can be written mathematically:

$$m_5 < \frac{1}{2}(m_3 + m_4) ?$$

↓ ↓
US B

Alternatively:

$$m_5 - \frac{1}{2}(m_3 + m_4) < 0 ?$$

To estimate this difference among mean values we would use:

$$L_4 : \bar{X}_5 - \frac{1}{2}(\bar{X}_3 + \bar{X}_4)$$

Then test whether L_4 was significantly different from 0.

Linear Combinations of Means

More specifically in this case we would be doing a 1-sided test, since we are only looking at whether there is evidence that μ_5 is less than the average of μ_3 and μ_4 .

To compare diet with the average of 'engineering' methods we would ask whether the average repair times due to diet, estimated by:

$$\frac{1}{2} \bar{X}_1 + \frac{1}{2} \bar{X}_2$$

differs from the average times taken when using tech, estimated by:

$$\frac{1}{3} \bar{X}_3 + \frac{1}{3} \bar{X}_4 + \frac{1}{3} \bar{X}_5 .$$

Linear Combinations of Means

The difference between these 2 estimates:

$$L_5 : \frac{1}{2} \bar{X}_1 + \frac{1}{2} \bar{X}_2 - \frac{1}{3} \bar{X}_3 - \frac{1}{3} \bar{X}_4 - \frac{1}{3} \bar{X}_5$$

estimates:

$$\frac{1}{2} m_1 + \frac{1}{2} m_2 - \frac{1}{3} m_3 - \frac{1}{3} m_4 - \frac{1}{3} m_5$$

Linear Combinations of Means

To reiterate, here are the linear combinations again:

$$L_1 : \bar{X}_1 - \bar{X}_2$$

$$L_2 : \bar{X}_3 - \bar{X}_4$$

$$L_3 : \frac{1}{2} \bar{X}_3 + \frac{1}{2} \bar{X}_4$$

$$L_4 : \bar{X}_5 - \frac{1}{2} (\bar{X}_3 + \bar{X}_4)$$

$$L_5 : \frac{1}{2} \bar{X}_1 + \frac{1}{2} \bar{X}_2 - \frac{1}{3} \bar{X}_3 - \frac{1}{3} \bar{X}_4 - \frac{1}{3} \bar{X}_5$$

Linear Combinations of Means

Now here are the same linear contrasts in tabular form:

Means	\bar{X}_1	\bar{X}_2	\bar{X}_3	\bar{X}_4	\bar{X}_5	
Coefficients:	λ_1	λ_2	λ_3	λ_4	λ_5	$\Sigma\lambda$
L ₁	+1	-1	0	0	0	0 ..
L ₂	0	0	+1	-1	0	0
L ₃	0	0	+1/2	+1/2	0	1 -
L ₄	0	0	-1/2	-1/2	+1	0
L ₅	+1/2	+1/2	-1/3	-1/3	-1/3	0

Linear Combinations of Means

Definition:

- Linear combinations with $\sum \lambda = 0$ are called contrasts.
- A sample contrast, denoted L , is an estimator of the population contrast.
- The Standard Error of this estimate is:

$$SE(L) = SE\left(\sum_i \lambda_i \bar{X}_i\right) = \sqrt{s^2 \times \frac{\left(\sum_i \lambda_i^2\right)}{n_i}}$$

Where s^2 is the MS(E) from the ANOVA table, $\sum \lambda_i^2$ is the coefficient sum of squares, and n_i is the number of samples in the i^{th} group.

Linear Combinations of Means

If, for example, each $n_i = 5$, then the standard errors for the different linear combinations are:

$$SE(L) = \sqrt{s^2 \cdot \frac{(\sum \lambda_i^2)}{n_i}}$$

Combination	$\Sigma \lambda^2$	$SE(L)$
L1	2.000	$s \times \sqrt{2.000 / 5}$
L2	2.000	$s \times \sqrt{2.000 / 5}$
L3	0.500	$s \times \sqrt{0.500 / 5}$
L4	1.500	$s \times \sqrt{1.500 / 5}$
L5	0.833	$s \times \sqrt{0.833 / 5}$

Hypothesis Testing Using Contrasts:

Consider L4, a comparison of ultrasound with the average of magnetic field induced repair times:

$$L_4 : \bar{X}_5 - \frac{1}{2}(\bar{X}_3 + \bar{X}_4)$$

The $\Sigma \lambda^2 = 1.500$. If all 3 means are based on samples of 5 times each, then:

$$SE(L) = \sqrt{s^2 \times \frac{(\lambda_i^2)}{n_i}} = s \sqrt{\frac{1.5}{5}} = s \cdot 0.5477$$

Hypothesis Testing Using Contrasts:

If s = pooled estimate of the population standard deviation (σ), based on 5 pooled variances $s_1^2, s_2^2, s_3^2, s_4^2, s_5^2$, each with 4 df

- then s will have 5×4 (i.e. 20) degrees of freedom.

Using a Student's t-table: $t_{\alpha=0.05,20} = 1.725$; $t_{\alpha=0.025,20} = 2.086$

The null hypothesis to be tested:

$$H_0 : m_5 - \frac{1}{2}(m_3 + m_4) = 0$$

Which can also be re-written as:

$$H_0 : m_5 = \frac{1}{2}(m_3 + m_4)$$

Hypothesis Testing Using Contrasts:

Test H_0 against one of the alternative hypothesis:

$$H_A : m_5 \neq \frac{1}{2}(m_3 + m_4) \quad (\text{2-sided alternative})$$

$$H_A : m_5 > \frac{1}{2}(m_3 + m_4) \quad (\text{1-sided alternative})$$

$$H_A : m_5 < \frac{1}{2}(m_3 + m_4) \quad (\text{another 1-sided alternative})$$

Hypothesis Testing Using Contrasts:

A t-test can be used here:

$$t = \frac{\text{Estimated Value} - \text{Hypothesized True Value}}{\text{Standard Error of Estimated Value}} \quad t = \frac{\left(\bar{X}_5 - \frac{1}{2}(\bar{X}_3 + \bar{X}_4) \right) - 0}{0.5477 \times s}$$

Depending on whether we have chosen a 1 or 2 sided test we would reject H₀ in favour of the selected alternative hypothesis:

- if $|t| > 2.086$ (i.e. $t > 2.086$ or $t < -2.086$) for the 2 sided test.
- if $t > +1.725$ for the 1 sided test for the alternative which predicted $\mu_5 > \frac{1}{2}(\mu_3 + \mu_4)$.
- if $t < -1.725$ for the 1 sided test for the alternative which predicted $\mu_5 < \frac{1}{2}(\mu_3 + \mu_4)$.

Hypothesis Testing Using Contrasts:

The 95% confidence interval for a linear combination:

(Estimated value – $t_{\alpha/2, df} \times$ stderr of estimated value) < Linear combination of true values < (Estimated value + $t_{\alpha/2, df} \times$ stderr of estimated value)

In this case:

$$\left(\bar{X}_5 - \frac{1}{2}(\bar{X}_3 + \bar{X}_4) \right) - (2.086 \cdot 0.5477s) \leq \left(m_5 - \frac{1}{2}(m_3 + m_4) \right) \leq \left(\bar{X}_5 - \frac{1}{2}(\bar{X}_3 + \bar{X}_4) \right) + (2.086 \cdot 0.5477s)$$

Orthogonal Contrasts:

Means	\bar{X}_1	\bar{X}_2	\bar{X}_3	\bar{X}_4	\bar{X}_5	
Coefficients:	λ_1	λ_2	λ_3	λ_4	λ_5	$\Sigma\lambda$
Combination						
L_1	+1	-1	0	0	0	0
L_2	0	0	+1	-1	0	0
L_4	0	0	-1/2	-1/2	+1	0
L_5	+1/2	+1/2	-1/3	-1/3	-1/3	0

- Notice each contrast looks at a different characteristic of the data
- Not all contrasts look at genuinely different characteristics !!

Hypothesis Testing Using Contrasts:

For example, the contrasts:

$$\bar{X}_1 - \bar{X}_2 \quad \bar{X}_1 - \bar{X}_3 \quad \bar{X}_2 - \bar{X}_3$$

compare 1 with 2, 1 with 3, and 2 with 3. The third contrast, however, really tells us nothing we couldn't have figured out with the other two, since:

$$\bar{X}_2 - \bar{X}_3 = \bar{X}_1 - \bar{X}_3 - (\bar{X}_1 - \bar{X}_2)$$

A way to ensure that contrasts are looking at completely different aspects of the data is to require that all contrasts be orthogonal.

Hypothesis Testing Using Contrasts:

The numerical verification that 2 contrasts are orthogonal is that the sum of the products of their corresponding coefficients is zero.

For example, L_1 and L_2 are orthogonal:

L_1	+1	-1	0	0	0	
L_2	0	0	+1	-1	0	
Products	0	0	0	0	0	Sum = 0

Also, L_4 and L_5 are orthogonal:

L_4	0	0	-1/2	-1/2	1	
L_5	+1/2	+1/2	-1/3	-1/3	-1/3	
Products	0	0	1/6	1/6	-1/3	Sum = 0

Hypothesis Testing Using Contrasts:

However, $X_1 - X_2$ and $X_1 - X_3$ are not orthogonal:

X1-X2	+1	-1	0	0	0	
→ X1-X3	+1	0	-1	0	0	
Products	+1	0	0	0	0	Sum = +1

i.e. if 2 contrasts, with q number of coefficients:

$$\sum_{i=1}^q \bar{a}_i A_i / \bar{a}_i B_i = 0$$

Then contrasts A and B are orthogonal.

Non-orthogonal contrasts do not provide any extra information !!

Hypothesis Testing Using Contrasts:

It's a bit tricky at first to come up with orthogonal contrasts. The best thing to do is think up contrasts which address specific and distinct questions- Then check for orthogonality.

e.g. How do 1 and 2 compare ?

$$\begin{matrix} 1 & 0 & 0 & 1 \\ KS & \beta \end{matrix}$$

How do 3 and 4 compare ?

How do 3 and 4 compare with 5 ?

How do 1 and 2 together compare with 3, 4, and 5 together ?

→ These questions led to the orthogonal contrasts L1, L2, L3, and L5

Hypothesis Testing Using Contrasts:

NOTES:

- if we have i treatments then there exactly $i-1$ possible orthogonal contrasts.
- the $i-1$ is exactly equal to the **df** for measuring variability among the treatment means.
- These orthogonal contrasts correspond to a decomposing of this variability !!
- each contrast has **1 df** associated with it.

Sum of Squares of Contrasts:

- a measure of a size of a contrast is it's sum of squares i.e. $SS(L)$

$$SS(L) = \frac{n \times (\text{estimated value of } L)^2}{\sum \lambda_i^2}$$

- the estimated value of L is calculated usually using mean values.
- Totals can also be used (and are equivalent)

$$L_{1.}^{Totalbased} : \overline{X}_{1.} + \overline{X}_{2.}$$

Where $\overline{X}_{1.}$ is the total of all observations taken for the first treatment

$$SS(L) = \frac{(\text{estimated value of } L^{Totalbased})^2}{n \times \sum \lambda_i^2}$$

Back to the fish example

We'd like to test the following:

$$\text{for } L_1 \quad H_0 : m_4 - m_2$$

$$\text{for } L_2 \quad H_0 : m_4 - 1/2(m_2 + m_3)$$

Note- try it yourself !! Check that L_1 and L_2 are orthogonal !!

Fish Example Contrasts

Contrast L_1 : $\hat{L}_1 = \bar{X}_4 - \bar{X}_2 = 132.2 - 65.8 = 66.4$

The coefficients are: $\lambda_1 = 0; \lambda_2 = -1; \lambda_3 = 0; \lambda_4 = +1$

The sum of squares of $\Sigma\lambda^2$

$$SE(L_1) = \sqrt{s^2 \times \frac{(\lambda)^2}{n_i}} = \sqrt{344.25 \cdot \frac{2}{5}} = 11.73$$
$$t = \frac{\hat{L}_1 - 0}{11.73} = \frac{66.4}{11.73} = 5.66$$

From t-table, with df = 16, $t = 2.120$ (2 tailed)

$$+_{table} = 2.12$$

Since $t_{calc} > t_{table}$ reject H_0 i.e. diets 4 and 2 are significantly different

Fish Example Contrasts

Similarly, Contrast L_2 :

$$\hat{L}_2 = \bar{X}_4 - \frac{1}{2}(\bar{X}_2 - \bar{X}_3) = 132.2 - \frac{1}{2}(65.8 - 83.4)$$

The coefficients are: $\lambda_1 = 0$; $\lambda_2 = -1/2$; $\lambda_3 = -1/2$; $\lambda_4 = +1$

The sum of squares of $\Sigma\lambda$ (i.e. $\Sigma\lambda^2$) = 1.5

$$SE(L_1) = \sqrt{s^2 \times \frac{(\lambda)^2}{n_i}} = \sqrt{344.25 \times \frac{1.5}{5}} = 10.16$$

~~diet4 dif avg d2~~
c13

$$t_{\text{calc}} = 6.535$$

More Experimental Designs

- CRD with subsampling
- Randomized Complete Block Design (RCBD)
- Analysis of Covariance (ANCOVA)

Subsampling

- the term used to describe the situation in which more than one observation is taken per experimental unit.
- such observations are made on sampling units.
- when subsampling is performed, the linear model and the ANOVA must be expanded to take into account the variation among samples (the source of sampling error)

CRD with Subsampling: Model I (Treatment Effects)

Consider the following experiment:

- A new drug phenytoin-HCl, was thought to enhance liver cyt-P450 in people with late-stage cirrhosis
- when given with alcohol the effect was thought to be diminished.
- a total of 6 randomly assigned patients were used to test this drug:
 - 2 controls (no drug)
 - 2 phenytoin-HCl
 - 2 phenytoin-HCl + ethanol
- After an appropriate time and dose, 4 liver biopsies were taken under ultrasound/MRI (co-registered) guidance from each patient and cyt-P450 was measured in piece.





(Source: <http://www.skills.uct.ac.za/activities.htm>)

Experimental Objectives:

- 1) To determine if there was a significant difference among the three treatments
- 2) To estimate the 2 variance components:
 - variation among measurements within patients
(i.e. sampling error).
 - variation among patients within a given treatment
(i.e. experimental error).

CRD with Subsampling: Model I

The data ($Y_{i,j,k}$) from 6 randomly chosen/assigned patients and 4 randomly selected pieces of liver from each.

Patients	Control		Phelphodyne		Phelphodyne +EtOH	
	1	2	1	2	1	2
cyt-P450 readings	131	148	157	152	124	140
	130	143	153	155	125	138
	125	150	154	162	136	138
	131	150	149	161	130	139
	517	591	613	630	515	555
	1108		+ 1243	+ 1070		

$$Y_{ij\cdot} \quad Y_{i..}$$

$$\sum Y_{ij\cdot}^2 = 1962489$$

$$\sum Y_{i..}^2 = 3917613$$

$$Y_{...} = 3421$$

CRD with Subsampling: Model I

Notes:

- an experimental unit here is a patient
- a sampling unit is a piece of patient liver

Degrees of Freedom

- $i = 3$ (treatments)
- $j = 2$ (replicates)
- $k = 4$ (subsamples)
- Total df = $(3 \times 2 \times 4) - 1 = 24 - 1 = 23$

i j k

There are 2 hypothesis that can be tested:

- 1) $H_0: \tau_i = 0$, for all i vs. $H_A: \tau_i \neq 0$ (i.e. treatment effect)
- 2) $H_0: \sigma_e^2 = 0$ vs. $H_A: \sigma_e^2 \neq 0$ (i.e. error variance)

Compute Sums of Squares:

$$Total(SS) = \sum Y_{ijk}^2 - \frac{(\bar{Y}_{...})^2}{i \times j \times k} = 490875 - \frac{3421^2}{24} = 3239.96$$

$$SS(\text{Treatments}) = SS(T) = \frac{\sum (\bar{Y}_{i..})^2}{j \times k} - \frac{(\bar{Y}_{...})^2}{i \times j \times k} = 2066.5833$$

$$SS(\text{Subsamples}) = SS(SS) = \frac{\sum Y_{ij..}^2}{k} - \frac{(\bar{Y}_{...})^2}{i \times j \times k} = 2987.2083$$

$$SS(\text{sampling error}) = SS(SE) = Total(SS) - SS(SS) = 252.75$$

$$SS(\text{experimental error}) = SS(EE) = SS(SS) - SS(T) = 920.625$$

ANOVA Table



Source	df	SS	MS	F _c	F _{i,j,a}
subsamples (AMONG PATIENTS)	5	2987.2083	-----	-----	-----
Treatment	2	2066.5833	1033.2417	3.3674	9.55
Exp. Error	3	920.625	306.875	21.85	3.16
Samp. Error	18	252.75	14.042	-----	-----
TOTAL	23	3239.9583	-----	-----	-----

*Note treatment df
+ EE df =
sub samp df

$$df_{TOTAL} = (i \cdot j \cdot k) - 1 = (3 \times 2 \times 4) - 1 = 23$$

$$df_T = i - 1 = 3 - 1 = 2$$

$$df_{SS} = (i \cdot j) - 1 = (3 \times 2) - 1 = 5$$

$$df_{EE} = i(j-1) = 3(2-1) = 3$$

$$df_{SE} = ij(k-1) = 3 \times 2 \times (4-1) = 18$$

ANOVA Analysis

$$F_c \text{ (among treatments)} = 1033.2417 / 306.875 = 3.367$$

Since $F_{0.05,2,3} = 9.55$

we ~~Accept~~ $H_0: \tau_i = 0$. Therefore there is ~~No~~ effect of treatment.

$$F_c \text{ (experimental error)} = 306.875 / 14.042 = 21.854$$

Since $F_{0.05,3,18} = 3.16$

We ~~REJECT~~ $H_0: \sigma_e^2 = 0$. Therefore there is ~~A~~ significant source of error between patients.

REMEMBER CRD what MS is estimating !!

	Model I	Model II
treatment	$S^2 + \frac{j}{i-1} \bar{t}_i^2$	$S^2 + jS_t^2$
error	S^2	S^2
treatment	$S^2 + \frac{\bar{a}_j(t_i)^2}{i-1}$	$S^2 + \frac{n - (\bar{a}_j^2)/n}{i-1} S_t^2$
error	S^2	S^2

CRD with Subsampling

For Fixed (Type I) Models, what does the mean (MS) estimate?

Treat Table

exp err

Samp error

ANOVA Table

$$E[MS(T)] = S^2 + kS_e^2 + \frac{j \times k}{i - 1} \bar{t}_i^2$$
$$E[MS(EE)] = S^2 + kS_e^2$$
$$E[MS(SE)] = S^2$$

Estimation of variance components:

$$\hat{S}_e^2 = s_e^2 = (306.875 - 14.042)/4 = 73.208$$

MS_{EE} - MS_{SE}

$$\hat{S}^2 = s^2 = 14.042 \quad S^2 + S_e^2 = 87.250$$

Where is variance?

σ^2 = variation among cyt-P₄₅₀ within patients

s_ϵ^2 = variation in patients within treatments

$\sigma^2 + s_\epsilon^2$ = total variance (within and among patients within a treatment)



The variation within patients (i.e. among liver samples) accounts for $100 \times (14.042 / 87.250) = 16.1\%$ of the estimated total variance.

The variation among patients of a given treatment accounts for $100 \times (73.208 / 87.250) = 83.9\%$ of the estimated total variance.

i.e. There is approximately 5.1x as much variation among patients as there is within patients.

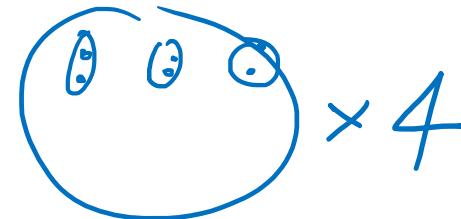
CRB SS NOTES:

- The critical F Value (FC) value for treatments was noted to be quite small. This is due to the large denominator
- The large MS(EE) indicates high variability among the patients for any one preparation.
- Although no differences in treatments were detected, any differences may actually have been hidden by the variation among experimental units.
- Future experiments should take into account heterogeneity of experimental units.
Also could performing a block design in which 2 blocks of 3 patients are examined too.

CRD with Subsampling (Model II- Random Effects)

Consider this experiment:

- an experiment was conducted to assess the precision to which EEG could be measured for application in a brain-computer interface.
 - 4 people were randomly selected
 - 3 brain regions were randomly chosen from each person
 - 2 samples taken per region for signal power analysis



Primary Objectives:

- 1) Estimate EEG signal power
- 2) Find the EEG signal power standard error
- 3) Isolate and estimate the sources of variation.

CRD with SS (Model II-Random Effects)

From this analysis:

- recommendation can be made with regard to optimizing future brain EEG sampling strategies.
- This would allow the researcher to reduce the standard error of their estimate in future studies.

Person	k	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}	k_{11}	k_{12}	k_{13}	k_{14}	k_{15}	k_{16}	k_{17}	k_{18}	k_{19}	k_{20}	k_{21}	k_{22}	k_{23}
Brain Region	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
Subsample	3.48	3.72	3.03	2.66	2.07	2.39	2.97	3.94	2.75	3.98	4.27	3.51	3.29	3.68	3	2.64	2.12	2.39	2.86	3.64	2.75	4.07	4.32	3.51
	6.77	7.4	6.03	5.3	4.19	4.78	5.83	7.58	5.5	8.05	8.59	7.02												
	20.2			14.27			18.91			23.66														

$$Y_{ij\cdot}$$

$$Y_{i\cdot\cdot}$$

$$Y_{\dots} = 77.09$$

$j = 4$ pat .

$j = 3$ region

$k = 2$ sub samp

total df = (ijk) - 1 = 23

CRD II SS Model Equation:

$$Y_{ijk} = m + t_i + e_{ij} + d_{ijk}$$

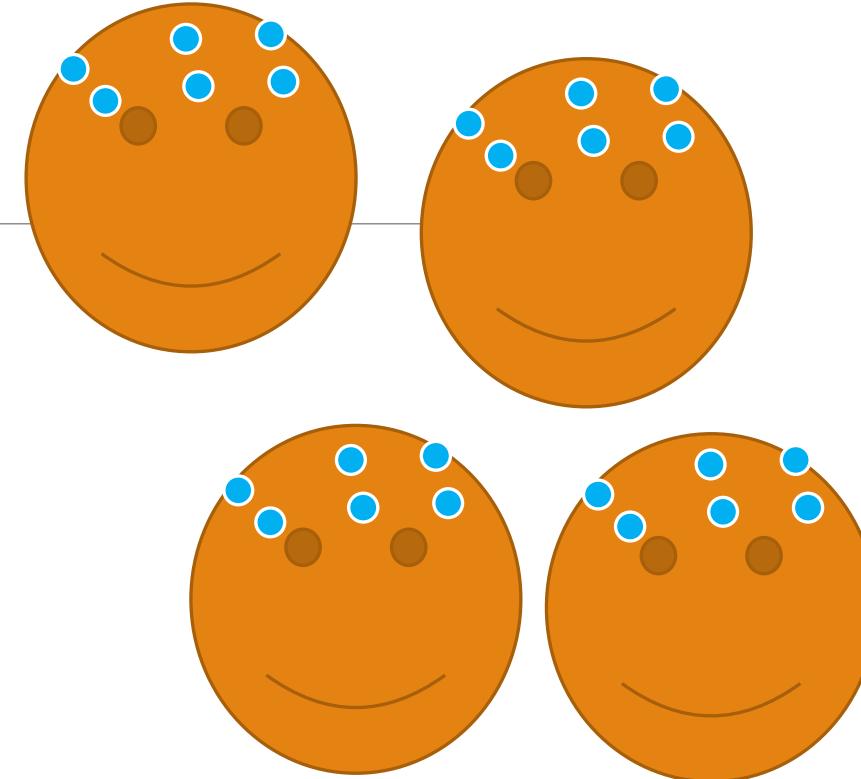
Treatment Effect

Experimental Error

Sampling Error

Necessary assumptions:

- τ_i are $N(0, \sigma_\tau^2)$, where σ_τ^2 is the variation between people
- ε_{ij} are $N(0, \sigma_\varepsilon^2)$, where σ_ε^2 is the variation between brain areas
- d_{ijk} are $N(0, \sigma_d^2)$, and σ_d^2 is variation among samples between brain areas.



CRD II SS ANOVA

Source	df	SS	MS	Fc	Fi,j, α	
subsamples (AMONG AREAS)	11	10.19055	---	---	---	
Treatment (AMONG PEOPLE)	3	7.56035	2.5201167	7.665	1 \mathcal{T}	4.07
Exp. Error	8	2.6302	0.320775	49.41	2 \mathcal{E}	2.85
Samp. Error	12	0.07985	0.0066542	---	---	
TOTAL	23	10.2704	---	---	---	

*Note treatment df +
EE df = subsamp df

Testable hypothesis

$H_0: \sigma_\tau^2 = 0$ (i.e. no significant difference between people)

$H_0: \sigma_\epsilon^2 = 0$ (i.e. no significant difference between brain areas)

CRD II SS ANOVA Results:

- variance component due to brain areas within people is significantly different from 0?

49.41 > 2.85

reject H_0 EEG power varies
between areas.

$H_0 \delta\sigma^2 = 0$

- variance component from person to person is significantly different from 0?

7.665 > 4.07

reject H_0 EEG power varies b/w
people

H

In CRD II with Subsampling

Random Effects (Type II) Models: what does the mean (MS) estimate?

$$E[MS(T)] = S_d^2 + kS_e^2 + jkS_t^2$$

people

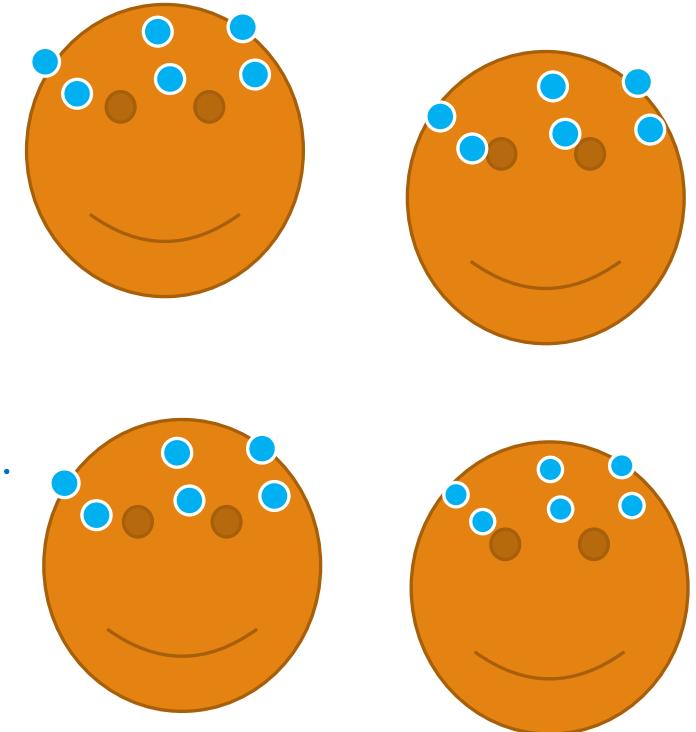
$$E[MS(EE)] = S_d^2 + kS_e^2$$

people

$$E[MS(SE)] = S_d^2$$

among areas

within areas



Estimates of the 3 variance components:

within
area

$$\hat{S}_d^2 = s^2 = MS(SE) = 0.0066$$

among
area

$$\hat{S}_e^2 = s_e^2 = \frac{MS(EE) - MS(SE)}{k} = \frac{0.328775 - 0.0066542}{2} = 0.1611$$

among
PPI

$$\hat{S}_\tau^2 = s_\tau^2 = \frac{MS(T) - MS(EE)}{jk} = \frac{2.5201167 - 0.320775}{6} = 0.365232$$

$$Total = s^2 + s_e^2 + s_\tau^2 = 0.0066 + 0.1611 + 0.3652 = 0.5329$$

CRD II with Subsampling Conclusions

Variation within brain areas represents = $100 * (0.0066 / 0.5329) = 1.24\%$ ✓

Variation among brain areas represents = $100 * (0.1611 / 0.5329) = 30.22\%$

Variation among people represents = $100 * (0.3652 / 0.5329) = 68.54\%$

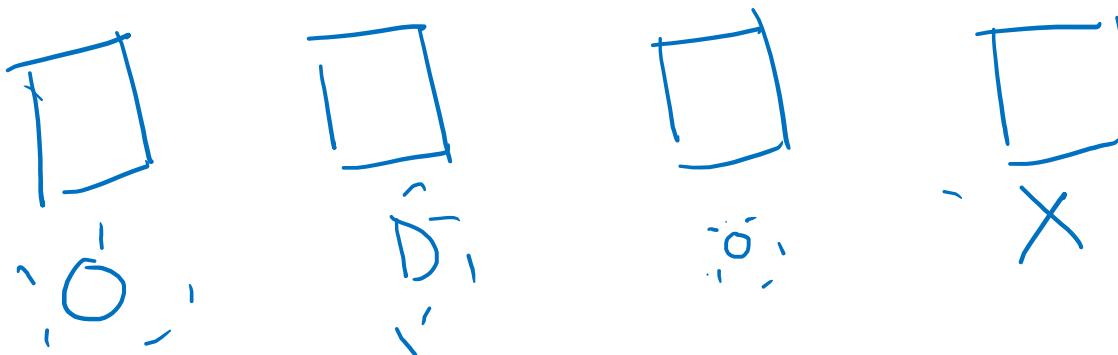
Estimate the total mean EEG power = 3.212 ± 0.01665

$$\text{Standard Error} = SE = \sqrt{\frac{MS(EE)}{n}} = \sqrt{\frac{0.0066542}{24}} = 0.01665$$

move ppl

Randomized Complete Block Design

- removes source of variation
- if it is known in advance that the experimental units are NOT homogeneous then the CRD is no longer appropriate.
- the RCBD is used to remove sources of heterogeneity among experimental units.
- here experimental units are allocated to blocks such that those assigned to the same block should be similar in response to their treatment (i.e. homogeneous as possible).

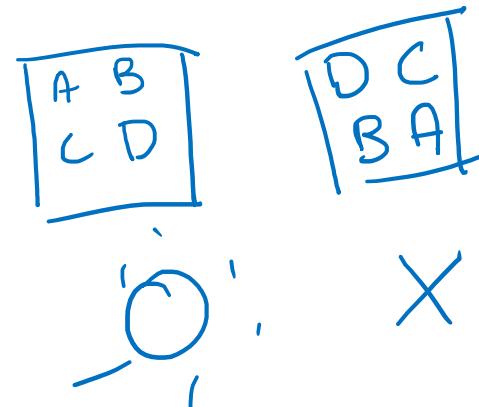


RCBD

- treatments are then allocated to the experimental units of each block, by a separate randomization that is carried out within each block.

Some blocking factors could include:

- DATE of experiment
- cage battery for animal housing
- plot of land
- incubator or oxygen chamber
- individual hospital



RCBD Fruit Fly Example

Consider the following genetics experiment.

- 5 people in the lab were all assigned to a project where they were to assess protein levels of CuZnSOD (superoxide dismutase) in *Drosophila melanogaster* (fruit fly) that had been transfected with human CuZnSOD. The 'boss' wanted to know which cross (i.e. homozygote (hom), heterozygote (het), or wild type (wld)) had higher CuZnSOD.

type(i)	PERSON (j)					\bar{Y}	$Y_{i\cdot}$	$\sum Y_{i\cdot}^2$
	1	2	3	4	5	40.078	200.39	8039.948
HOM	39.87	38.16	42.08	40.84	39.44	42.79	213.95	9161.273
HET	42.51	40.82	44.17	43.46	42.99	45.206	226.03	10224.2
WLD	45.76	43.14	46.29	44.95	45.89			
	128.14	122.12	132.54	129.25	128.32			
	5490.695	4983.518	5864.479	5577.18	5509.546	640.37	27425.42	
$Y_{\cdot j}$								
$\sum Y_{\cdot j}^2$								

RCBD Fruit Fly Example

For this RCBD design:

A block is an individual person τ^B

A treatment is a genotype τ^T

An experimental unit is a fruit fly ϵ^E

A model equation for the RCBD design:

$$Y_{ij} = \mu + \tau_i + B_j + \epsilon_{ij}$$

Null hypothesis: H_0 : all $\tau_i = 0$

Alternative: H_A : all $\tau_i \neq 0$

Null hypothesis: H_0 : all $B_j = 0$

Alternative: H_A : all $B_j \neq 0$

RCBD Fruit Fly Example

$$\begin{aligned}
 df_{\text{block}} &= j - 1 = 5 - 1 = 4 \\
 df_{\text{treat}} &= i - 1 = 3 - 1 = 2 \\
 df_{\text{total}} &= (i \cdot j) - 1 = 3 \cdot 5 - 1 = 14 \\
 df_{\text{error}} &= \text{total} - (\text{block} + \text{treat}) = (i-1)(j-1) = 8 \\
 i &= 3 \text{ (treatments); } j = 5 \text{ (blocks)}
 \end{aligned}$$

$$\begin{aligned}
 Total(SS) &= \sum Y_{\bullet j}^2 - \frac{\bar{Y}_{\bullet \bullet}^2}{ij} = 27425.417 - \frac{(640.37)^2}{5 \cdot 3} = 87.168
 \end{aligned}$$

$$\begin{aligned}
 SS(\text{Blocks}) &= \frac{\sum Y_{\bullet j}^2}{i} - \frac{\bar{Y}_{\bullet \bullet}^2}{ij} = \frac{(128.14)^2 + \dots + (128.32)^2}{3} - = 18.948
 \end{aligned}$$

RCBD Fruit Fly Example

$$\begin{aligned}df_{\text{block}} &= j - 1 = 5 - 1 = 4 \\df_{\text{treat}} &= i - 1 = 3 - 1 = 2 \\df_{\text{total}} &= (i \cdot j) - 1 = 3 \times 5 - 1 = 14 \\df_{\text{error}} &= \text{total} - (\text{block} + \text{treat}) = (i-1)(j-1) = 8 \\i &= 3 \text{ (treatments); } j = 5 \text{ (blocks)}\end{aligned}$$

$$SS(\text{Treat}) = \frac{\sum Y_{i \cdot}^2}{j} - \frac{Y_{\cdot \cdot}^2}{ij} = \frac{(200.39)^2 + (213.95)^2 + (226.03)^2}{5} - \frac{(640.3)^2}{3 \cdot 5} = 65.814$$

$$SS(\text{Error}) = \text{Total}(SS) - [SS(\text{blocks}) + SS(\text{Treat})]$$

$$= 87.168 - [18.948 + 65.814]$$

RCBD Fruit Fly Example

Source	df	SS	MS	F_c	$F_{i,j,\alpha}$
Block _{person}	4	18.947706	4.73926	15.76	3.84
treatment _{type}	2	65.81397	32.906987	109.4	4.46
Exp. Error	8	2.406194	0.3007743	----	----
TOTAL	14	87.167873	----	----	----

Here $F_{i,j,\alpha} = F_{2,8,0.05} = 4.46$

Conclusion: Reject H_0 , (i.e. $F_c > F_{\text{table}}$) - - l.f.
i.e. treatments are different !

Here $F_{i,j,\alpha} = F_{4,8,0.05} = 3.84$

Conclusion: Reject H_0 , (i.e. $F_c > F_{\text{table}}$)
i.e. Blocks are NOT homogeneous !

RCBD Fruit Fly Example

If did not block on person, we could potentially contaminate the real source of differences in the data with differences between the people's lab techniques.

So, where's the source of differences ?

- can use lsd, Scheffe, etc.

e.g. critical value for lsd:

$$lsd = t_{v, \alpha/2} \times \sqrt{\frac{E(MS) \times \frac{2}{j}}{5}} = 2.306 \times \sqrt{0.3007743 \times \frac{2}{5}} = 0.799852$$

Where $t_{v, \alpha/2} = t_{8, 0.025} = 2.306$

v (nu) = df for $E(MS) = (i-1)(j-1) = 8$

RCBD Fruit Fly Example

	HET (42.79)	WLD (45.206)
HOM (40.078)	2.712*	5.128*
HET (42.79)	-----	2.416 *

* = significantly different

Final Statements

- the ‘boss’ had hoped CuZnSOD transfection would work. Obviously it didn’t !!
- If anything the resultant flies had less.
- the lab has 5 people with significantly differing technical skills.

Differences between CRD and RCBD

- if we didn't block on person E(MS) would have been equal to $21.3539/12 = 1.7795$
- therefore the $F_c = 18.49$ (re: $F_{table} = F_{2,12,0.05} = 3.89$).
- the CRD doesn't partition the Error.

Let's revisit those trout fry:

Before, we had 4 diets, 5 fish/diet. Now let's suppose we obtained 4 fish (1/diet) from each of 5 hatcheries.

- this time, 4 fish had been randomized to 4 diets at each fish hatchery.
- fish are homogeneous in their response to treatment, but the hatchery may be a source of heterogeneity we wish to remove.

Fish Example: CRD

Example: Want to examine the effect of 4 diets on the growth of rainbow trout fry.

Dietary Treatment				
1	2	3	4	
59	70	93	124	
47	59	85	135	i = 4 treatments
40	52	79	167	j = 5 replicates
32	87	72	83	$\sum_j Y_{ij} = Y_{i\bullet} =$
39	61	88	152	$217 \quad 329 \quad 417 \quad 661 \rightarrow 1624$
217	329	417	661	$= Y_{\bullet\bullet}$
				$\sum Y_{ij}^2 = 9835 \quad 22375 \quad 35043 \quad 91483 \rightarrow 158736$
				$\bar{Y}_{i\bullet} = 43.4 \quad 65.8 \quad 83.4 \quad 132.2 \rightarrow 81.2$
				$= \bar{Y}_{\bullet\bullet}$

Replicates
(mass in grams)

Fish Example: RCBD

Example: Want to examine the effect of 4 diets on the growth of rainbow trout fry in n 4 different hatcheries

i = 4 treatments

j = 5 blocks

Hatchery	Dietary Treatment				$\sum_1^i Y_{ij} = Y_{\bullet j}$
Replicates (mass in grams)	1	2	3	4	
A	59	70	93	124	346
B	47	59	85	135	326
C	40	52	79	167	338
D	32	87	72	83	274
E	39	61	88	152	340
	217	329	417	661	totals
					$\sum_1^j Y_{ij} = Y_{i\bullet} = 217 \quad 329 \quad 417 \quad 661$
					$\sum Y_{ij}^2 = 9835 \quad 22375 \quad 35043 \quad 91483$
					$\bar{Y}_{i\bullet} = 43.4 \quad 65.8 \quad 83.4 \quad 132.2$
					$1624 \quad 158736 \quad 81.2 = \bar{Y}_{..}$

Compute Sums of Squares CRD:

$$Total(SS) = \sum Y_{ij}^2 - \frac{(Y_{..})^2}{i \times j} = 158736 - \frac{1624^2}{4 \times 5} = 26867.2$$

SS(Treatments)=

$$SS(T) = \frac{\sum (Y_{i.})^2}{j} - \frac{(Y_{..})^2}{i \times j} = \frac{217^2 + 329^2 + 417^2 + 661^2}{5} - \frac{1624^2}{4 \times 5} = 21359.2$$

$$SS(error) = SS(E) = Total(SS) - SS(T)$$

ANOVA: CRD

Source	df	SS	MS	F_c	$F_{i,j,\alpha}$
treatment	3	21359.2	7119.73	20.68	3.24
error	16	5508.0	344.25	-----	-----
Total	19	26867.2	-----	-----	-----

$$df_{\text{treatment}} = (\# \text{ treatments} - 1) = (i - 1)$$

$$df_{\text{TOTAL}} = (i \times j) - 1$$

$$df_{\text{error}} = df_{\text{TOTAL}} - df_{\text{treatment}}$$

$$\begin{aligned} MS &= SS / df \\ F_c &= T(MS) / E(MS) \end{aligned}$$

$F_{i,j,\alpha}$ = from table

e.g. $F_{3,16,0.05} = 3.24$

ANOVA: RCBD

Source	df	SS	MS	F_c	$F_{i,j,\alpha}$
Block _{hatchery}	4	859.2	214.8	0.554	3.26
treatment _{diet}	3	21359.20	7119.733	18.38	> 3.49
Exp. Error	12	4648.8	387.4		
Total	19	26867.2	-----	-----	-----

Here $F_{i,j,\alpha} = F_{3,12,0.05} = 3.49$ Conclusion: Reject H_0 , (i.e. $F_c > F_{\text{table}}$)

The difference?

- Error SS from CRD gets divided up into Error(SS) & Block(SS) in the RCBD design.
- The RCBD design removes some of the experimental error as error due to block effect.

CRD

Source	df	SS	MS	F _c	F _{i,j,a}
treatment	3	21359.2	7119.73	20.68	3.24
error	16	5508.0	344.25	-----	-----
Total	19	26867.2	-----	-----	-----

RCBD

Source	df	SS	MS	F _c	F _{i,j,a}
Block _{hatchery}	4	859.2	214.8	0.554	3.26
treatment _{diet}	3	21359.20	7119.733	18.38	3.49
Exp. Error	12	4648.8	387.4	-----	-----
Total	19	26867.2	-----	-----	-----

Assessing the Efficiency of Blocking

$$\hat{\sigma}_{RCBD}^2 = MS(E) = 387.4$$

$$\hat{S}_{CRD}^2 = \frac{(j-1)s_{block}^2 + j(i-1)s^2}{(i-1)(j-1)} = \frac{(5-1)214.8 + 5(4-1)387.4}{(4-1)(5-1)} = 344.25$$

$$\frac{\hat{\sigma}_{RCBD}^2}{\hat{\sigma}_{CRD}^2} = \frac{387.4}{344.25} = 1.125$$

- if this ratio > 1.0 then the RCBD is not any more efficient.

Fruit Fly Example

$$\hat{\sigma}_{RCBD}^2 = MS(E) = 0.3007743$$

$$\hat{\sigma}_{CRD}^2 = \frac{(j-1)s_{block}^2 + j(i-1)s^2}{(i-1)(j-1)} = 1.7795$$

$$\frac{\hat{\sigma}_{RCBD}^2}{\hat{\sigma}_{CRD}^2} = \frac{0.3007743}{1.7795} = 0.169 = 16.9\%$$

Interpretation:

- a CRD design with, say, 100 experimental units not assembled into blocks will give answers that are about as precise as those for a RCBD with about 17 experimental units !!

RCBD with Subsampling

weanling rats fed 5 diets for 2 weeks

A = ZnDF

B = PEM

C = ZnPf

D = ZnAL

E = +ZnAL

- measured final weight (all started at exactly the same weight)

BLOCK	DIET					Totals	
	A	B	C	D	E		
1	72	82	110	117	138	519	56741
	61	87	105	103	116	472	46380
	58	79	99	110	113	459	44275
subtotal	191	248	314	330	367	1450	
2	54	82	106	117	127	486	50694
	55	76	97	108	119	455	44035
	61	80	102	114	131	488	50682
subtotal	170	238	305	339	377	1429	
3	65	83	110	122	139	519	57419
	53	80	99	104	117	453	43515
	50	75	98	125	125	473	48979
subtotal	168	238	307	351	381	1445	442720
TOTALS	529	724	926	1020	1125	4324	

Y...

$\sum Y_{ij}^2$

RCBD with Subsampling Model Equation

Assumptions:

1) ε_{ij} are $N(0, s\varepsilon^2)$

2) $\sum \beta_j = 0$

3) $\sum \tau_i = 0$

4) δ_{ijk} are $N(0, s\delta^2)$

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} + \delta_{ijk}$$

↓
treat
↑ Block
↓
↑ SS

How many parameters does the model try to fit ?

$$= (\# \text{ treatments}) + (\# \text{ blocks}) + 3 = 11$$

(i.e. $\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \beta_1, \beta_2, \beta_3, \mu, \sigma\delta^2, \sigma\varepsilon^2$)

RCBD with Subsampling :Calculations

$$Total(SS) = \sum Y_{ijk}^2 - \frac{Y_{\dots\dots\dots}^2}{ijk} = \frac{4324^2}{(5)(3)(3)} = 442720 - 415488.36 = 27231.65$$

i = 5 (treatments) = diet
j = 3 (blocks) = cage battery
k = 3 (subsamples) = rat

$$SS(AmongRats) = \frac{\sum Y_{ij\dots}^2}{k} - \frac{Y_{\dots\dots\dots}^2}{ijk} = \frac{191^2 + 248^2 + \dots + 381^2}{3} - \frac{4324^2}{(5)(3)(3)} = 25600.98$$

$$SS(Error) = Total(SS) - SS(AmongRats) = 27231.65 - 25600.98 = 1630.67$$

$$SS(Blocks) = \frac{\sum Y_{\dots j\dots}^2}{ik} - \frac{Y_{\dots\dots\dots}^2}{ijk} = \frac{1450^2 + 1429^2 + 1445^2}{(5)(3)} - \frac{4324^2}{(5)(3)(3)} = 16.04$$

$$SS(T) = \frac{\sum Y_{i\dots\dots}^2}{jk} - \frac{Y_{\dots\dots\dots}^2}{ijk} = \frac{529^2 + 724^2 + \dots + 1175^2}{(3)(3)} - \frac{4324^2}{(5)(3)(3)} = 25346.98$$

$$SS(ExpError) = SS(AmongRats) - [SS(Blocks) + SS(T)] = 25600.98 - [16.04 + 25346.98] = 237.96$$

RCBD with SS ANOVA Table:

Source	df	SS	MS	F_c	$F_{i,j,\alpha}$
Among ExpUnits _{Rats}	14	25600.98	-----	-----	-----
Block _{battery}	2	16.04	4.01	-----	-----
treatment _{diet}	4	25346.98	6336.74	213.07	3.84
Exp. Error	8	237.96	29.745	0.5479	2.27
ERROR	30	1630.67	54.36	-----	-----
TOTAL	44	27231.65	-----	-----	-----

$$df_{\text{expUnits}} = ij-1 = (5)(3)-1 = 14 \quad df_{\text{ExpError}} = (i-1)(j-1) = 4 \times 2 = 8$$

$$df_{\text{Block}} = j-1 = 2$$

$$df_{\text{treatment}} = i-1 = 4$$

$$df_{\text{Error}} = ij(k-1) = (5)(3)(3-1) = 30$$

$$df_{\text{Total}} = ijk-1 = (5)(3)(3)-1 = 44$$

RCBD with Subsampling: ANOVA

$$F_C(\text{treatment}) = \frac{MS(T)}{MS(EE)} = \frac{6336.74}{29.745} = 213.07$$

$$F_C(\text{ExpError}) = \frac{MS(EE)}{MS(E)} = \frac{29.745}{54.36} = 0.5427$$

- To test the null hypothesis of no differences between treatments one should use $MS(T)/MS(EE)$ as above.
- However, if $MS(EE) \leq MS(E)$ it is recommended that you should use the pooled error:

$$MS(\text{PooledError}) = \frac{SS(EE) + SS(E)}{df_{EE} + df_E} = \frac{237.96 + 1630.67}{8 + 40} = 49.2$$

RCBD with Subsampling: ANOVA

$$F_C(\text{treatment}_{\text{pooled}}) = \frac{MS(T)}{MS(PE)} = \frac{6336.74}{49.2} = 128.86$$

This parameter would be used to test the null hypothesis concerning treatment effects.

- this situation could happen when the variation among experimental units is insignificant, and error is only within experimental units, as measured by the sampling error (i.e. $MS(E)$).
- in other words $MS(E)$ and $MS(EE)$ are essentially measuring the same thing (i.e. the quantity σ^2)

RCBD with Subsampling: MS

$$MS(B) = \sigma^2 + k\sigma_{\varepsilon}^2 + \frac{ik}{(j-1)} \sum \beta_j^2 = \sigma^2 + 3\sigma_{\varepsilon}^2 + 7.5 \sum \beta_j^2$$

$$MS(T) = \sigma^2 + k\sigma_{\varepsilon}^2 + \frac{jk}{(i-1)} \sum \beta_j^2 = \sigma^2 + 3\sigma_{\varepsilon}^2 + 2.25 \sum \beta_j^2$$

$$MS(EE) = \sigma^2 + k\sigma_{\varepsilon}^2$$

$$MS(E) = \sigma^2$$

RCBD with Subsampling: Results

The best estimate of $\sigma\epsilon^2 = 0$

→ This is because we accepted the null hypothesis that experimental error was not significant.

(i.e. $F_c < F_{table} = 0.5479 < 2.27$)

The best estimate of σ^2 is the pooled error 49.2

Is there a significant difference among diets ?

→ YES reject H_0 since $F_C > F_{table}$ (i.e. $213.1 > 3.84$)

RCBD with Subsampling: Results

So, there is a significant diet effect. Where is(are) the differences?

To evaluate the nature of these differences one can use the lsd method:

$$lsd = t_{v,\alpha/2} \times \sqrt{MS(EE) \times \frac{2}{jk}} = 2.306 \times \sqrt{29.74 \times \frac{2}{(3)(3)}} = 5.9282$$

Note that k is added, only in subsampling

Note: other uses of the lsd method uses $E(MS)$, the error mean SS. For subsampling the $MS(\text{EE})$ is used instead.

RCBD with Subsampling: Post Hoc

Diet	B _{PEM}	C _{ZnPF}	D _{ZnAL}	E _{+ZnAL}
A _{ZnDF}	21.666*	44.111*	54.555*	66.222*
B _{PEM}		22.445*	32.889*	44.556*
C _{ZnDF}			10.444*	22.111*
D _{ZnAL}				11.667*

* = significant

Where,

$$A_{ZnDF} = 58.778 \text{ g}$$

$$B_{PEM} = 80.444 \text{ g}$$

$$C_{ZnPF} = 102.889 \text{ g}$$

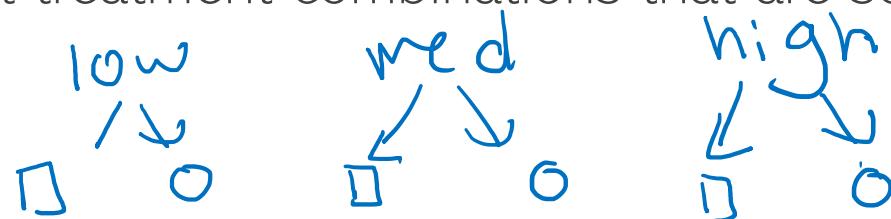
$$D_{ZnAL} = 113.333 \text{ g}$$

$$E_{+ZnAL} = 125.0 \text{ g}$$

Other Important Designs

1). Factorial Design

- to this point only one factor has been investigated (e.g. diet on weight gain)
- what about treatment combinations that are somehow jointly responsible for the response



2). Analysis of Covariance (ANCOVA)

- any of the other models. However, include a covariate term (e.g. initial age, initial weight, scalp/skin impedance, SNR, etc.)

Introduction

TAYLOR DEVET MASC.

PHD. CANDIDATE BIOLOGICAL AND BIOMEDICAL ENGINEERING

MCGILL UNIVERSITY

SHRINERS HOSPITAL FOR CHILDREN

A bit about me...

BEng Electrical and Biomedical Engineering Coop, McMaster (2018)

2015-2017 Venture Engineering and Science

2014-2018 Welcome Week Rep, Cooc

2014-2020 TA ENG1D04, ELEC ENG 4BC3, IBIO 1P10, IBIO 4P04

MASc. Biomedical Engineering McMaster (2020)

Development of an electrical stimulation device for bone cells

Supervisor Dr Greg Wohl

PhD Biological and Biomedical Engineering McGill (2021-)

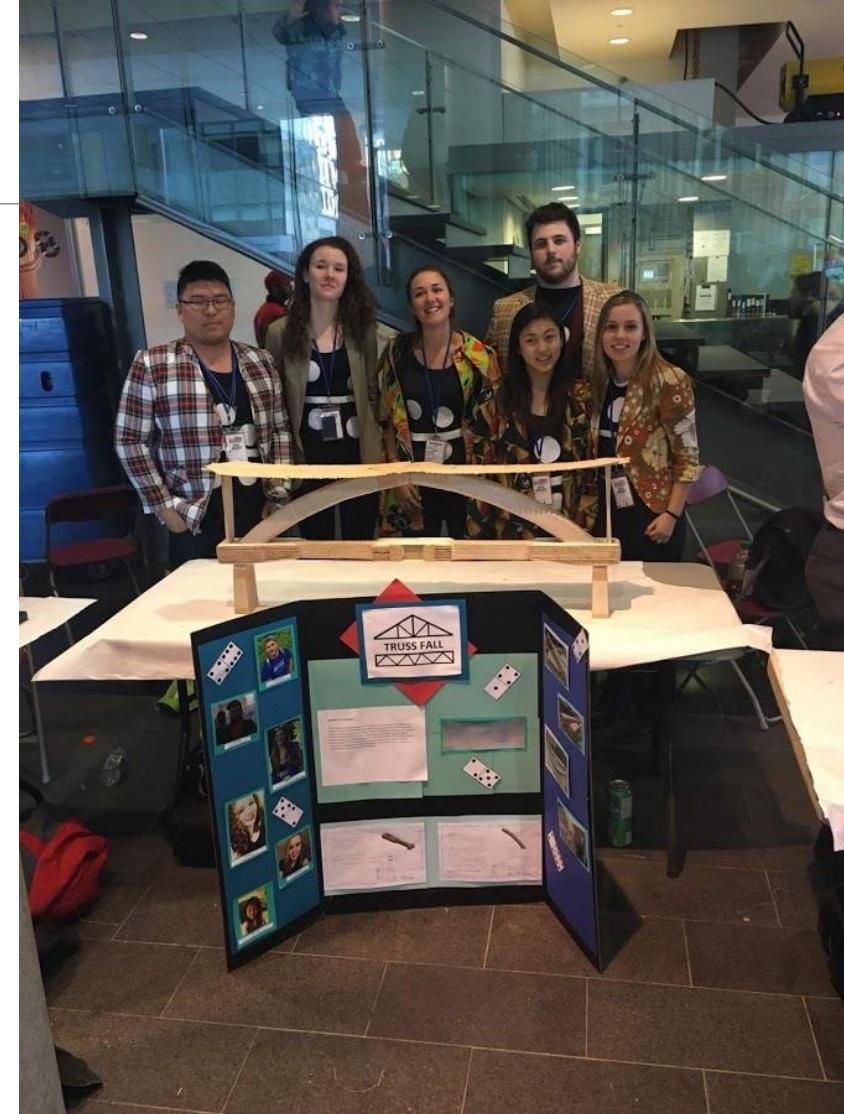
Investigating the osteocyte-lacuna-canalicular network

Spaceflight, Craniosynostosis, and SOST



Revamped Civil Structure: 2019

Troitsky Bridge Building 2016



Contact Information

Email: devett@mcmaster.ca

Campus Office: n/a

Office Hours

Fridays 10:30am - 12:30 noon (Teams)

Teaching Assistants:

Noor Abu Jarad abujaran@mcmaster.ca

Andrew Lofts loftsa@mcmaster.ca

Course Content

From the Undergraduate Calendar:

Introduction to experimental design and variance associated with biological systems and analysis of biological data, mathematical and engineering methods for describing and predicting the behaviour of biological systems; statistical models of biological functions;

- Variance within biological systems and measurement systems
- Modelling basics and assumptions
- Types of models and systems
- Machine Learning

Course Objectives

Mathematical approaches to modelling biological systems and the challenges associated with it.

- Understanding biological data
- Statistical significance of models

Linear time invariance (LTI)

- short-time Fourier transform (STFT)
- Wavelets
- PCA/ICA
- Nonlinear dynamics (fractal and chaotic models)

Real life examples

- imaging systems
- physiological recording systems

Course Content

- 1) Course intro & Introduction to modelling
- 2) Biostatistical analysis , regression, Experimental design
- 3) CRD RCBD Post hoc testing
- 4) Cardiac modelling
- 5) Pharmacokinetics (1, 2, and 3 compartments), indicator dilution, contrast agents and other tracers used for assessing microvascular and metabolic kinetics.
- 6) PCA ICA
- 7) Chronobiology , LTI
- 8) Fourier, STFT, Wavelets
- 9) Scaling allometry Chaos and fractals
- 10) Chaos cont. , phase space , attractors
- 11) Machine Learning

Course Schedule and Location

One 3hr lecture and one tutorial weekly.

Tutorials: Wednesday.1:30pm-2:20pm BSB-137

Lectures: Monday .7:00pm–10:00pm Teams

Textbooks and Recommended Readings

There is no one book to cover all the material for this course. Necessary papers and readings to complement the course material will be posted to the course website as suggested readings. The following texts may be helpful:

- Modeling and Simulation in Medicine and the Life Sciences 2nd Edition – Frank Hoppensteadt, Chasles S.Peskin
 - Dynamic Systems, Biology Modeling and Simulation – Joseph DiStefano III
 - Biosignal and Medical Image Processing 3rd Edition - John L. Semmlow, Benjamin Griffel
 - Physiological Control Systems – Michael C.K. Khoo
- Class notes will be posted on the course website before 12:00noon the day of the lecture: The class website is on Avenue to Learn: <http://avenue.mcmaster.ca/>

Grading

Individual assignments (4) worth 5% each	20%
#1 due October 3	
#2 due October 17	
#3 due November 7	
#4 due November 21	
Quizzes (12, drop 2)	5%
Group Analysis Project	25%
In-class group presentation	
Midterm (October 19-22,)	20%
Final Exam (TBA)	30%

NOTE #1: Assignments are due digitally (i.e. upload to Avenue) at 11:59pm on the due date.

NOTE #2: Late assignments will be deducted 0.01389% per minute. (20% /day)

Assignments (5% Each)

Based on material learned in lectures and tutorials

Programming in Matlab or Python

- If you don't have access to a computer to run these, let me know and we can get you access

Individual assignments to be done as an individual effort

Analysis Project (20%)

- project and presentation are done in groups (of up to 4)
- model some form of physiology or physiological response using 2 different methods
- can be disease and/or therapy related
- must choose how to model the data and why an approach is chosen (why appropriate)
- must include proper statistical analysis from multiple subjects
- must include error analysis where appropriate.
- The report should end with interpretations, conclusions, and possible future directions and should be 2000-3000 words (not including figure/table captions or programming code)
 - I.e. should look like a scientific journal
- figures, programming code, etc. must be included

Presentation (5%)

- At the conclusion of your project there will be a presentation day where groups will share their project in breakout rooms
- Grading will be done by the instructor or TA.
- Group self assessment can be done and submitted to the TA the day after the presentation (optional).
- Groups will be decided upon by the students and submitted to the TA by Tuesday September 17th. Whoever does not get into a group will be randomly assigned into one.
- Presentations will be graded according to: Style, Clarity, Knowledge, Quality, Organization and ability to answer questions. The [rubric](#) is on avenue

Analysis Project

Where to get Data??

- 1). You can acquire it yourself. BUT- please let me know asap if this is what you wish (you will need ethical approval)
- 2). Download data from a repository:

<https://physionet.org/physiobank/physiobank-intro.shtml>

https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html

<https://github.com/meagmohit/EEG-Datasets>

<https://www.re3data.org>

<https://lionbridge.ai/datasets/18-free-life-sciences-medical-datasets-for-machine-learning/>

<https://www.cancerimagingarchive.net/>

<https://www.kaggle.com/uciml/pima-indians-diabetes-database>

<https://www.kaggle.com/uciml/breast-cancer-wisconsin-data#data.csv>

<https://www.kaggle.com/tags/pharmaceutical-industry>

Due December 5th.

	Instructions: Fill out completely and email to the TA for each assignment and presentation		
	NOTE: This is an integral part of grading.		
	NOTE: This form is optional. Use it if you think a member(s) has not shared in doing the work		
A	YOUR Name:		
	Student #:		
	Assignment:		

B	Peer evaluation of contributions to presentation by my group members		
	Fill in the table below with a score indicating the RELATIVE CONTRIBUTION of each member in your group, including yourself. Total of all your scores must equal zero.		
	The relative contributions should be indicated on a scale of -10 to +10, where negative scores indicate the contribution was below average and positive scores indicate the contribution was above average.		
	Each assignment and presentation are worth 9% of your total grade. Each point counts as +/- 10% of this (up to a max of +/-100%). So if your group assignment grade is 8/9 and your group members give you an average of -4, your assignment grade is $(8/9)-40\% \times 9 = 4.4/9$		
	If you score anyone's contribution as < -2 or > +2, make sure you justify your score.		
	NOTE: YOU can't get higher than 9/9 final grade on an assignment!!!		
	Last name	First name	Relative Contribution
		total	0
C	Justification		

Exams

Midterm Exam: (October 19th -22nd) 20%

Final Exam: Date, TBA 30%

- covers from Midterm onwards

BOTH EXAMS are “OPEN BOOK”

EMERGENCIES ON CAMPUS

LOCKDOWN



Be familiar with Lockdown Procedures



EMERGENCY LOCKDOWN

- Exit the building if it is safe to do so
- If you are unable to leave:
 - Close, lock and barricade the door
 - Cover the door windows
 - Turn off the lights
 - Lay on the floor, under furniture if possible
 - DO NOT answer the door
 - Wait for Police or security to assist you out of the building



CAMPUS VIOLENCE

If you encounter a violent act:
(Shooter or hostage situation)

- **RUN:** if safe to do so, move quickly away from the area; OR
- **HIDE:** if not safe to move, hide as best you can, locking or barricading the doors; OR
- **FIGHT:** as a last resort, try to disrupt or incapacitate the perpetrator
- If hiding, stay hidden until an all clear is given

EMERGENCIES ON CAMPUS LOCKDOWN

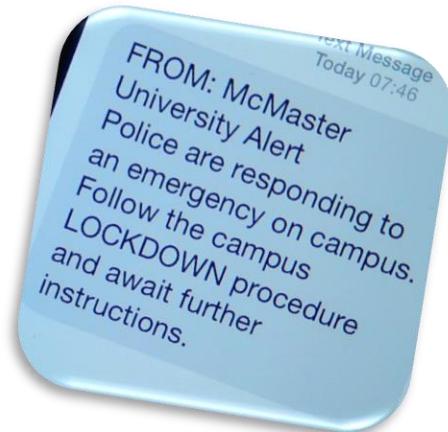
Be familiar with Lockdown Notifications



Listen for
Sirens



Register to receive text
messages



Read Building LCD
Screens



EMERGENCIES ON CAMPUS



The screenshots show the app's main screen with various emergency contact buttons and a detailed view of the 'On-Campus Emergency' section.

Main Screen:

- Emergency Contacts (red button)
- There are no current alerts.
- Emergency First Response Team (red button)
- Student Walk Home Attendant Team (grey button)

On-Campus Emergency Section:

For on-campus emergencies please use the button below to contact McMaster University's Campus Security

Dial McMaster Security

Off-Campus Emergency Section:

For off-campus emergencies please call 911 using the button below.

Dial 911

Other Features:

- Campus Safety ...
- Supporting Students In Difficulty
- Campus Violence
- Emergency SMS Notification Sign-up
- Lockdown Procedures

At the bottom of the main screen are navigation buttons: Home, Cab, Transit, More, and Share.

Download the MUSST Safety App

- Available in all app stores
- FREE
- Quick dialing for emergencies
- Quick emergency references

EMERGENCIES ON CAMPUS



MEDICAL

If someone requires medical assistance:

- Request Emergency First Response Team (EFRT) or an ambulance
- Describe the medical emergency to the dispatcher
- Keep the individual calm and provide information to responders



**For ALL
Emergencies on
Campus:
905-522-4135
Dial 88
from campus phones
or **The MUSST
safety App****



FIRE

- Pull the nearest fire alarm and exit the building
- Leave the building immediately
- **DO NOT** use elevators
- Keep all accesses clear
- **DO NOT** re-enter the building
- Await instructions from McMaster Security Services



Integrated Biomedical
Engineering & Health
Sciences Program

IBEHS - 4QZ3
Modelling of Biological Systems

Lecture 4

TAYLOR DEVET MASC.

PHD. CANDIDATE BIOLOGICAL AND BIOMEDICAL ENGINEERING

MCGILL UNIVERSITY

SHRINERS HOSPITAL FOR CHILDREN

Todays Aims...



Cardiac Modelling

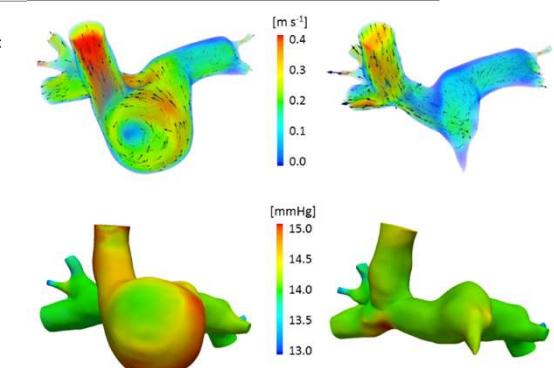
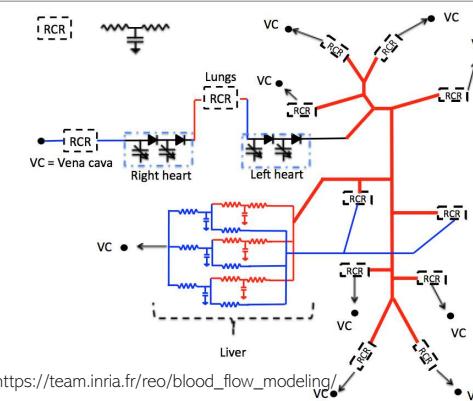
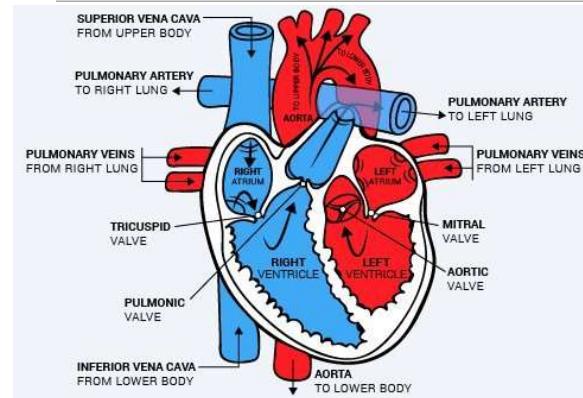


Macro Vasculature → Micro
Vasculature modeling



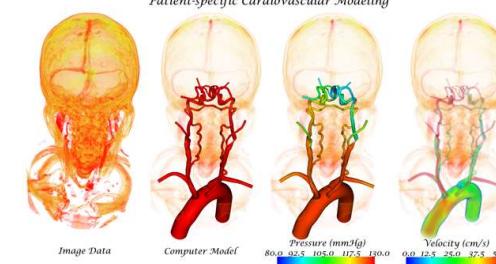
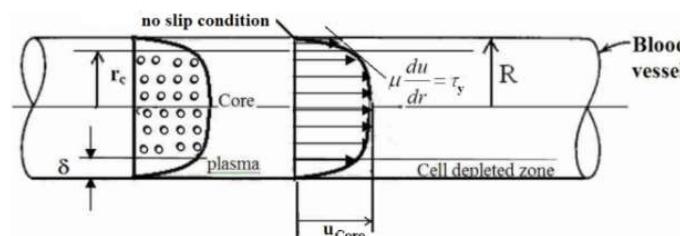
Intro to Pharmacokinetics

Modelling Blood Flow



https://team.inria.fr/reo/blood_flow_modeling/

<https://royalsocietypublishing.org/doi/10.1098/rsif.2018.0486>



<https://bloodflow.engin.umich.edu/gallery/patient-specific-blood-flow-simulation/>

Macro vasculature

Modelling Blood Flow

Stroke Volume

- $SV = L/\text{beat} = 70\text{cm}^3/\text{beat} = 0.070L/\text{beat}$

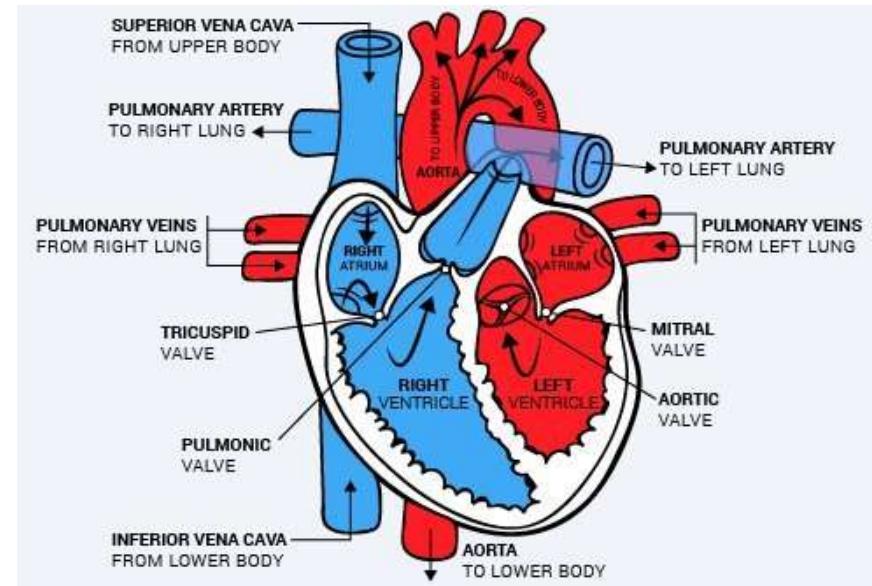
Heart Rate

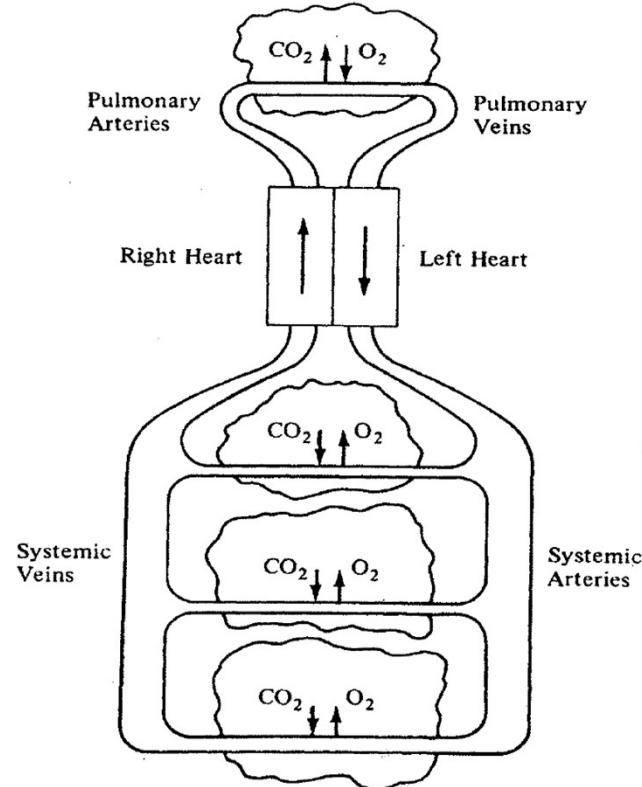
- $HR = \text{beats}/\text{min} = 80 \text{ beats}/\text{min}$

Cardiac Output

- $CO = \text{Cardiac Output (L/min)} = 5.6L/\text{min}$

$$CO = SV \times HR$$



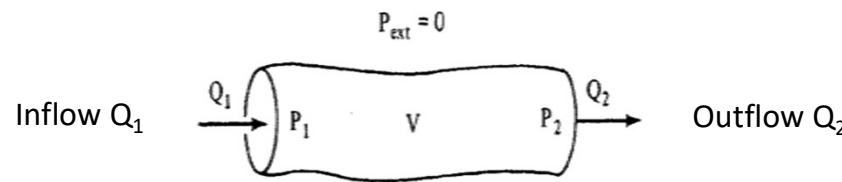


Normal resting blood pressures and volumes

	P(mmHg)	V(liters)
sa	100	1.0
sv	2	3.5
pa	15	0.1
pv	5	0.4

(s=systemic, p=pulmonary, a=artery, v=vein)

Resistance and Compliance Vessels



Blood vessels are characterized by 2 main characteristics

- 1) Resistance – How they resist the blood flowing though them
- 2) Compliance – how they deform due to changes in pressure

P_{ext} = external reference pressure,

At steady state $Q_1 = Q_2$. Is this true? Why?

How are Q , P_1 , P_2 , and V related?

Resistance Vessel

Smaller vessels in the microvasculature

Resists the flow of blood

Assumptions:

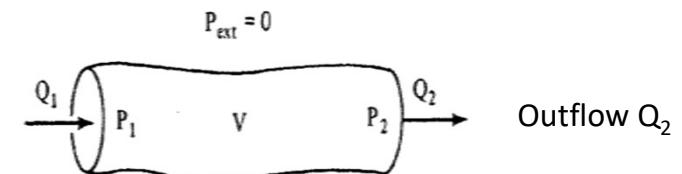
- V has to be constant
- only pressure differences matter

NOTE resistive only vessel are idealizations!

- real vessels need to exhibit compliance as well
- here everything is assumed as linear

$$Q = \frac{P_1 - P_2}{R} \quad I = \frac{V}{R}$$

I = Current in Amperes (A)
V = Voltage in Volts (V)
R = Resistance in Ohms (Ω)



Compliance Vessel

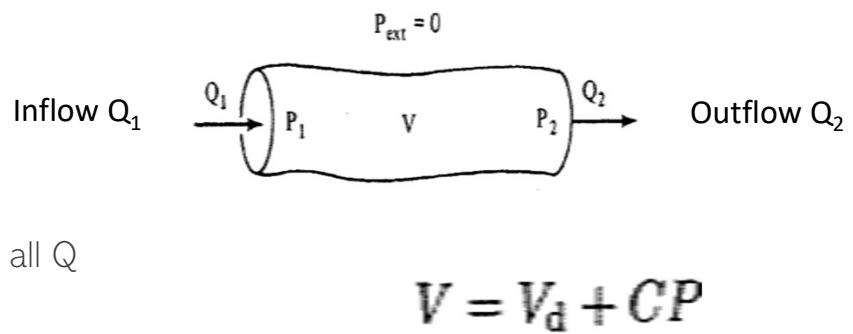
- Larger vessels
- Have resistance to pressure and store it

Assumptions

- assume zero resistance to blood flow
- pressures on both ends of vessel are equal for all Q
- V_d is the “dead volume” at $P=0$

NOTE compliance vessels are idealizations!

- real vessels need to exhibit both
- here everything is assumed as linear



However...

- large arteries and veins are “primarily” compliance vessels and changes in V are highly significant
- resistance is in the tissues
 - Primarily at the arteriole level.
 - Here volume changes are less important and large pressure drops are observed.
- Assumptions:
 - Linearity?
 - Steady state?

Vascular Networks

Arterioles

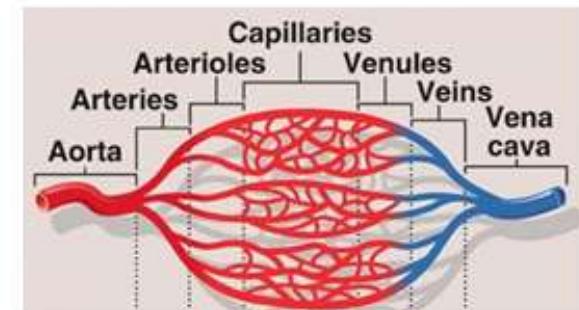
- Flow
 - 95% - 100%
- 25 μm diameter.
- <15% total vessel blood volume

Capillaries

- Flow
 - 80% - 90%
- 8 μm diameter
- 40% total vessel blood volume
- primary O₂ blood-tissue exchange site

Venules

- flow
 - 60% - 90%
- 25-50 μm diameter.
- 40% total vessel blood volume



■ ■ ■ **Transit Time = 2-3 s** ■ ■ ■

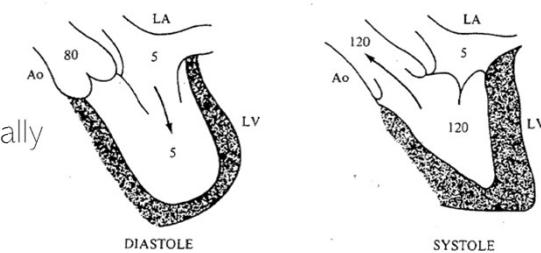
Source: Chris Thomas' Slides

The Heart as a Pair of Pumps

- equipped with an inflow (mitral) valve and an outflow (aortic) valve.

- Diastole

- ventricle is relaxed
- the inflow valve is open and the outflow valve is closed
- left ventricle receives blood from the left atrium at a pressure that is essentially that of the pulmonary veins



- Systole

- ventricle contracts the inflow valve closes and outflow valve opens.
- left ventricle actively pumps blood into the systemic arterial tree.
- left ventricle pressure systemic arterial pressure

$$V(t) = V_d + C(t)P(t)$$

The Heart as a Pair of Pumps

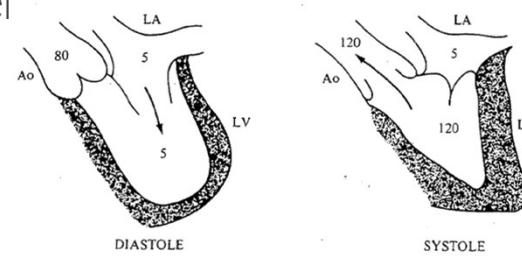
- What determines left ventricle output?

- consider ventricle is a compliance vessel

- $C(t)$ is compliance at time= t

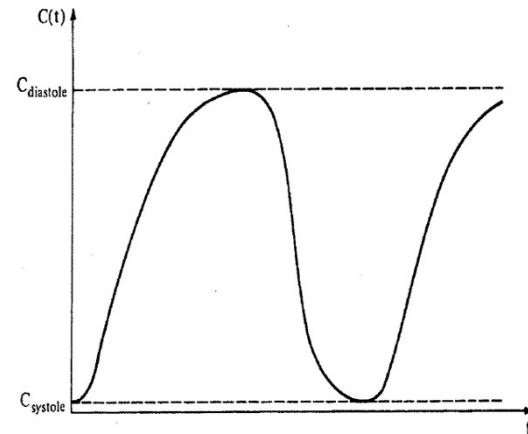
- Systole ? $C(t) \uparrow \quad \downarrow$

- Diastole? $C(t) \uparrow \quad \downarrow$



$$V(t) = V_d + C(t)P(t)$$

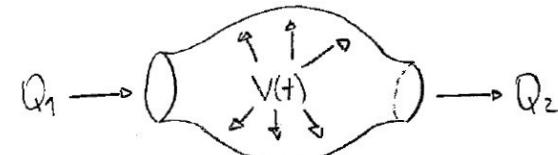
Compliance vs time



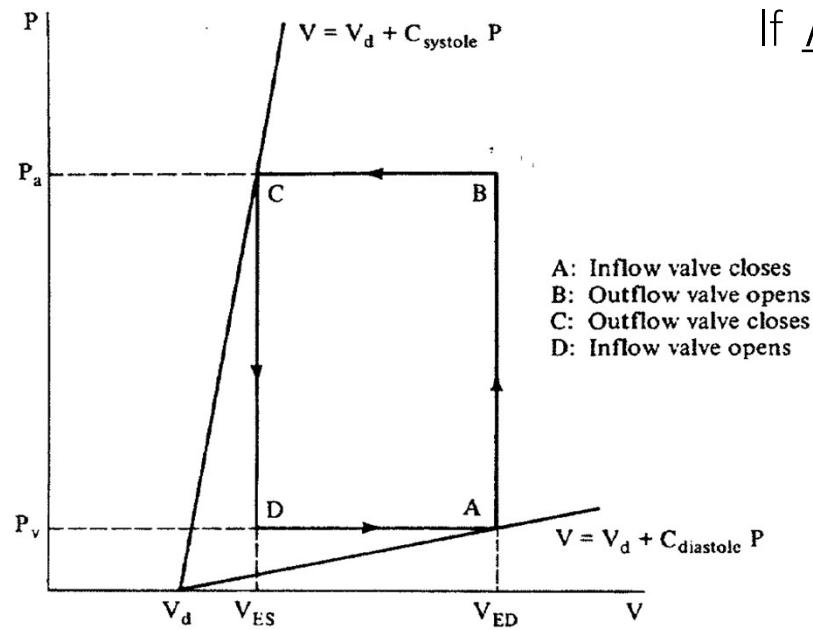
Maximal volumes at end diastole (ED) and end systole (ES):

Therefore Stroke Volume (SV) is:

P_a is pressure in the arteries supplied by the ventricle, P_v is pressure in the veins that fill it.



Ventricle Pressure-volume diagram



If $F = \text{heart rate}$ (frequency, in beats/min):

$$Q = FV_{\text{stroke}} = FC_{\text{diastole}} P_v$$

Ventricular Efficiency (K)

$$K = FC_{diastole}$$

F(HR) is the same for the two sides of the heart

Constant C_{diastole} is greater for right ventricle than in the left ventricle

- hence K is _____ on the right than on the left.
- Also, the two sides of the heart are connected to different venous systems.

Therefore Right and Left Cardiac Outputs are:

sv = systemic venous

pv = pulmonary venous

Mathematical Model of the Uncontrolled Circulation

1. to study the self-regulating properties of the circulation, independent of external control mechanisms
2. to explain the need for external control mechanisms;
3. to serve as a foundation on which we can construct a simple model of circulation control

Recall

Q = flow (mL/min)

F = heart rate (beats/min)

P = Pressure (mm Hg)

P = Pressure (mm Hg)

V = Volume (L)

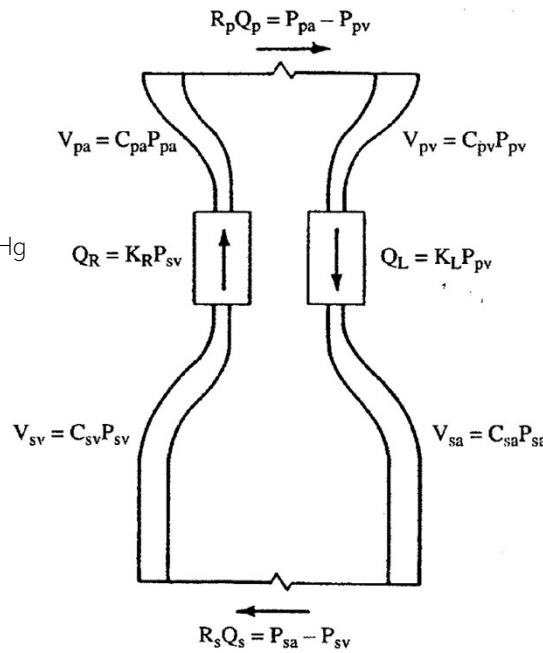
K = efficiency in (L/min)/mm Hg

C = Compliance (L/mm)

Put it all together...

Recall

Q = flow (mL/min)
 P = Pressure (mm Hg)
 V = Volume (L)
 C = Compliance (L/mm)
 F = heart rate (beats/min)
 P = Pressure (mm Hg)
 K = efficiency in (L/min)/mm Hg



$$V_{sa} = C_{sa} P_{sa}$$

$$V_{sv} = C_{sv} P_{sv}$$

$$V_{pa} = C_{pa} P_{pa}$$

$$V_{pv} = C_{pv} P_{pv}$$

$$\text{Recall } V = V_d + CP \text{ (ignore } V_d)$$

assume that the systemic and pulmonary tissues act like resistance vessels, so that:

$$Q_s = \frac{1}{R_s} (P_{sa} - P_{sv}),$$

$$Q_p = \frac{1}{R_p} (P_{pa} - P_{pv}).$$

Solving the Variables

Now we have an equation for each element of the circulation, with 12 unknowns:

$$Q_R, Q_L, Q_s, Q_p; P_{sa}, P_{sv}, P_{pa}, P_{pv}; V_{sa}, V_{sv}, V_{pa}, V_{pv}.$$

Can calculate total blood volume:

assume that the circulation is in a steady state, so that the flow into each of the compliance vessels must equal the flow out.

→ Thus:

Therefore now we have 9 equations and 9 unknowns:

$$Q, P_{sa}, P_{sv}, P_{pa}, P_{pv}, V_{sa}, V_{sv}, V_{pa}, V_{pv}$$

Solving the Equations

1. From the pump equations, we get the venous pressures in terms of Q :

Substituting this result in the resistance equations, we get the arterial pressures in terms of Q :

Combining equations...

Substituting all four pressures into the compliance equations, we obtain:

$$V_{sa} = C_{sa} P_{sa}$$

$$V_{sv} = C_{sv} P_{sv}$$

$$V_{pa} = C_{pa} P_{pa}$$

$$V_{pv} = C_{pv} P_{pv}$$

$$V_{sv} = \frac{C_{sv}}{K_R} Q,$$

$$V_{pv} = \frac{C_{pv}}{K_L} Q,$$

$$V_{sa} = \left[\frac{C_{sa}}{K_R} + C_{sa} R_s \right] Q,$$

$$V_{pa} = \left[\frac{C_{pa}}{K_L} + C_{pa} R_p \right] Q$$

$$V_i = T_i Q,$$

$$i = sv, pv, sa, pa$$

To save writing, we introduce the following combinations of parameters

$$T_{sv} = C_{sv}/K_R,$$

$$T_{pv} = C_{pv}/K_L,$$

$$T_{sa} = (C_{sa}/K_R) + C_{sa} R_s,$$

$$T_{pa} = (C_{pa}/K_L) + C_{pa} R_p$$

Combining equations...

$$(T_{\text{sa}} + T_{\text{sv}} + T_{\text{pa}} + T_{\text{pv}})Q = V_0,$$

substitute previous in to equations for the total blood volume and solve for Q:

$$Q = \frac{V_0}{(T_{\text{sa}} + T_{\text{sv}} + T_{\text{pa}} + T_{\text{pv}})}.$$

$$V_i = \frac{T_i V_0}{(T_{\text{sa}} + T_{\text{sv}} + T_{\text{pa}} + T_{\text{pv}})},$$

$$P_i = \frac{T_i V_0}{C_i(T_{\text{sa}} + T_{\text{sv}} + T_{\text{pa}} + T_{\text{pv}})}$$

Therefore, now have a formula for each unknown. Just need some normal starting values!

Normal Resting Values:

	Systemic	Pulmonary
R :	$R_s = 17.5$	$R_p = 1.79 \text{ mmHg}/(\text{liter}/\text{min})$
C :	$C_{sa} = 0.01$	$C_{pa} = 0.00667 \text{ liters/mmHg}$
	$C_{sv} = 1.75$	$C_{pv} = 0.08 \text{ liters/mmHg}$
	Right	Left
K :	$K_R = 2.8$	$K_L = 1.12 \text{ (liters/min)}/\text{mmHg}$
V :	$V_0 = 5.0 \text{ liters}$	

Balancing the Two Sides of the Heart

- Essentially have 2 separate systems that are joined together at the heart
- How are the two sides of the heart and the two circulations coordinated?

- What keeps the outputs of the right and left hearts equal?

- What mechanisms control the partition of blood volume between the systemic and pulmonary circulations?

Example

Assume : K_R Decreases

- Temporarily, $QR < QL$
- This will increase P_{sv} and lower the P_{pv} .
- The overall effect of pressure changes will be to drive the cardiac outputs back toward equality.
- A net rate of transfer of volume will persist until equality of output [of both sides] is restored.
- This results in a new equilibrium with a different partition of the blood volume

In the steady-state model, we only need to compute the end result of this process.

Recall:

$$T_{sv} = C_{sv}/K_R$$

$$T_{pv} = C_{pv}/K_L$$

$$T_{sa} = (C_{sa}/K_R) + C_{sa}R_s$$

$$T_{pa} = (C_{pa}/K_L) + C_{pa}R_p$$

Recall

Q = flow (mL/min)

P = Pressure (mm Hg)

V = Volume (L)

C = Compliance (L/mm)

F = heart rate (beats/min)

P = Pressure (mm Hg)

K = efficiency in (L/min)/mm Hg

$$\frac{V_p}{V_s} = \frac{V_{pa} + V_{pv}}{V_{sa} + V_{sv}} = \frac{T_{pa} + T_{pv}}{T_{sa} + T_{sv}}$$

$$= \frac{\left(\frac{C_{pa} + C_{pv}}{K_L} + C_{pa}R_p \right)}{\left(\frac{C_{sa} + C_{sv}}{K_R} + C_{sa}R_s \right)}$$

The key to the success of this intrinsic control mechanism is the dependence of cardiac output on venous pressure

External Control Mechanisms

- 1) Arterioles dilate.
 - Result: decreased systemic resistance R_s .
 - Concurrently cardiac output rises
 - Systemic arterial pressure (P_{sa}) is maintained.
- 2) The ↑ in cardiac output comes primarily from an ↑ in heart rate while stroke volume remains fairly constant.

Uncontrolled circulation flaws

- uncontrolled model response to change in R_s different response from the observed response
- In the uncontrolled circulation a decrease in R_s results in only a modest increase in cardiac output.
- The most noticeable effect is a substantial decrease in systemic arterial pressure (P_{sa}).
- This shows the need for the external circulatory control mechanisms

Model response to control mechanisms

What happens when we change in R_s ?

Only appears in one equation... but affects the rest

- *This is a system of equations*

What happens to Q and P_{sa} when R_s is reduced to 50% of its normal value (while leaving the other parameters unchanged) ?

$$Q = \frac{V_0}{T_{sa} + T_{sv} + T_{pa} + T_{pv}} \quad P_{sa} = \frac{V_0}{C_{sa}} \cdot \frac{T_{sa}}{T_{sa} + T_{sv} + T_{pa} + T_{pv}}$$

In the model reduce: R_s 50%:

<u>Normal</u>	$R_s = 50\%R_s^{normal}$
Q 5.6	6.2
P _{sa} 100.0	57

Effect of changing systemic resistance (R_s) on cardiac output (Q) and systemic arterial pressure (P_{sa}) in an uncontrolled circulation model.

Assess changes via sensitivity analysis:

If Y depends on X, and X changes, then the sensitivity of Y to X is defined to be:

$$\begin{aligned}\sigma_{YX} &= \frac{\Delta \log Y}{\Delta \log X} = \frac{\log Y' - \log Y}{\log X' - \log X} \\ &= \frac{\log(Y'/Y)}{\log(X'/X)}\end{aligned}$$

where $X' = X + \Delta X$ and Y' is the value that Y takes on when X is changed to X' .

From previous table we find:

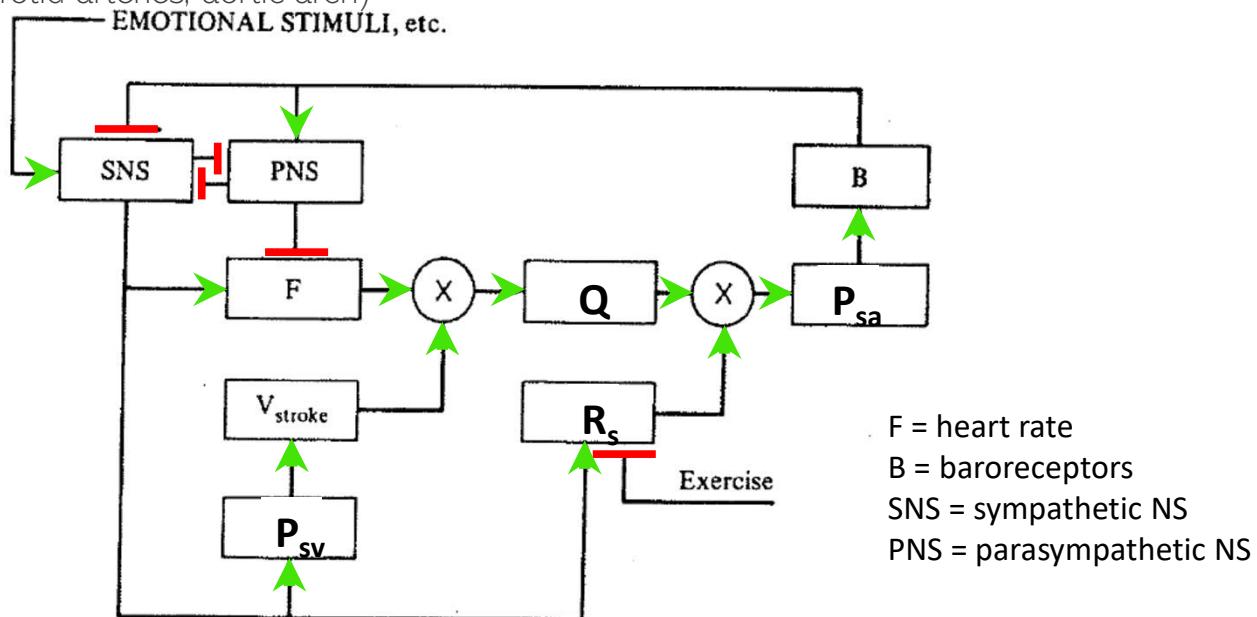
Conclusions about Rs

Therefore the sensitivity of Q to Rs and P to Rs are directly linked

- Any mechanism that accomplishes one will automatically accomplish the other.
- There has to be other factors at work

Physiological Control

- Baroreceptors (carotid arteries, aortic arch)



Baroreceptor Notes

- 1) The PNS is excited by activity of the baroreceptors → slows heart rate (F).
- 2) The SNS is inhibited by baroreceptor activity. This has several effects on circulation, including:
 - (a) increased heart rate
 - (b) increased venous pressure, and so increased stroke volume
 - (c) increased systemic resistance

The loop is closed through the mechanics of the circulation:

$$Q = F * V_{stroke} \text{ and } P_{sa} = Q * R_s$$

Baroreceptors adjust heart rate until systemic arterial pressure hits a target value

Introducing Heart Rate (F)

Recall:

$$Q_R = K_R P_{sv}$$

$$Q_L = K_L P_{pv}$$



$$Q_R = FC_R P_{sv}$$

$$Q_L = FC_L P_{pv}$$

Now this is a model of the controlled circulation where the above 2 equations can be included with the steady state relationship:

$Q_R = Q_P = Q_S = Q_L$ and:

$$V_{sa} = C_{sa} P_{sa}$$

$$V_{sv} = C_{sv} P_{sv}$$

$$V_{pa} = C_{pa} P_{pa}$$

$$V_{pv} = C_{pv} P_{pv}$$

$$Q_s = \frac{1}{R_s} (P_{sa} - P_{sv}),$$

$$Q_p = \frac{1}{R_p} (P_{pa} - P_{pv}).$$

How to Solve: make some approximations!

- 1) Neglect P_{sv} compared to P_{sa} in the equation of the systemic resistance to simplify

$$QR_S = P^*$$

- 2) neglect pulmonary volumes in comparison with the systemic volumes in the equation of the total blood volume.

$$V_{sa} + V_{sv} = V_0 \quad \text{Or:} \quad C_{sa}P^* + C_{sv}P_{sv} = V_0$$

Use above to determine Q and P_{sv} :

Simplified Model

$$Q = \frac{P^*}{R_s}$$

$$P_{sv} = \frac{V_0 - C_{sa} P^*}{C_{sv}}$$

Substitute into →

$$Q_R = F C_R P_{sv}$$

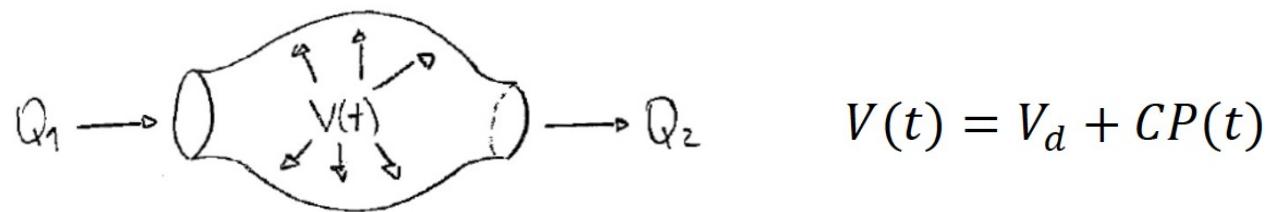
Resultant is solution for heart rate:

$$F = \frac{P^* C_{sv}}{R_s C_R (V_0 - C_{sa} P^*)}$$

Changes in R_{sv} ?

Increase in Q_R ?

Let's Consider this as a dynamic system



Compliant Vessel expanding due to blood flow

$$\frac{dV}{dt} = C \frac{dP}{dt} \quad \frac{dV}{dt} = Q_1 - Q_2$$

$$C \frac{dP}{dt} = Q_1 - Q_2 \quad (\text{Eqn.1})$$

1st dynamic cardiovascular model

- o monitor changes in blood pressure over time in a systemic artery.
- o assume artery is a compliant vessel with inflow of blood Q1
- o blood outflow (Q2) flows out to the microcirculation
- o assume arterioles and capillaries are lumped together as 'resistance vessels'

$$C_{SA} \frac{dP_{SA}}{dt} = Q_{Ao} - Q_S \quad (\text{Eqn.2})$$

$$Q_S = \frac{P_{SA} - P_{SV}}{R_S} \quad (\text{Eqn.3})$$

Dynamic model cont

$$Q_s = \frac{P_{SA} - P_{SV}}{R_s} \quad (\text{Eqn.3}) \quad \longrightarrow \quad Q_s = \frac{P_{SA}}{R_S} \quad (\text{Eqn.4})$$

Putting Eqn.4 and 2 together:

$$C_{SA} \frac{dP_{SA}}{dt} = Q_{Ao} - \frac{P_{SA}}{R_S}$$

$$\frac{dP_{SA}}{dt} = \frac{1}{C_{SA}} \left(Q_{Ao} - \frac{P_{SA}}{R_S} \right) \quad (\text{Eqn.5})$$

Dynamic model cont

- Now, the outflow from the heart into the systemic artery = $Q_{AO}(t)$
- BUT.... outflow is pulsatile!!
- Simplify as a triangle

Q_{max} = maximum blood flow)

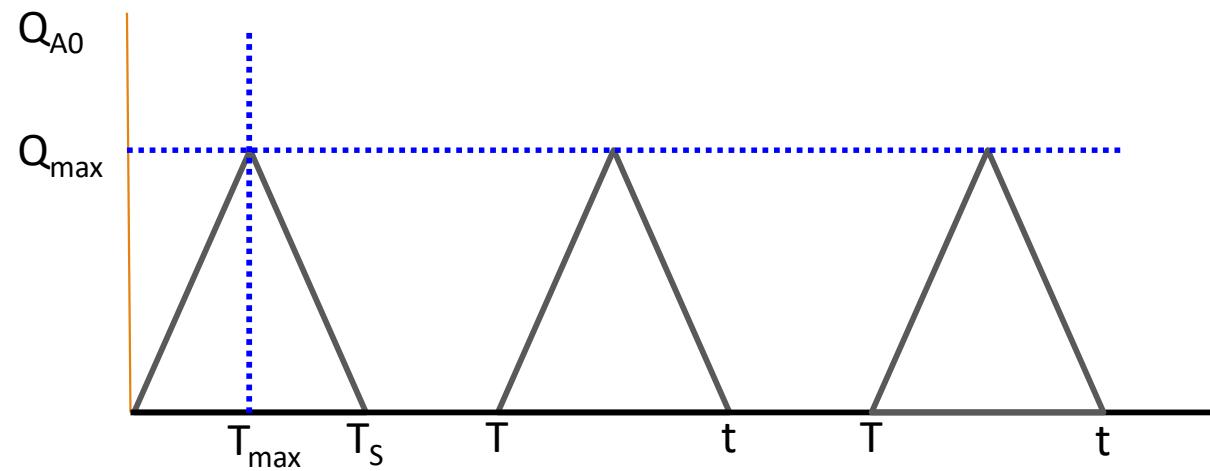
T_{max} = time of max. blood flow (i.e. corresponds with Q_{max} , the waveform peak)

T_s = duration of systole (corresponds to length of base of triangle)

T = duration of one heart beat (period of signal)

(Eqn.6)

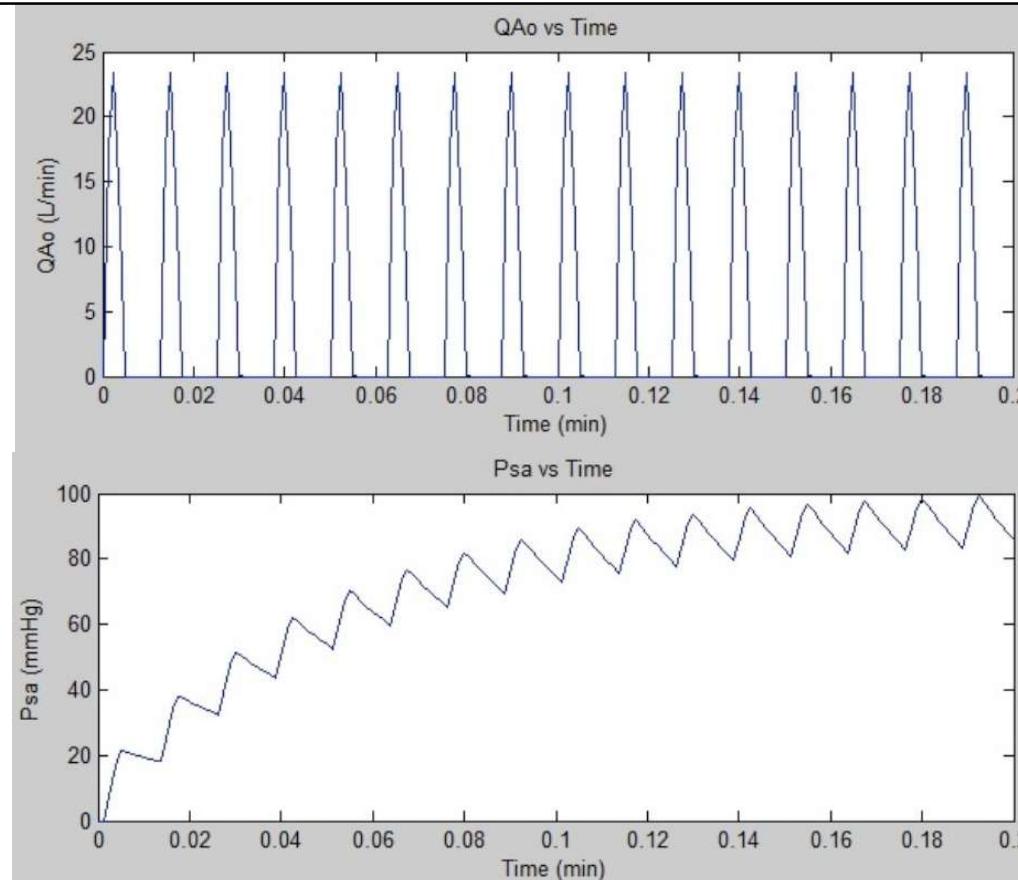
$$Q_{AO}(t) = \begin{cases} Q_{max}t/T_{max} & 0 \leq t \leq T_{max} \\ Q_{max}(T_S - t)/(T_S - T_{Max}) & T_{max} \leq t \leq T_S \\ 0 & T_S \leq t \leq T \end{cases}$$



In starting any model we need to give the model initial conditions:

```
T = 0.0125; % Duration of heartbeat: min
Ts = 0.0050; % Duration of systole: min
Stroke_volume = 70e-3; % Volume ejected by 1 heart beat: L
Tmax = 0.0020; % Time at which flow is max: min
Qmax= Stroke_volume/(0.5*Ts); % Max flow through aortic valve: L/min
dt = .01*T; % i.e. 100 time pts per cardiac cycle
% Compliance and resistance values for arteries
Csa = .00175; % Systemic arterial compliance: L/mmHg
Rs = 17.86; % Systemic resistance: mmHg/(L/min)

% To begin with, QA0 and Psa are zero
% will model 16 heart beats, with a step size of 75ms
```



Done with a fixed step ode1 (Euler) solver using Matlab.

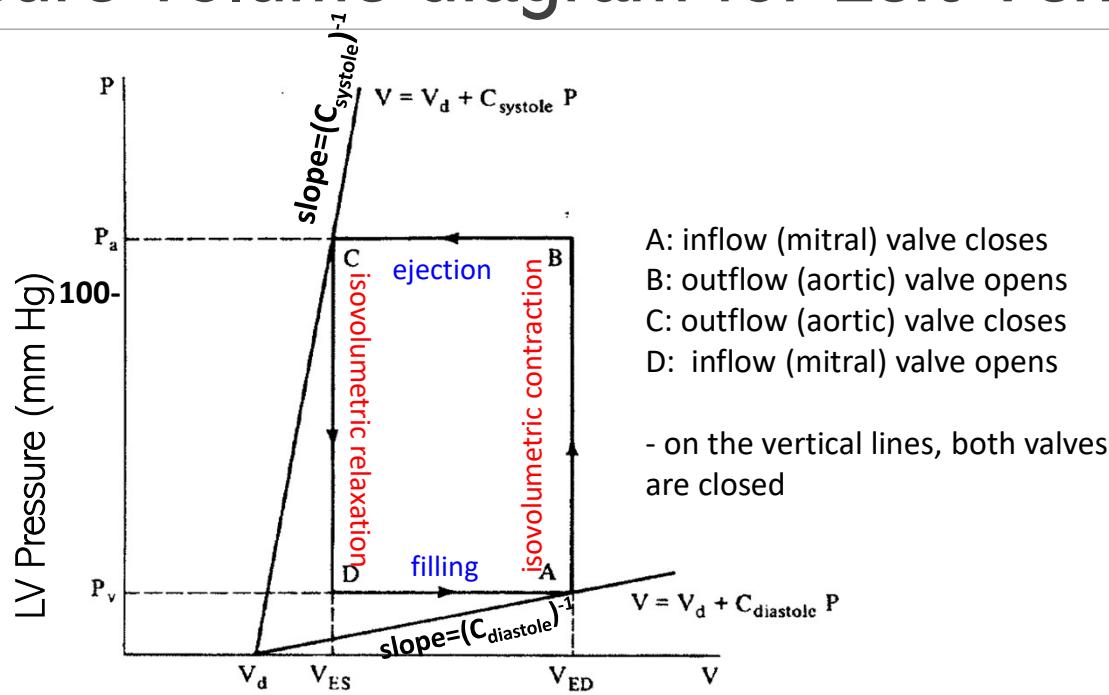
NOTES:

- 1) Initial condition of the blood pressure was set to 0 mmHg.
- 2) after several simulated heart beats the model settles to a stable solution where Psa (systolic pressure) achieves ~100 mmHg with a diastolic pressure of 80 mmHg.
- 3) Increasing pressure would correspond to increase in ejection volume
- 4) It does not reach typical 120/80 mmHg due to initial conditions for arterial compliance and systemic resistance
 - These adjust pulse pressure and mean arterial pulse

What about the Left Ventricle?

- Requires modeling of cardiac contraction and valves opening/closing of the valves.
- Thus, expand model to include left ventricle (LV)
 - a basic pump
 - accepting fluids at low pressure (P_1)
 - Transfers to a region where pressure is higher ($P_2 > P_1$).
 - muscular contractions will change LV wall compliance

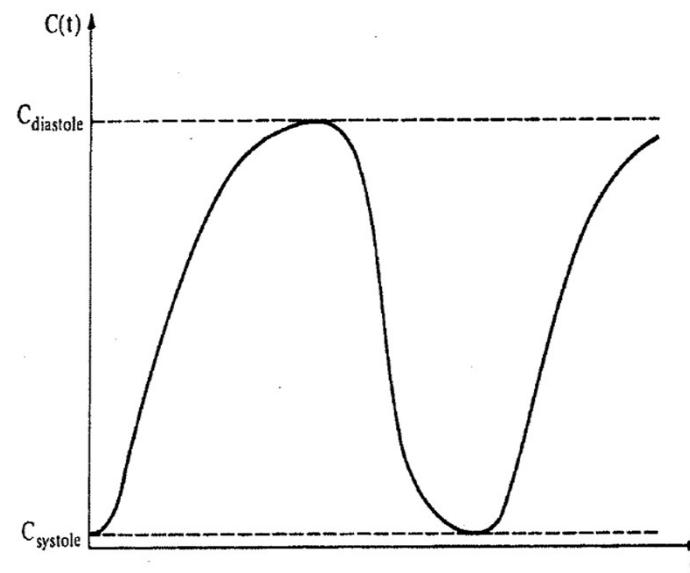
Pressure-volume diagram for Left Ventricle

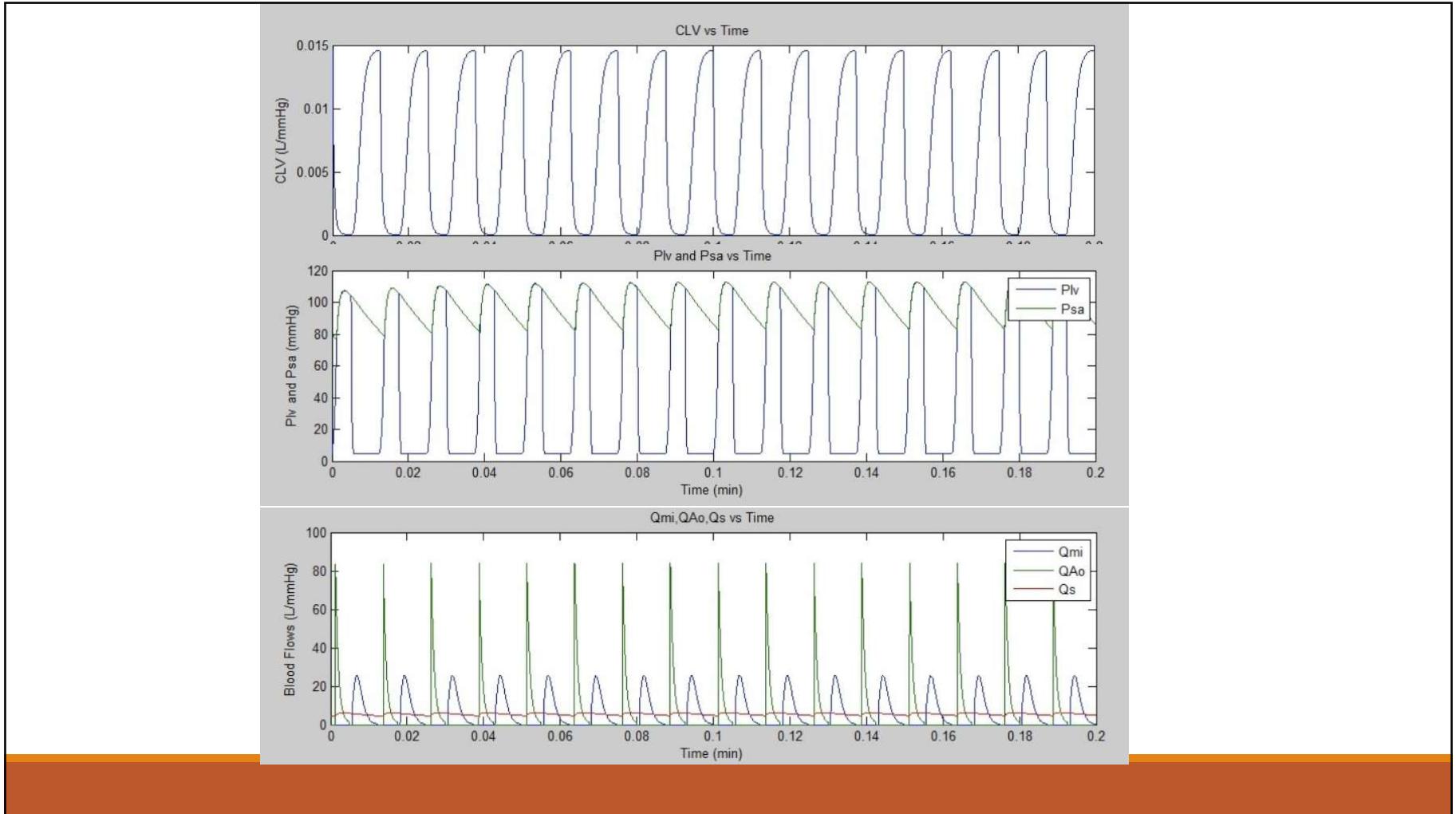


Compliance over time

LV can now be treated similarly to the compliant vessels (model above). However, compliance is no longer static.

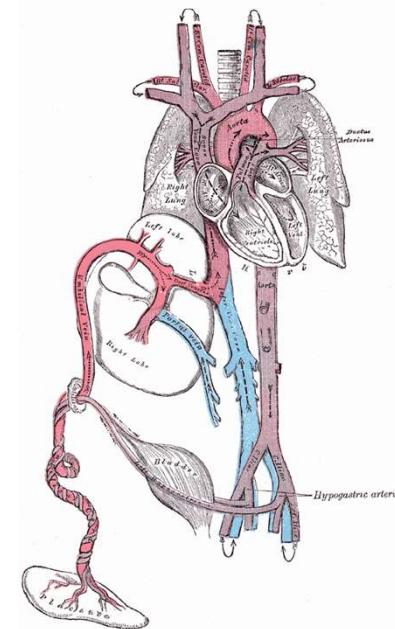
- During diastole, \uparrow compliance, accommodating \uparrow blood volumes.
- During systole, \downarrow compliance to be able to contract ejecting blood with higher pressure.





Fetal Circulation

- 1) Ductus Venosus- shunt placental blood to vena cava
- 2) Foramen Ovale- hole between atria to short circuit blood that would go to lungs
- 3) Ductus Arteriosus – portal between aorta and pulmonary artery.



Fetal Circulation

At birth, when the infant breathes for the first time the following happens:

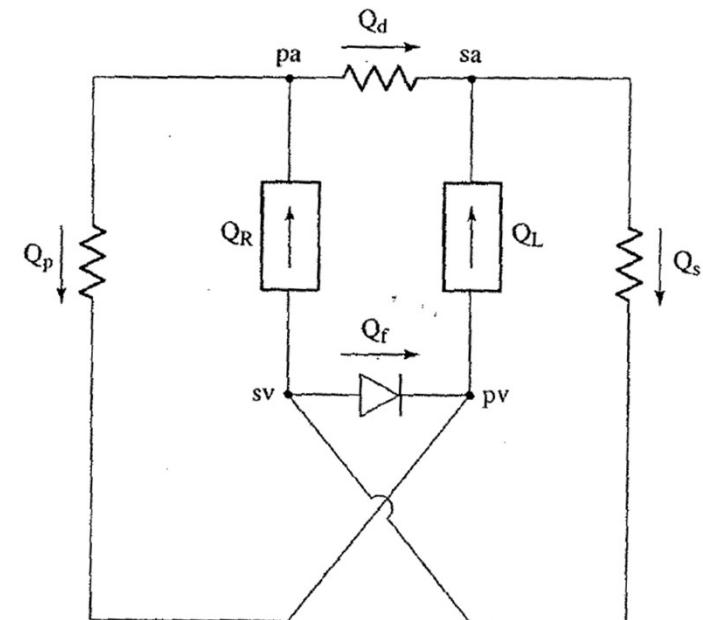
- 1) There is a decrease in the resistance in the pulmonary vasculature
- 2) This causes the pressure in the left atrium to increase relative to the pressure in the right atrium.
- 3) This leads to closure of the foramen ovale
- 4) The newly increased rise in blood oxygen leads to a decrease in prostaglandins, causing closure of the ductus arteriosus.

These closures prevent blood from bypassing pulmonary circulation, and therefore allow the neonate's blood to become oxygenated in the newly operational lungs.

Fetal Circulation

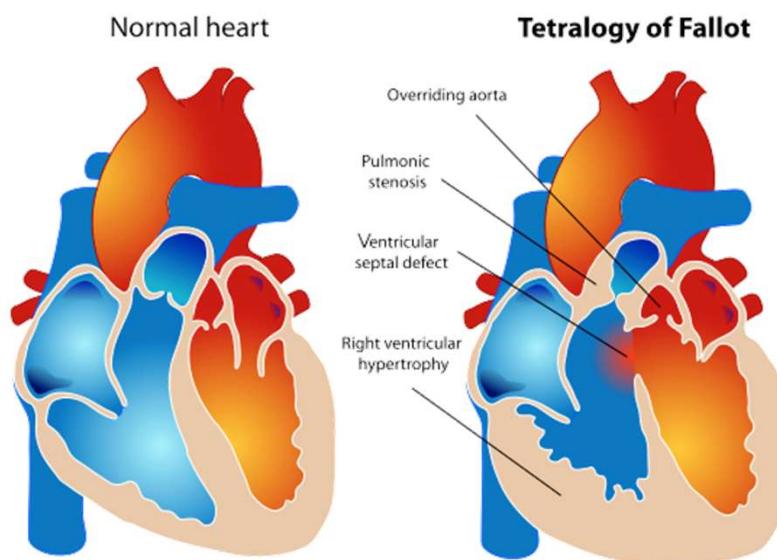
For the fetal circulation there are 6 flows to consider:

- QL (left heart output)
- QR (right heart output)
- Qp (pulmonary flow=small)
- Qs (systemic flow=large since includes placenta)
- Qd (flow through ductus arteriosus)
- Qf (flow through foramen ovale)



Tetralogy of Fallot

How is this modelled?



https://en.wikipedia.org/wiki/Tetralogy_of_Fallot

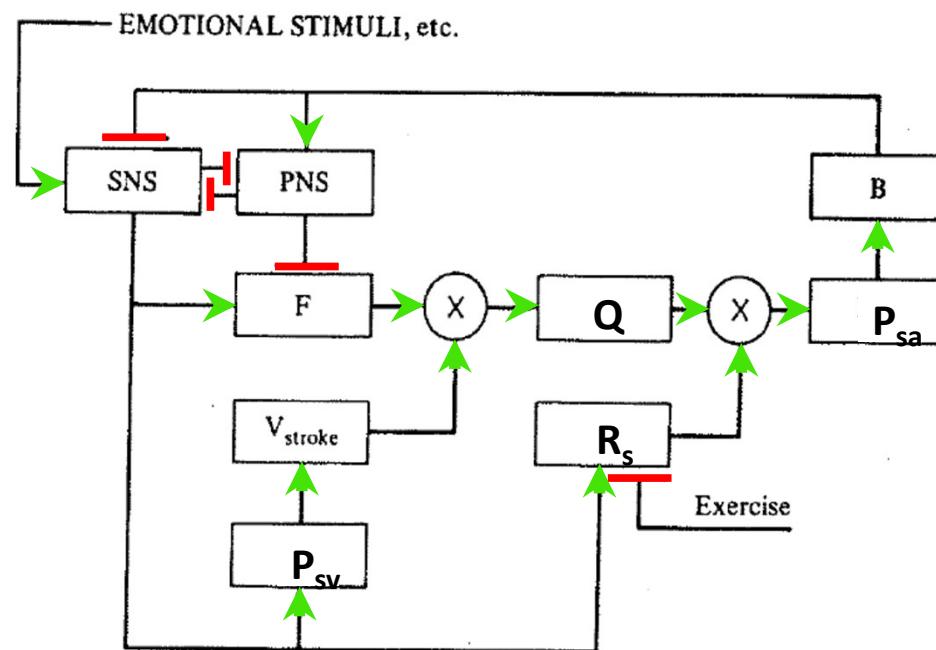
Engineering vs Physiological Control Systems

- 1) An Engineering control system is designed to accomplish a defined task, and frequently, the governing parameters have been fine-tuned extensively so that the system will perform its task in an "optimal" manner
 - physiological control systems are built for versatility and may be capable of serving several different functions (e.g. respiratory system provides gas exchange, and secondarily helps facilitate heat dissipation)
- 2) As engineering control systems are synthesized by a designer, the characteristics of the various components are generally known.
 - a physiological control system usually consists of components that are unknown and difficult to analyze.
 - constant need to apply system identification techniques to determine how these various subsystems behave before we are able to proceed to analyzing the overall control system

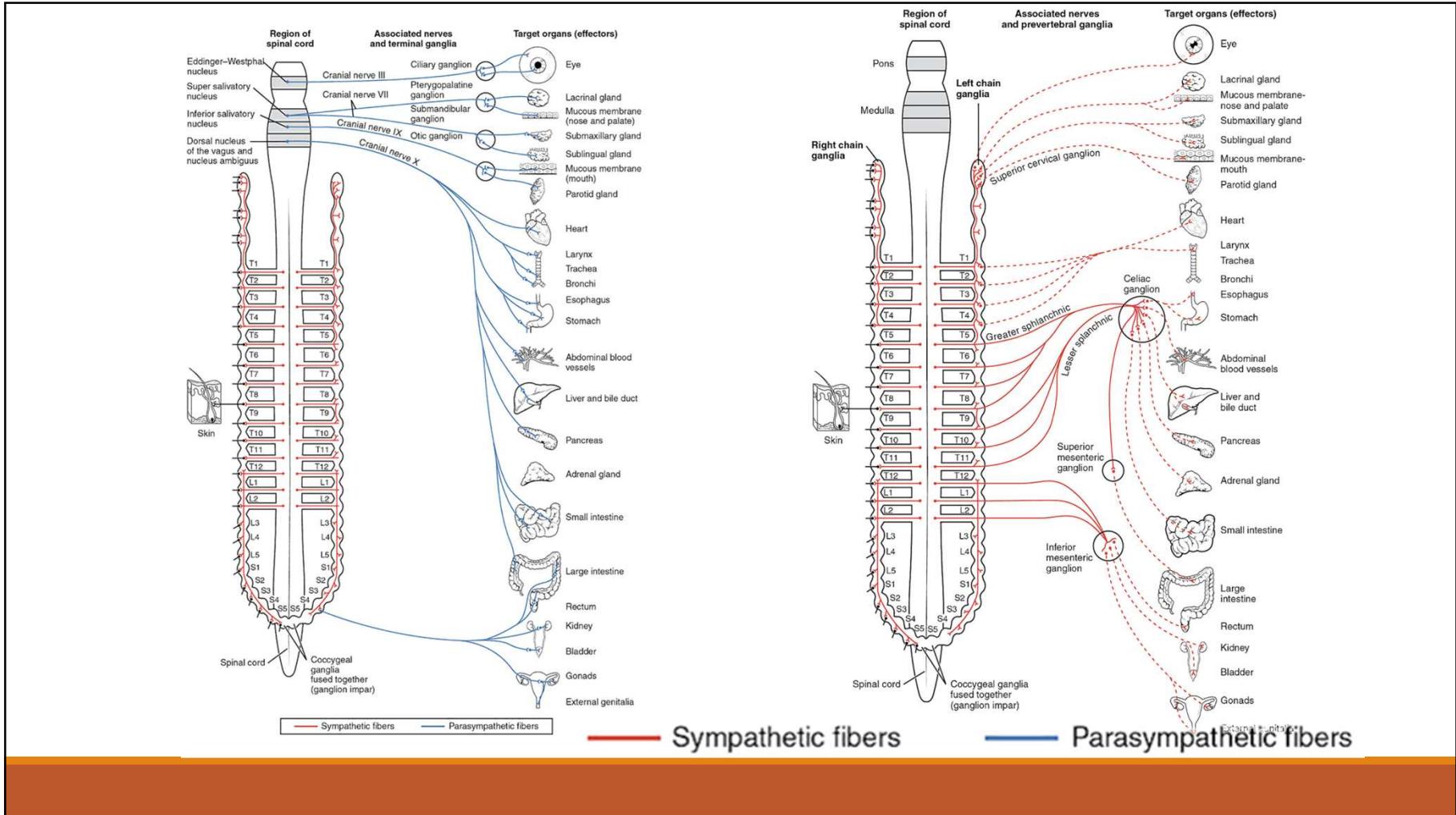
Engineering vs Physiological Control Systems

- 3) Occasionally physiological control systems are non-optimal, or poorly designed!! However, both engineering and physiological control systems break down. Both can have simple failures and catastrophic failures.
- 4) There is an extensive degree of cross-coupling or interaction among different physiological control systems.
e.g. the healthy cardiovascular system is largely dependent on interactions with respiratory, renal, endocrine, and other organ systems.
- 5) Physiological control systems, in general, are adaptive. This means that the system may be able to offset any change in output not only through feedback but also by allowing system characteristics to change.
- 6) physiological systems are generally nonlinear, while engineering control systems can be linear or nonlinear.
 - The engineering designer prefers the use of linear system components since they have properties that are well-behaved and easy to predict.

Consider Balance of PNS and SNS regulation



F = heart rate
B = baroreceptors
SNS = sympathetic NS
PNS = parasympathetic NS



System Complexity

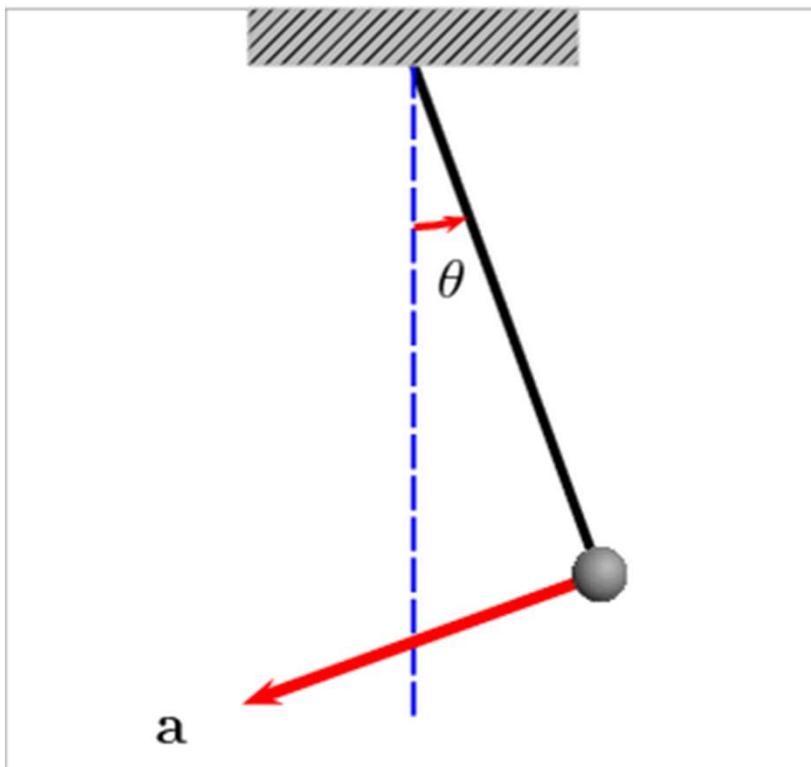
Consider each regulatory system is like a pendulum that influences cardiac contractility

The "simple pendulum" is an isolated system having the following assumptions:

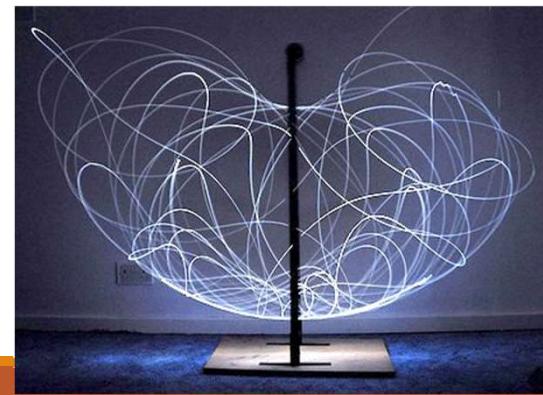
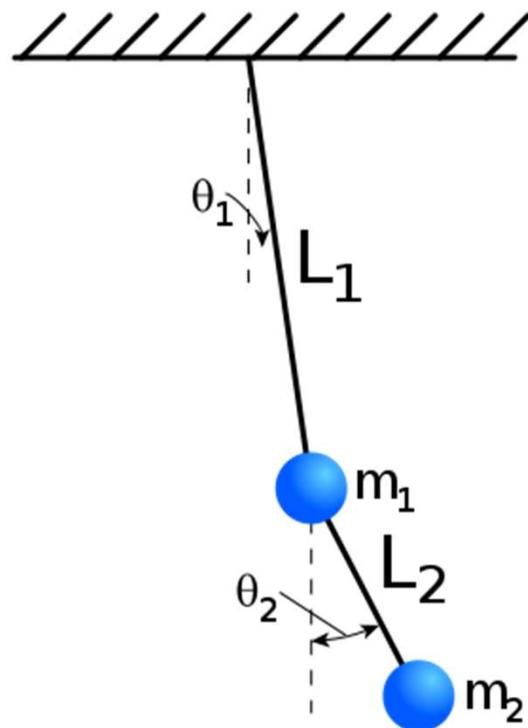
- 1) The cord is massless, inextensible and always taut
- 2) The hanging bob is a point mass
- 3) Motion occurs only in two dimensions (i.e. along an arc).
- 4) The motion does not lose energy to friction or air resistance.

The differential equation which represents the motion of a simple pendulum is

$$\frac{d^2\theta}{dt^2} + \frac{g}{\ell} \sin \theta = 0$$



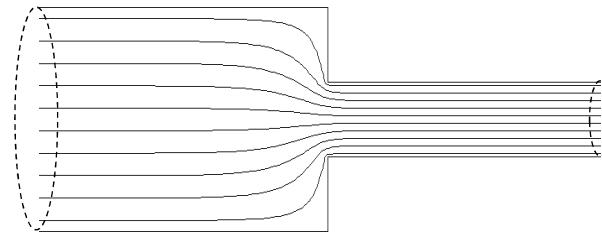
$$\frac{d^2\theta}{dt^2} + \frac{g}{\ell} \sin \theta = 0$$



Micro Vasculation

Laminar Fluid Flow

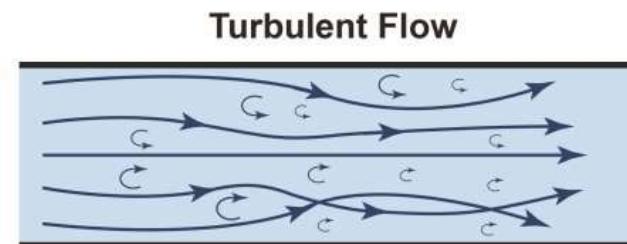
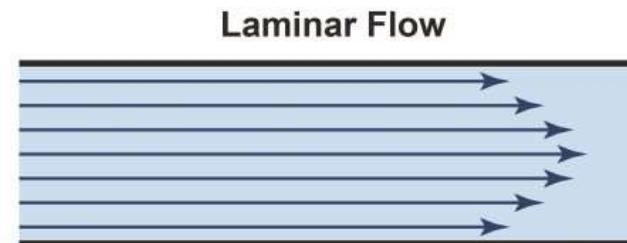
- the smooth motion of adjacent fluid layers sliding continuously past each other without breaking into whirlpools or vortices.
- Also known as streamline flow
- Visualized by lines of constant fluid speed (streamlines).
- Laminar streamlines never cross
- Steady state fluid flow is a special case of laminar flow where fluid velocity is everywhere time independent.



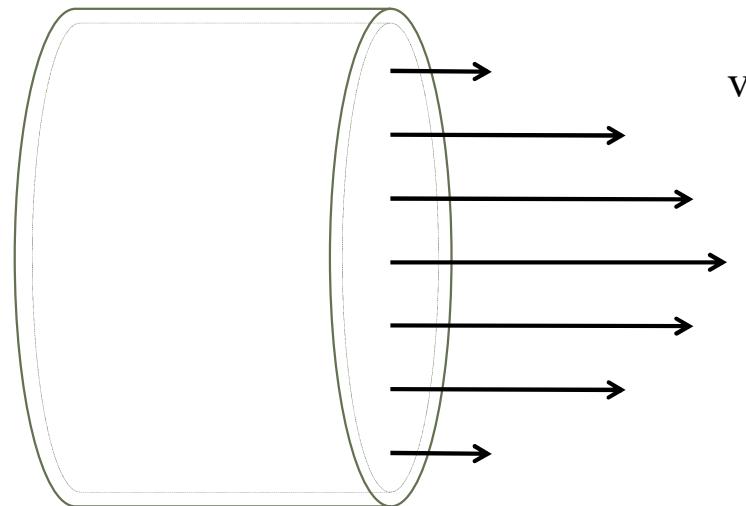
Turbulent Fluid Flow

Fluid experiences mixing or fluctuations in fluid flow

Changes in speed or direction of flow



<https://www.automation.com/en-us/articles/2018/demystifying-fluid-turbulence-velocity-and-flow-me>



Fluid velocity profile in a blood vessel. Blood flows faster at the center of the vessel than near the walls of the vessel.

Reynolds Number

Reynolds number is a useful measure of the fluid flow proportional to the ratio of inertial to viscous forces.

$$\text{Re} = \frac{\text{inertial forces}}{\text{viscous forces}} = \frac{\rho v a}{\eta}$$

ρ = density of the fluid (kg/m^3)

v = velocity (m/s)

a = length

η = fluid viscosity ($\text{Pa}\cdot\text{s}$ or $\text{N}\cdot\text{s}/\text{m}^2$ or $\text{kg}/(\text{m}\cdot\text{s})$)

Reynolds Number

- if $Re < 2000$, flow is laminar
- $Re > 3000$ flow is turbulent.
- in-between is a transition zone
- "transition" flow depends on other factors, such as pipe roughness and flow uniformity.

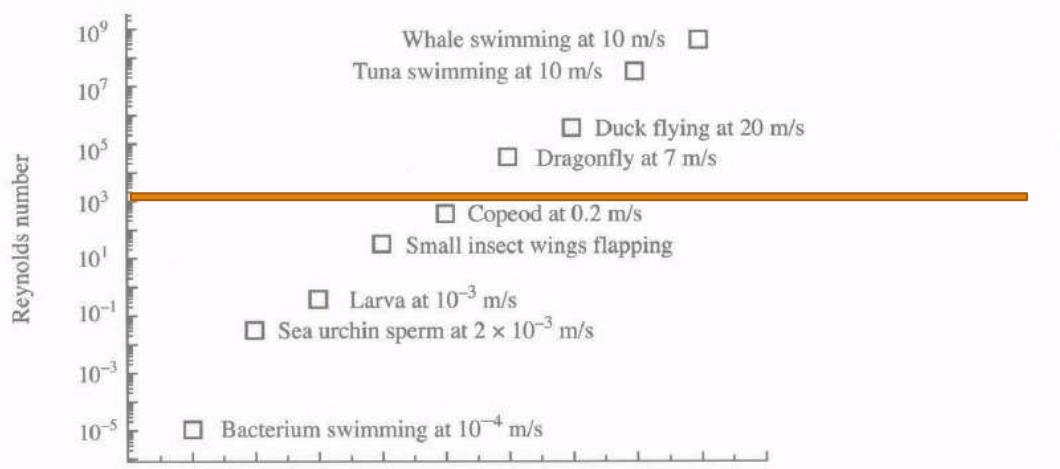


FIGURE 6.10 Reynolds numbers of moving organisms ranging over 14 orders of magnitude. (Data from Vogel, 1994. *Life in Moving Fluids*. Princeton University Press, Princeton, New Jersey.)

Poiseuille's Law

- Here flow (Q) is directly proportional to the pressure gradient
- In turbulent flow:
 - Based on Reynolds number: ↑ pipe diameter and/or ↑ flow, with high density blood, tends towards turbulence
 - Rapid changes in vessel diameter may lead to turbulent flow (e.g. narrower vessel widening to a larger one)

Q = volumetric flow rate

ΔP = pressure drop across the capillary

L = capillary length

μ_e = effective viscosity

R = vessel radius

Assumptions

- 1) Fluid is incompressible and Newtonian
 - i.e. viscous stresses arising from flow, at every point, are linearly proportional to the local strain rate,
- 2) laminar flow through a pipe of constant circular cross-section that is substantially longer than its diameter
- 3) Acceleration of fluid within the pipe doesn't happen.

For velocities and pipe diameters above a threshold, flow is driven to turbulent flow leading to incorrectly larger pressure drops than calculated by the Poiseuille equation

Basic version of Bernoulli's equation:

$$P_1 + \frac{1}{2}\rho v_1^2 + \rho g y_1 = P_2 + \frac{1}{2}\rho v_2^2 + \rho g y_2$$

so that the quantity

$$P + \frac{1}{2}\rho v^2 + \rho g y = \text{const}$$

- between any two points (y_1 and y_2) along the streamline flow.
- an increase in fluid velocity occurs simultaneously with a decrease in pressure or a decrease in the fluid's potential energy.
- Bernoulli's equation can be regarded as a statement of conservation of energy density where each term has dimensions of energy/volume.

Venturi Effect

- describes the change in pressure when fluid flows through a pipe with changing cross-sectional area.
- For simplicity consider the constant height Bernoulli equation $Y_1 = Y_2$ with no change in gravitational potential:

$$P + \frac{1}{2} \rho v^2 = \text{const}$$

Comparing two points along the streamline flow, we can calculate the change in pressure:

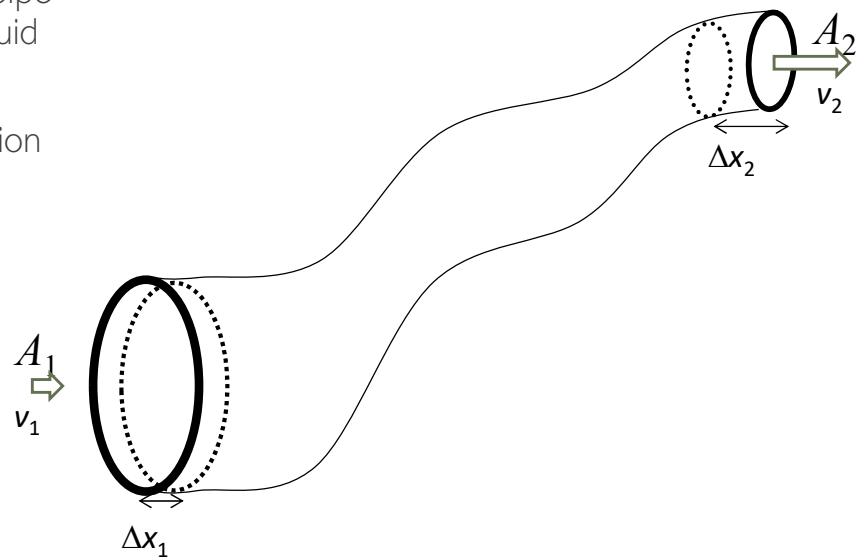
$$P_1 - P_2 = \frac{1}{2} \rho (v_2^2 - v_1^2)$$

so that a constriction in a blood vessel results in a drop in pressure.

$$P_1 - P_2 = \frac{1}{2} \rho v_1^2 \left[\left(\frac{A_1}{A_2} \right)^2 - 1 \right] \quad \text{From the continuity equation}$$

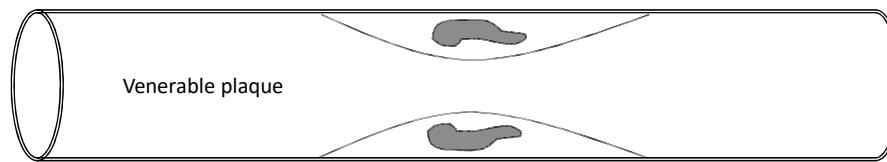
Fluid flow inside a pipe with varying cross section

- Fluid will flow faster through a constriction in a pipe or a blood vessel since the product of area and fluid speed is a constant.
- The increase in fluid speed may lead to a transition from laminar to turbulent flow.



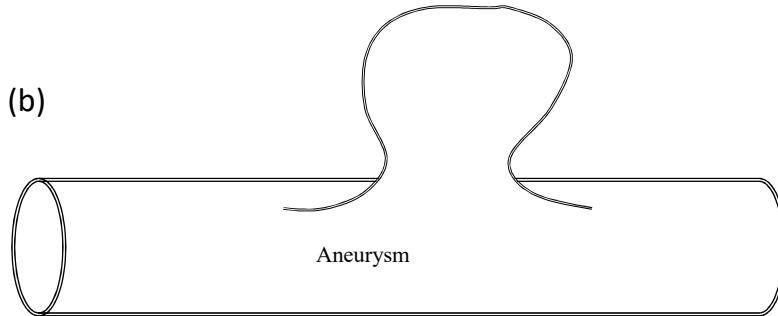
Vessel Volume Changes

(a)



(a) constricted by vulnerable plaque

(b)



(b) expanded by aneurysm

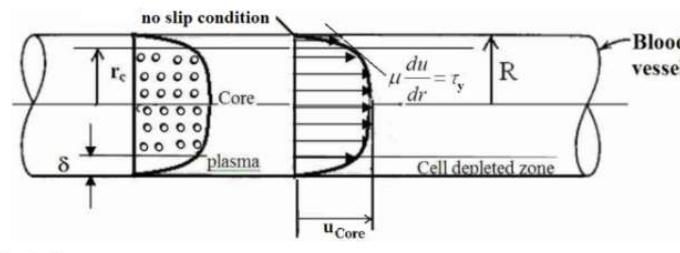
Very Small Vessels

Fåhræus–Lindqvist effect:

- viscosity of a fluid (blood) changes with tube diameter (only if the vessel diameter is between 10 and 300 micrometers).
- erythrocytes move over the center of the vessel, leaving plasma at the wall of the vessel.

Plasma cell-free layer

- also known as skimming layer
- thin layer adjacent to the capillary wall depleted of red blood cells.
- effective viscosity is lower than that of whole blood
- Because the cell-free layer is very thin (approximately 3 μm) this effect is insignificant in larger capillaries

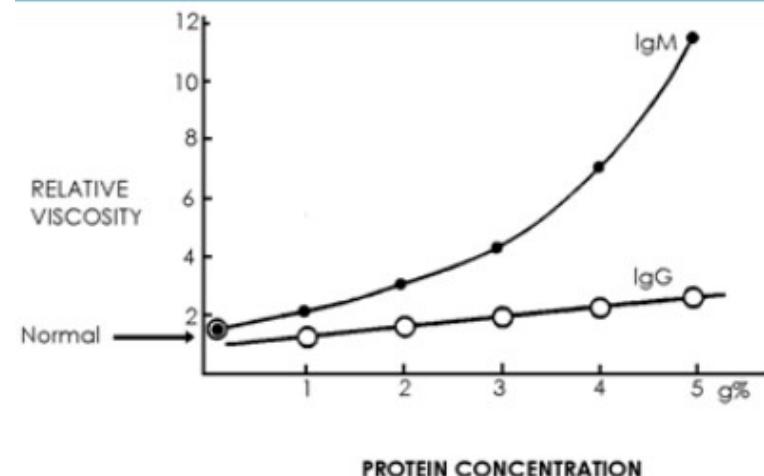


Very Small Vessels

- in truth, blood is non-Newtonian (in smaller vessels!)
- viscosity depends on cell counts and plasma viscosity
- viscosity also depends on shear rate:
 - when low results in Rouleaux formations and sedimentation



Rouleaux formation



Cardiac Modelling Conclusions

Blood behaves very differently based on the scale we analyze it by

Macro level

- Model as a Newtonian fluid
- Look at pressure, volume, HR etc

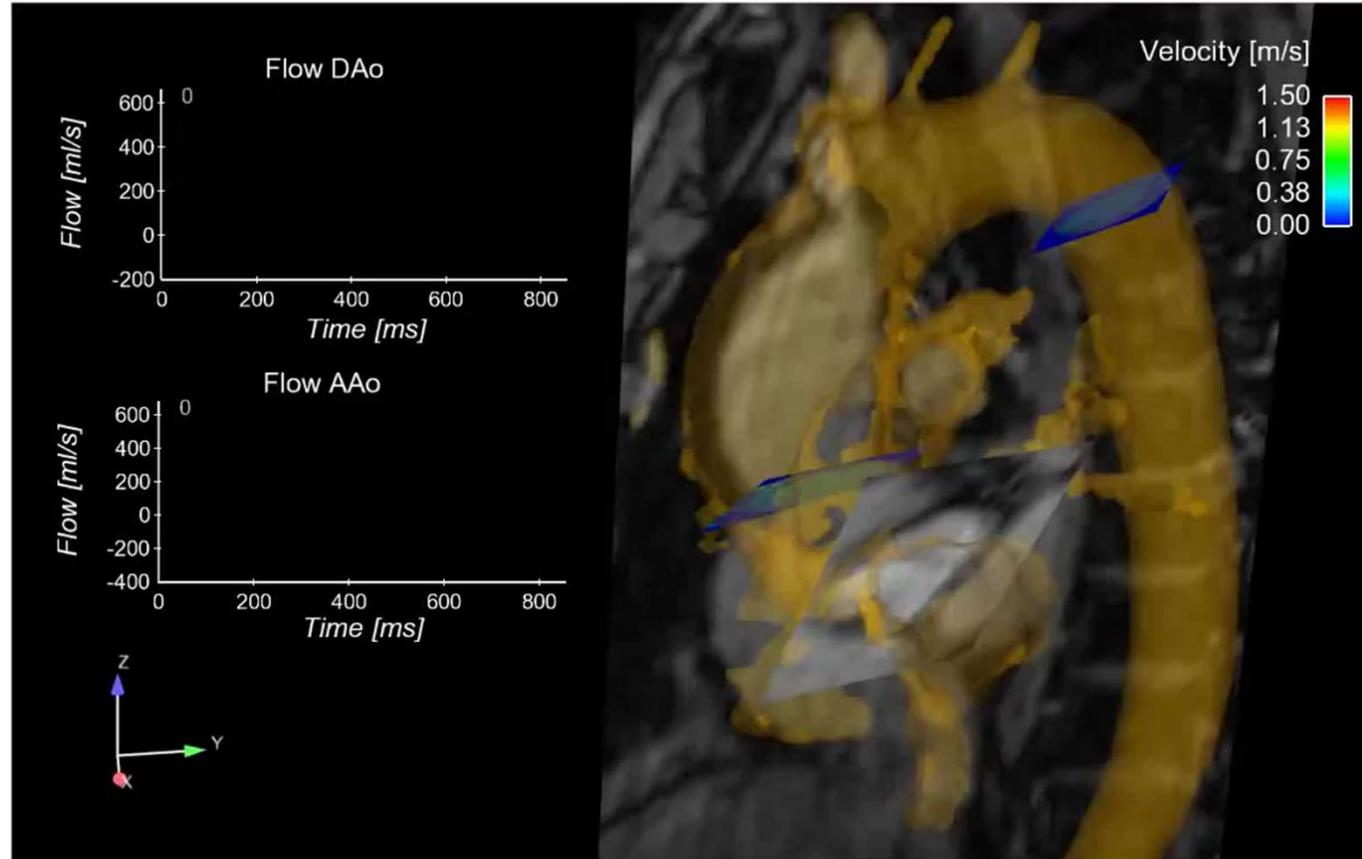
Micro level

- Turbulent vs Laminar flow
- Skimming layer

Capillary Level

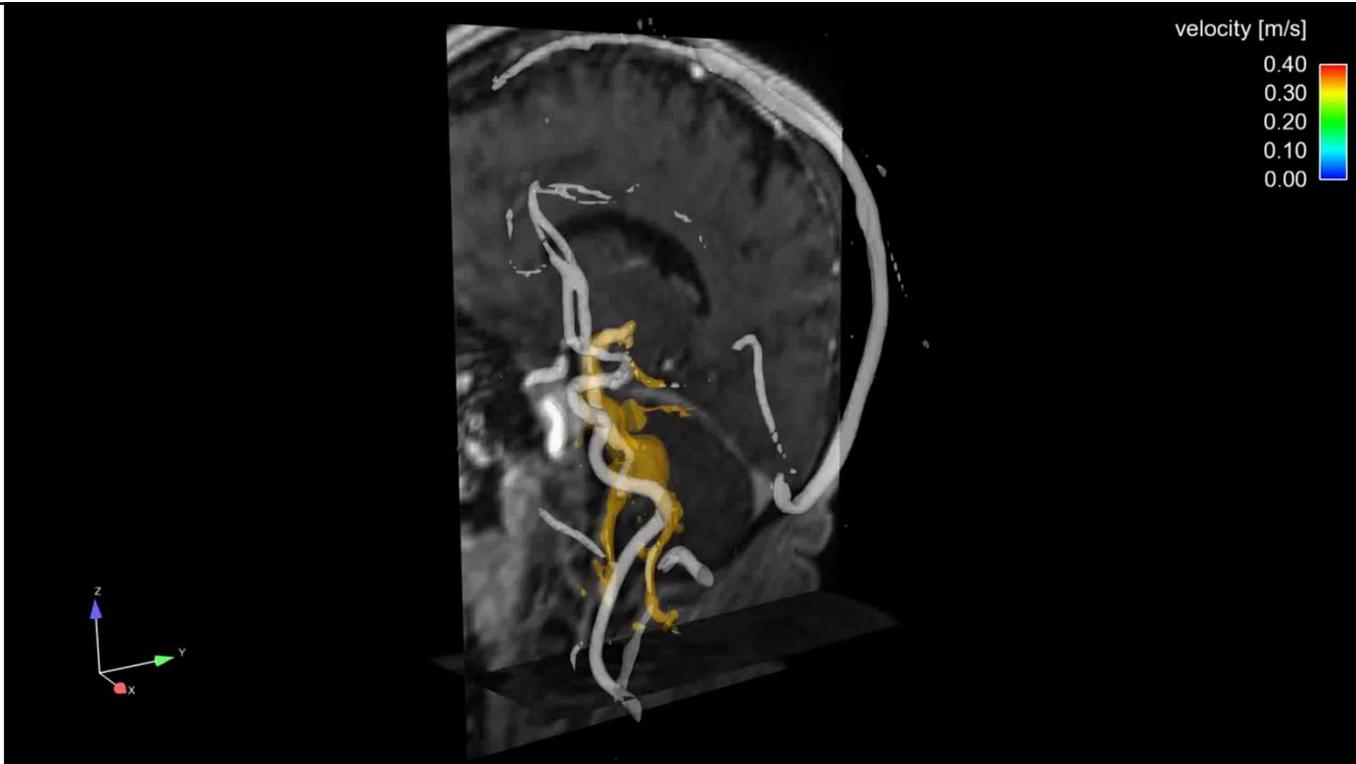
- Collection of solid shapes

Different considerations based on different scales



MRI 4D flow: Patient with a bicuspid aortic valve

(courtesy Dr. Zhaoyang Fan, UCLA)



MRI 4D flow: Patient with a intracranial aneurism

(courtesy Dr. Zhaoyang Fan, UCLA)

Pharmacokinetics

INTRODUCTION

Pharmacokinetics vs Pharmacodynamics

Pharmacokinetics

- the action of drugs in the body over a period of time
- including the processes of absorption, distribution, localization in tissues, biotransformation, and excretion.

Pharmacodynamics

- the study of the biochemical and physiological effects of drugs and the mechanisms of their actions
- including the correlation of their actions and effects with their chemical structure.

<http://medical-dictionary.thefreedictionary.com/>

Why is This Important?

- 1) We use drugs in hospitals!
 - 2) Chemical Engineering, implanted devices, drug delivery systems, etc.
 - 3) Anaesthetics
 - 4) Imaging with contrast agents
 - 5) Imaging drug delivery systems
 - 6) Toxicity, bio-elimination
- Etc.

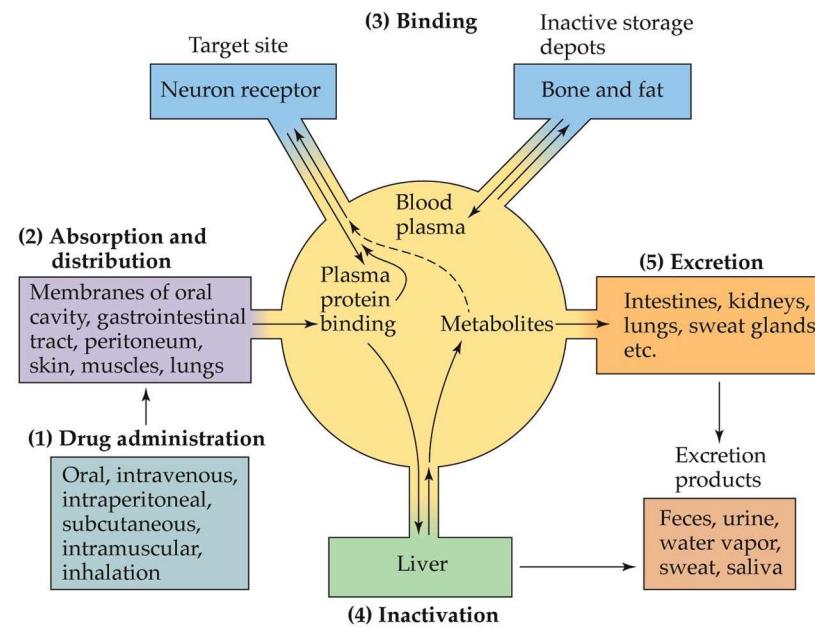
Pharmacokinetics

Drug molecules interact with target sites to affect the nervous system

Pharmacokinetics is the study of drug absorption, distribution within body, and drug elimination **over time**.

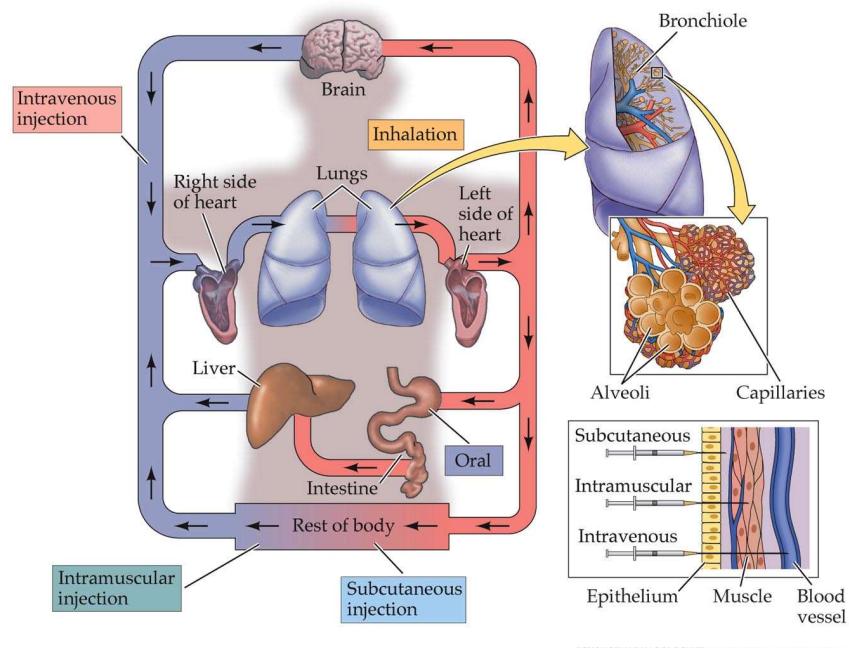
- Absorption depends on the route of administration
- Drug distribution depends on how soluble the drug molecule is in fat (to pass through membranes) and on the extent to which the drug binds to blood proteins (albumin)
- Drug elimination is accomplished by excretion into urine and/or by inactivation by enzymes in the liver

Pharmacokinetics



PSYCHOPHARMACOLOGY, Figure 1.1 © 2005 Sinauer Associates, Inc.

Routes of Administration



Drug Delivery Systems

Tablets

Candy

Injections (Syringe)

Gum

Cigarettes

Implants

Beverages

Gas

Patches

Creams

Suppositories

Others?

- Stamps
- Bandana

Important Terms

Exposure

- A measure for the amount of drug that an organism has really "seen"

Bioavailability

- A measure for the proportion of the dose that reaches the systemic circulation (not the same as exposure)

Clearance

- A measure of the elimination of a compound from the blood given as volume cleared/time
- Usually if <1% left, it is considered cleared

Volume of Distribution

- A measure of the theoretical volume that a compound distributes to.

Unbound Fraction

- The fraction of drug not bound to proteins: $C_{unbound} = f_u \times C_{total}$

Half-Life

- A measure of the time it takes for the organism to decrease the concentration of the drug by 50%

Drug Absorption

The rate at which a drug reaches its site of action depends on:

- **Absorption**
 - involves the passage of the drug from its site of administration into the blood
- **Distribution**
 - involves the delivery of the drug to the tissues

Dr. R. Copeland, Howard University, Washington, D.C

Drug Absorption

Absorption is the process by which a drug enters the bloodstream without being chemically altered

Factors which influence the rate of absorption

- types of transport
- the physicochemical properties of the drug
- protein binding
- routes of administration
- dosage forms
- circulation at the site of absorption
- concentration of the drug

Dr. R. Copeland, Howard University, Washington, D.C

Drug Absorption

Mechanisms of solute transport across membranes

- passive diffusion
- filtration and bulk flow
- endocytosis
- ion-pairing
- active transport

Dr. R. Copeland, Howard University, Washington, D.C

Membranes

Cell Membranes:

- Permeable to many drug molecules but not to others, depending on their lipid solubility
- Small pores, 8 angstroms, permit small molecules such as alcohol and water to pass through.

Walls of Capillaries

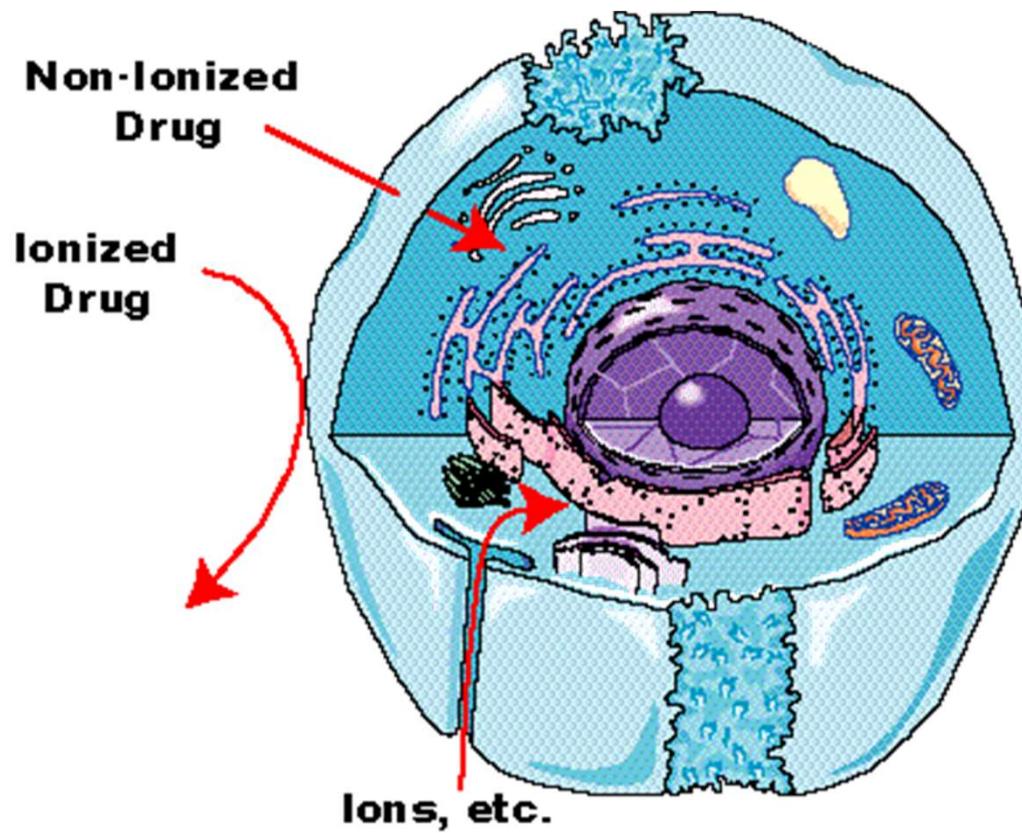
- Pores between the cells are larger than most drug molecules, allowing them to pass freely, without lipid solubility being a factor.

Blood/Brain Barrier

- This barrier provides a protective environment for the brain.
- Speed of transport across this barrier is limited by the lipid solubility of the psychoactive molecule.

Placental Barrier

- This barrier separates two distinct human beings but is very permeable to lipid soluble drugs.



Ion Trapping cont:

Higher concentration of a chemical built up because of a pH difference and the pKa value

Alters urine pH to inhibit toxins from being reabsorbed in the tubules in the kidney

Trap a toxin in an ionized form where it can be excreted

Body fluids where a pH difference from blood pH will favor trapping or reabsorption:

- stomach contents
- small intestine
- breast milk
- aqueous humor (eye)
- vaginal secretions
- prostatic secretions

Dr. R. Copeland, Howard University, Washington, D.C

Ion Trapping:

Kidney:

- Nearly all drugs filtered at the glomerulus:
- Most drugs in a lipid-soluble form will be absorbed by passive diffusion.

To increase excretion: change the urinary pH to favor the charged form of the drug:

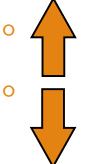
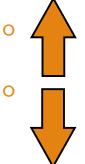
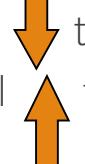
- Weak acids: excreted faster in alkaline pH (anion form favored)
- Weak bases: excreted faster in acidic pH (cation form favored)

Dr. R. Copeland, Howard University, Washington, D.C

Lipid-Water Partition Coefficient

- The ratio of the concentration of a drug in two immiscible phases:
 - a nonpolar liquid or organic solvent (representing the membrane)
 - an aqueous buffer, pH 7.4 (representing the plasma)

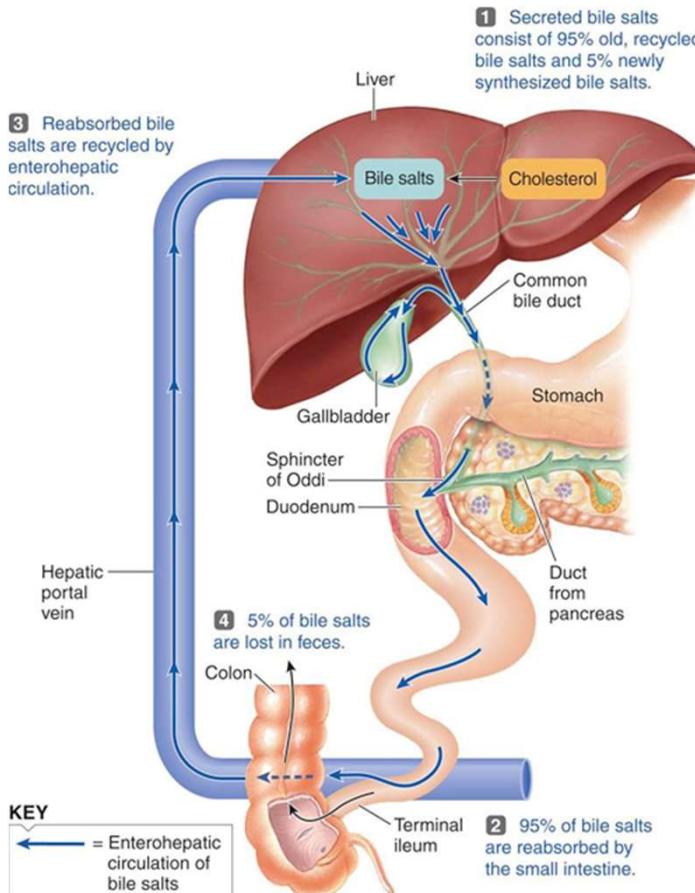
The higher the lipid/water p.c. the greater the rate of transfer across the membrane

-  polarity of a drug, by increasing ionization will  the lipid/ water p.c.
-  polarity of a drug, by suppressing ionization will  the lipid/ water p.c.

Dr. R. Copeland, Howard University, Washington, D.C

First-pass Effect

- Term used for the hepatic (liver) metabolism of a pharmacological agent when it is absorbed from the gut and delivered to the liver via the portal circulation.
- The greater the first-pass effect
 - the less the agent will reach the systemic circulation when the agent is administered orally.
 - the lower the bioavailability of the drug (the rate and extent of drug reaching systemic circulation).



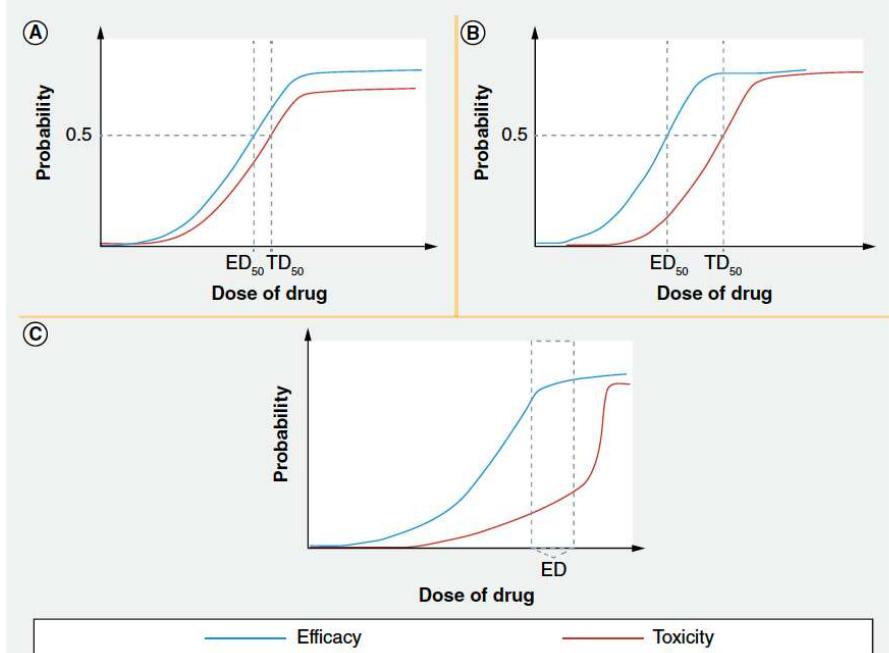
Magnitude of first pass hepatic effect:

Extraction ratio (ER):

Q = hepatic blood flow (~ 90 L per hour).

Systemic drug bioavailability (F) may be determined from the extent of absorption (f) and the extraction ratio (ER):

Therapeutic Index



Therapeutic index (TI) = TD_{50}/ED_{50}

(B) Better TI when the effect curve is displaced to left ($ED_{50} \ll TD_{50}$).

(C) Increased TI by preferentially directing the drug to tumor cells and or reducing toxicity

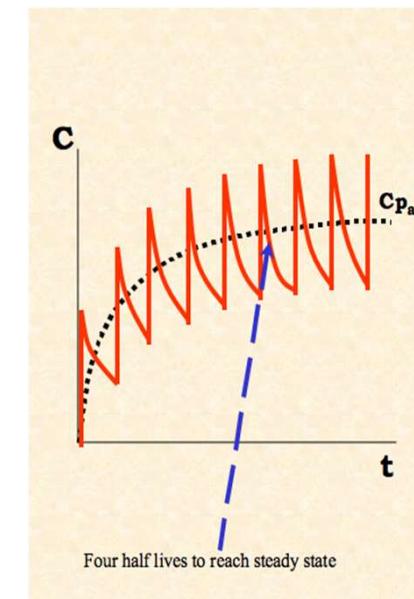
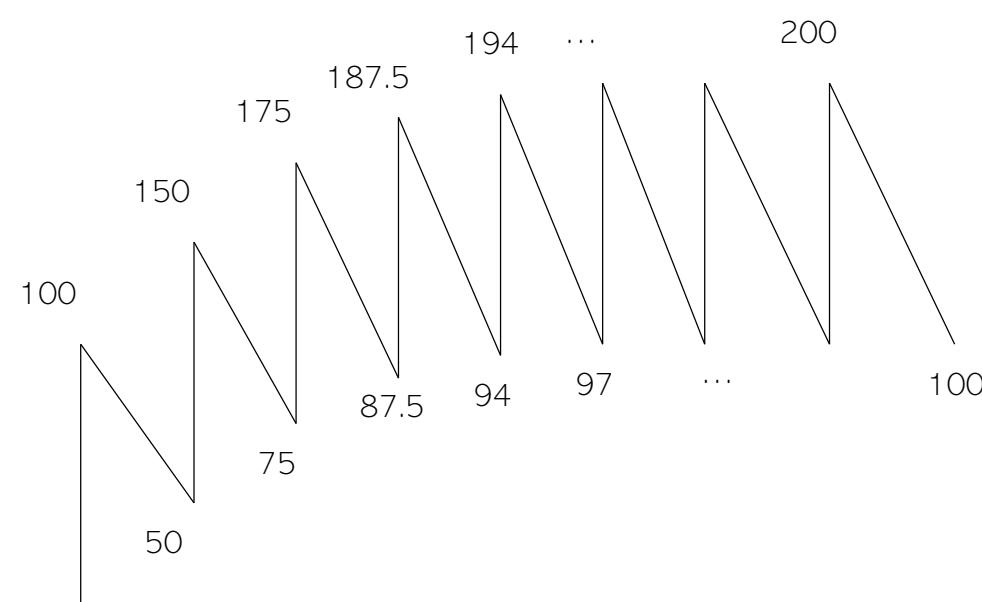
Steady-State

Steady-state occurs after a drug has been given for ~ 5 to 6 elimination half-lives.

- Rate in = Rate Out
- Reached in ~ 5 or 6 half-lives (linear kinetics)
- Important when interpreting drug concentrations in time-dependent manner or assessing clinical response
- Repeated doses are used to maintain a steady state as the body metabolizes/excretes the previous dose

Accumulation to Steady State

100 mg given every half-life

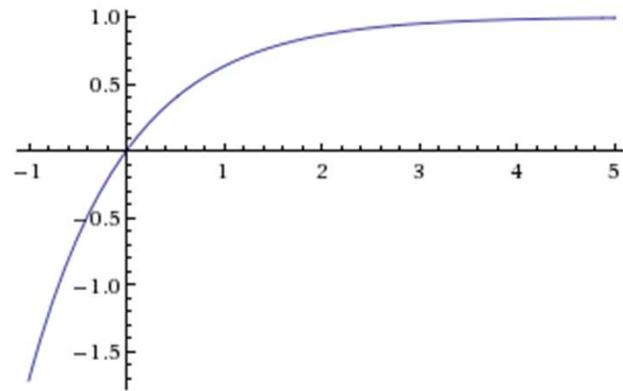


6 doses to reach steady state

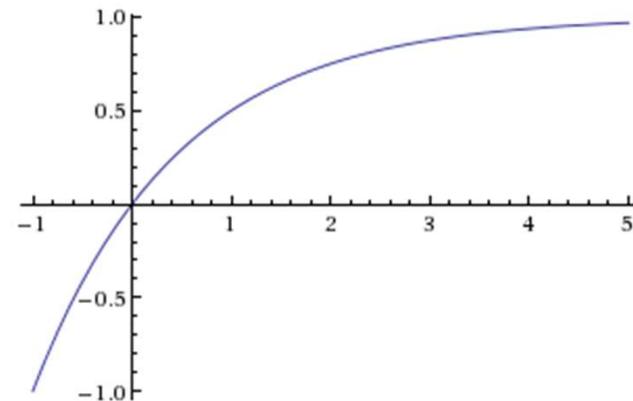
Rate of Absorption

exponential VS power model:

k = rate constant (time $^{-1}$)



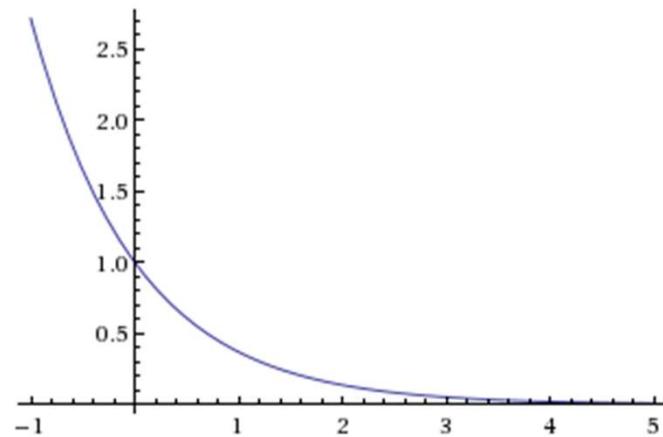
$$C(t) = C_i (1 - e^{-t/k})$$



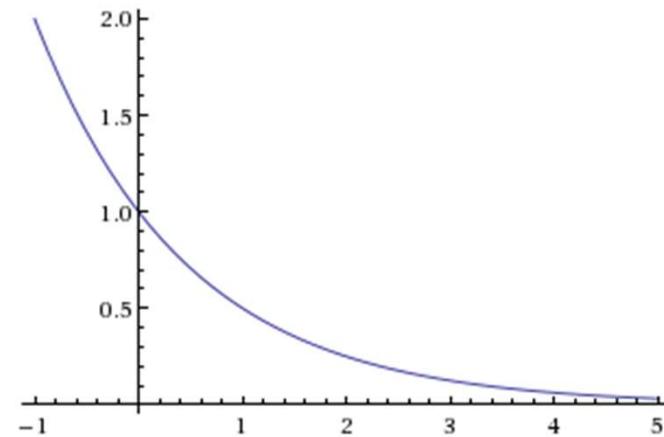
$$C(t) = C_i (1 - 2^{-t/k})$$

Rate of excretion

exponential VS power model:



$$C(t) = C_i (e^{-t/k})$$

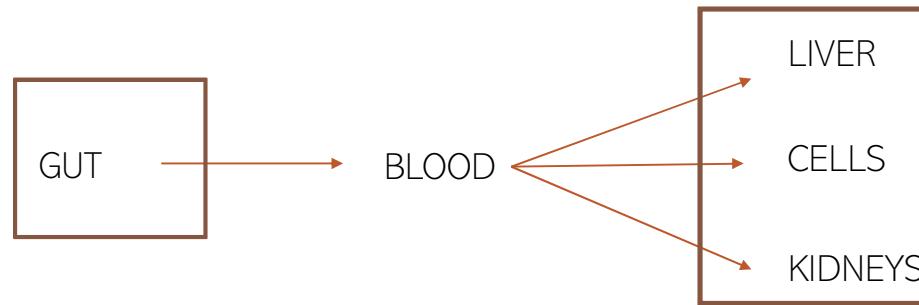


$$C(t) = C_i (2^{-t/k})$$

Pharmacokinetics

Outputs VS inputs – PRODUCT

Multiple inputs/outputs - SUM



Routes of Drug Administration

The **route of administration** (ROA) that is chosen may have a profound effect upon the speed and efficiency with which the drug acts

Definition

- A route of administration is the path by which a drug, fluid, poison or other substance is brought into contact with the body.

Dr. R. Copeland, Howard University, Washington, D.C

Classification

Routes of administration can broadly be divided into:

Topical

- Drugs are applied topically to the skin or mucous membranes, mainly for local action.
- Oral (aka PER OS)
 - used for systemic (non-local) effect, substance is given via the digestive tract.

PARENTERAL

- A drug administered parenterally is one injected via a hollow needle into the body at various sites and to varying depth.

Rectal

- Drugs given through the rectum by suppositories or enema.

Inhalation

- The lungs provide an excellent surface for absorption when the drug is delivered in gaseous, aerosol or ultrafine solid particle form.

Topical route

I Skin

A-Dermal

- cream, ointment (local action)

B- Transdermal

- absorption of drug through skin (i.e. minimal systemic action)
 - I. stable blood levels (controlled drug delivery system)
 - II. No first pass metabolism
 - III. Drug must be potent or patch needs will be [too] large

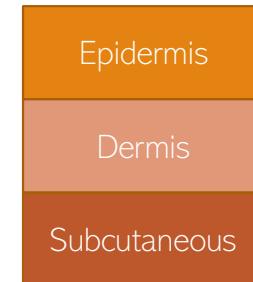


II Mucosal membranes

- eye drops (onto the conjunctiva)
- ear drops
- intranasal route (into the nose)

Topical route

- Drugs applied locally on the skin are poorly absorbed through the epidermis.
- However, the dermis is permeable to many solutes. Thus systemic absorption of drugs occurs more readily through abraded, burned or denuded skin.
- Inflammation and other conditions that enhance cutaneous blood flow also promote absorption.
- Drugs are applied in the form of ointments, pastes, poultice and cream to the skin for their local action.
- Absorption through skin can be increased by suspending the drug in an oily vehicle and rubbing the preparation into the skin. This method of administration is called inunction.
- To increase absorption drugs are applied onto the various mucous membranes for their local action.



Topical Route

ADVANTAGES

- Easy to apply
- Low risk of drug interactions
- High concentration of antibiotic to affected area
- Lack of effects on intestinal flora
- Low cost
- Avoid first pass metabolism
- Easy termination

DISADVANTAGES

- Most drugs have too high of a molecular weight and are not lipid soluble
- Local skin irritation
- Contact dermatitis with some drugs may occur
- Skin enzymes can break down drug
- Can be used with drugs that require low plasma concentration

Oral route

(aka: Per Os)

- By swallowing.

intended for systemic effects resulting from drug absorption through the various epithelia and mucosa of the gastrointestinal tract.



Oral route (Continued):

- Food and gastrointestinal motility can affect drug absorption.
 - Absorption may be slow, unpredictable and irregular because of the presence of variable amounts of food at various stages of digestion
 - Absorption is slower with food (milk and milk products) for **tetracyclines** and **penicillins**, etc. However, for **propranolol** bioavailability is higher after food, and for **griseofulvin** absorption is higher after a fatty meal.

Oral route (Continued):

Sometimes may have adverse reactions – e.g. Antibiotics may kill normal gut flora and allow overgrowth of fungal varieties. Thus, antifungal agent may be included with an antibiotic.

Not suitable for unconscious patient - Patient must be able to swallow solid dosage forms. Liquids may be given by tube.

- May cause irritation to gastric mucosa, nausea and vomiting.
- Effect too slow for emergencies.
- Some drugs are destroyed by intestinal enzymes e.g. insulin is destroyed by intestinal enzymes.
- There is a necessity for cooperation on the part of patient.

Oral route

ADVANTAGES

- 1) Convenient - portable, no pain, easy to take.
- 2) Cheap - no need to sterilize, compact, multi-dose bottles, automated machines produce tablets in large quantities.
- 3) Variety - tablets, capsules, suspensions, mixtures .

DISADVANTAGES:

- 1) Sometimes inefficient - low solubility drugs may suffer poor availability e.g. Griseofulvin
- 2) First-pass effect - drugs absorbed orally are transported to the general circulation via the liver. Thus drugs which are extensively metabolized will be metabolized in the liver during absorption. e.g. propranolol

Buccal/Sublingual route:

Some drugs are taken as smaller tablets which are held in the mouth (buccal tablet) or under the tongue (sublingual tablet).

Buccal tablets are often harder tablets [4 hour disintegration time], designed to dissolve slowly.

Buccal/Sublingual route:

- drug is very potent so needs few molecules to be absorbed in order to produce the therapeutic effect.
- major advantage of this route is that venous drainage from mouth (buccal cavity) is poured into the superior vena cava and the drug is saved from first-pass effect.

Buccal/Sublingual route:

ADVANTAGES

1. Avoid hepatic first pass - The liver is by-passed thus there is no loss of drug by first pass effect for buccal administration. Bioavailability is higher.
2. Rapid absorption - Because of the good blood supply to the area, absorption is usually quite rapid.
3. Drug stability - pH in mouth relatively neutral (stomach - acidic). Thus a drug may be more stable.

DISADVANTAGES

- 1- Holding the dose in the mouth is inconvenient.
- 2- Small doses only can be accommodated easily.



Lecture 5

TAYLOR DEVET MASC.

PHD. CANDIDATE BIOLOGICAL AND BIOMEDICAL ENGINEERING

MCGILL UNIVERSITY

SHRINERS HOSPITAL FOR CHILDREN

Todays Aims...

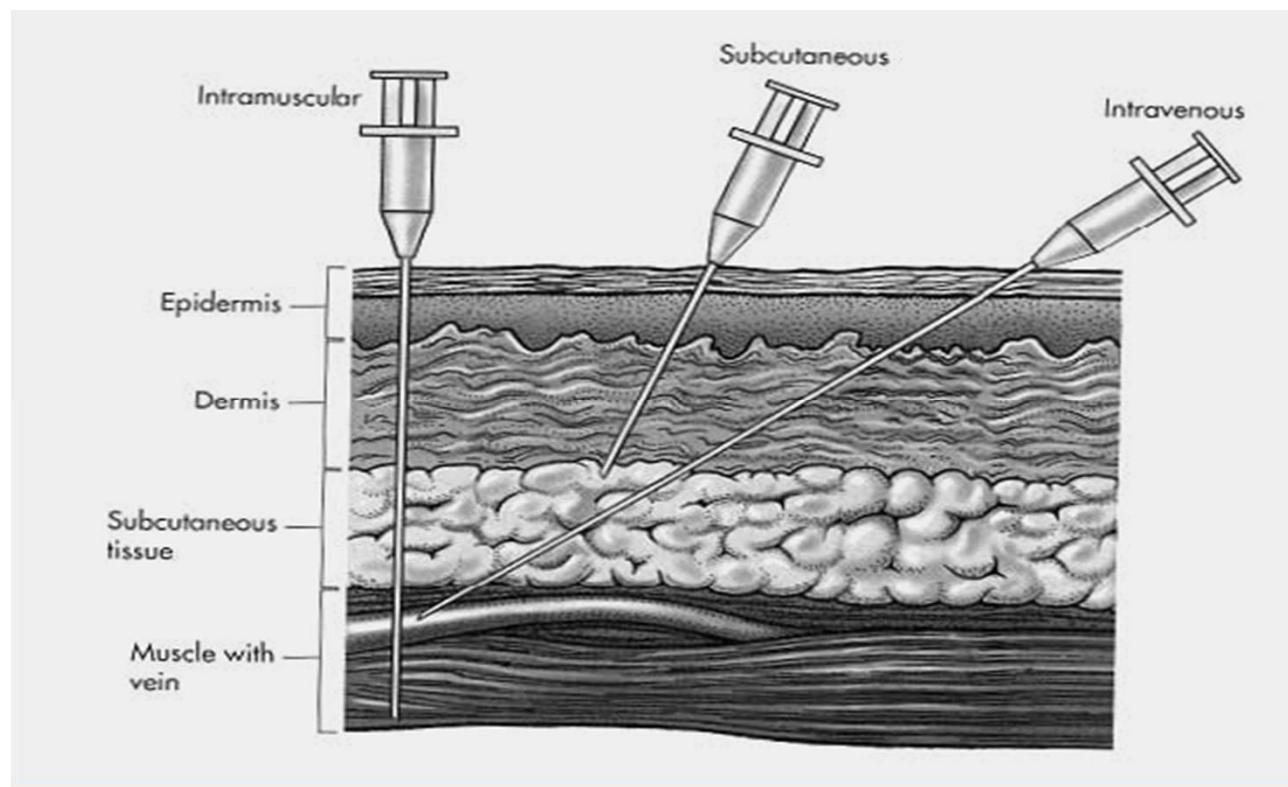


Drug delivery routes cont



Drug concentration models

Parenteral route:

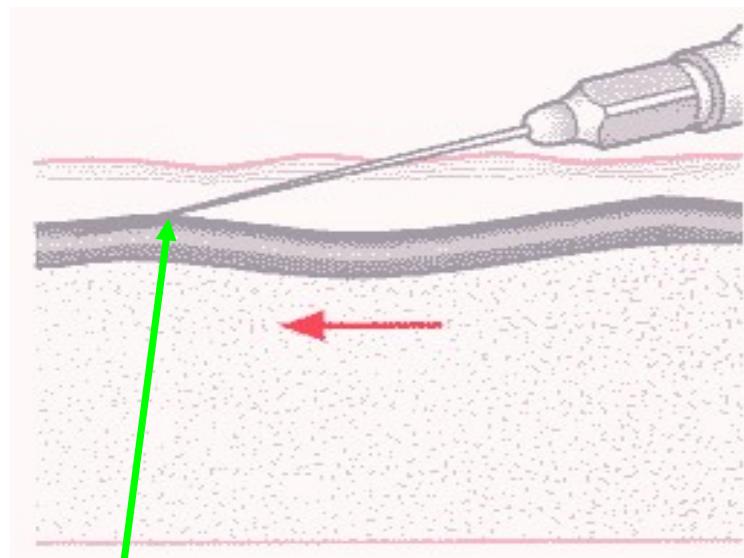


Parenteral route – Intravascular (IV/IA)

- placing a drug directly into blood stream.
- May be -intravenous (into a vein) or - intraarterial (into an artery).
- This route is of prime importance in emergency.
- this is the only route for giving large volume of drugs e.g. blood transfusion.
- However, there are certain disadvantages of this procedure.
 - Once the drug is injected nothing can be done to prevent its action.
 - I/V injection requires technical skill to minimize the risk of leakage of irritant solution into the surrounding tissues.
 - Air embolism may cause serious problems.



www.shutterstock.com - 1181263



Intravenous Administration

Parenteral route – Intravascular (IV)

ADVANTAGES

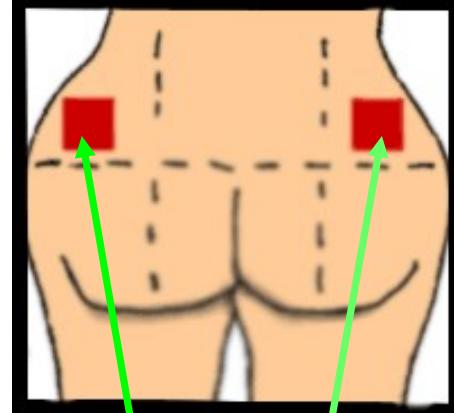
- 1- precise, accurate and immediate onset of action
- 2 - 100% bioavailability.

DISADVANTAGES

- 1- risk of embolism.
- 2- high concentrations attained rapidly leading to greater risk of adverse effects.
- 3- need a trained professional for administration

Parenteral route - Intramuscular

- Into skeletal muscle
- In humans, the best site is deltoid muscle in the shoulder or the gluteus muscle in the buttocks.
- is suitable for the irritating substances that cannot be given by subcutaneous route.
- speed of absorption from site of injection is dependent on the vehicle used
 - absorption is quick from aqueous solutions
 - slow from oily preparations.
- Absorption is complete, predictable and faster than subcutaneous route.



Intramuscular injection in deltoid and gluteal muscles

Parenteral route - Intramuscular

ADVANTAGES

1 - suitable for injection of drug in aqueous solution (rapid action) and drug in suspension or emulsion (sustained release).

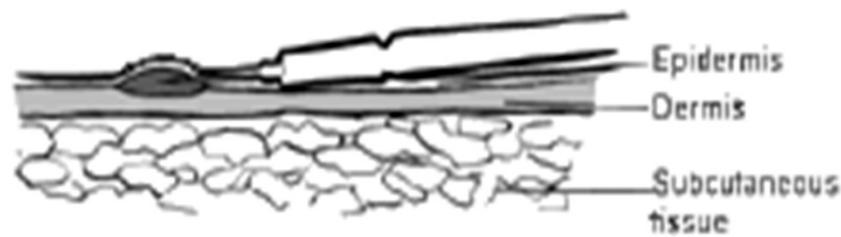
DISADVANTAGES

- 1- Pain at injection sites for certain drugs.
- 2 - Need a trained professional for administration

Parenteral route – Subcutaneous

C- Subcutaneous

- under the skin
- e.g. insulin.
- Includes embedded pellets



Parenteral route – Subcutaneous

ADVANTAGES

Good for skin infections

Safer than IV and IM

Slower than IM and IV

Easier for people to self administer

Can do depot injections

DISADVANTAGES

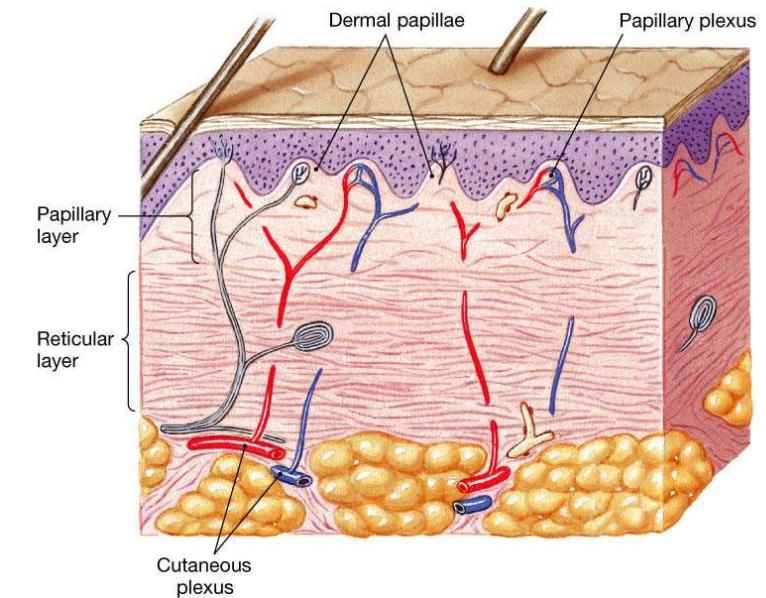
Not good for irritating drugs

Slow absorption

Smaller volumes at once

Parenteral route - Intradermal

- skin testing for some allergens
- drug is injected into papillary layer of skin.
- Tuberculin injection for montoux test
- BCG vaccination for active immunization against tuberculosis.





Intradermal Injection

Parenteral route - Intradermal

ADVANTAGES

- Can give lower dose than sub-cut or IM
- Better response for some vaccines
- Reduced cost
- Better response in immunocompromised or elderly

DISADVANTAGES

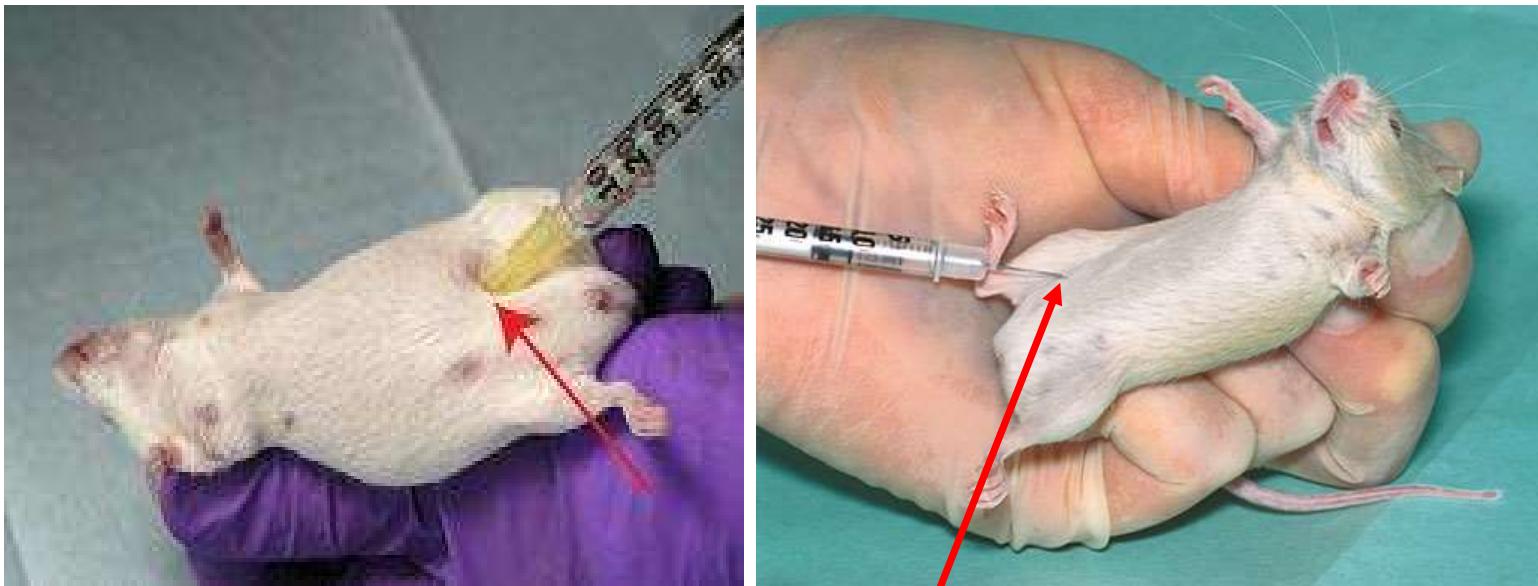
- Small doses
- Trained professional to get needle to correct depth
- Reactions at injection site
- Scarring



Parenteral route - Intraperitoneal

- infusion or injection into the peritoneum
- The peritoneum offers a large absorbing surface area from which drugs enter circulation rapidly but primarily by way of portal vein.
- This is probably the most widely used route of drug administration in laboratory animals.
- In human, it is very rarely employed due to the dangers of infection and injury to viscera and blood vessels.

Parenteral route - Intraperitoneal



Intraperitoneal Injection

Parenteral route - Intrapерitoneal

ADVANTAGES

Easy in mice

Direct targeting of peritoneal cancers

Fluids to infants

Can be used when large amounts of fluids are needed but IV is not an option

Better outcomes for dialysis

DISADVANTAGES

More difficult in humans

Easier ways

Rectal route

Most commonly by suppository or enema.

ADVANTAGES

1. By-pass liver
 - Some of the veins draining the rectum lead directly to the general circulation thus by-passing the liver.
 - Reduced first-pass effect.
2. This route may be most useful for patients unable to take drugs orally
 1. unconscious patients
 2. younger children
 3. patient is nauseous or vomiting

DISADVANTAGES

1. Erratic absorption
 - Absorption is often incomplete and erratic.
2. Not well tolerated by patients.



Inhalation route

- Used for gaseous and volatile agents and aerosols.
- solids and liquids are excluded if larger than 20 micron.

ADVANTAGES

1. Large surface area
2. thin membranes separate alveoli from circulation
3. high blood flow
4. As result of that a rapid onset of action due to rapid access to circulation.

DISADVANTAGES

1. Most addictive route of administration because it reaches the brain so quickly.
2. Difficulties in regulating the exact amount of dosage.
3. Sometimes patients have difficulties in giving themselves a drug by inhaler.



Other Routes

- a) Subcutaneous (S/C)
- b) Intramuscular (I/M)
- c) Intravenous (I/V)
- d) Intraperitoneal (I/P)
- e) Intradermal
- f) Intra Medullary
- g) Intrathecal
- h) Intraarticular
- i) Intra-cardiac
- j) Intra arterial

Intra Medullary

- The needle is introduced into marrow cavity
- effects are similar to those following intravenous injection
- Route is used when veins are not available especially in children.
- In adults the injection is made into marrow cavity of sternum
- Under 3 years of age into tibia or femur.

Intrathecal

- Theca = dura
- Thecal sac surrounds spinal cord and cauda equina
- Blood brain barrier often prevents the entry of certain drugs into the central nervous system
- effects of the drugs are then localized to the spinal nerves and meninges
- injection of local anesthetics for the induction of spinal anesthesia is given by this route.

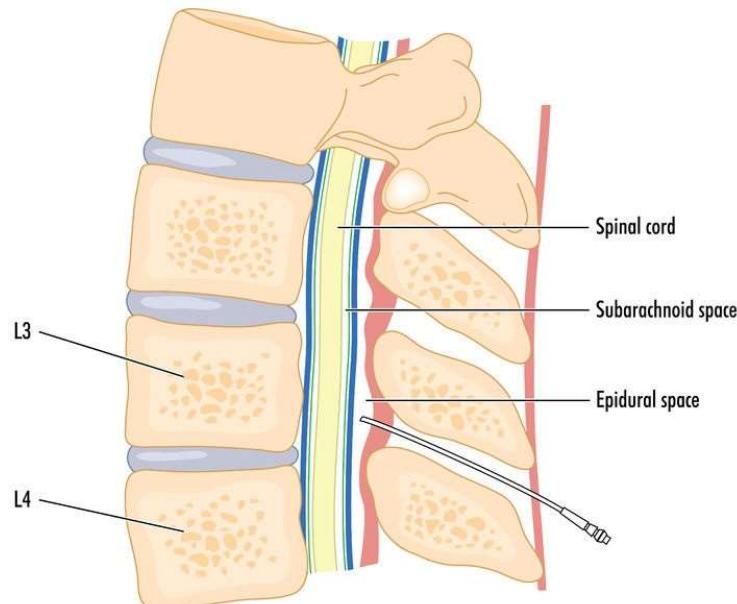
Parenteral route - Intrathecal

ADVANTAGES

- Effective pain relief
- Decrease surgical intervention for nerve pain
- Helps with functional disability due to pain, illness etc

DISADVANTAGES

- Difficult to do
- Painful to administer
- Effects BP
- May induce fever
- Potential o permanent nerve damage



Intra-articular

- also known as intrasynovial
- Sometimes drugs are injected into the joint cavity to localize their action at the site of administration
- e.g. Hydrocortisone acetate in the treatment of rheumatoid arthritis.
- Local anesthetic is added to minimize pain of injection.
- Strict asepsis must be maintained to avoid joint-infection.

Intra Cardiac

- Into the cardiac muscle
- In cardiac arrest intra-cardiac injection of adrenaline is made for resuscitation
 - naloxone for heroin OD

Intra-arterial

- Sometimes a drug is injected directly into an artery to localize its effects in a particular tissue or organ.
- the therapeutic value of such practice is doubtful.

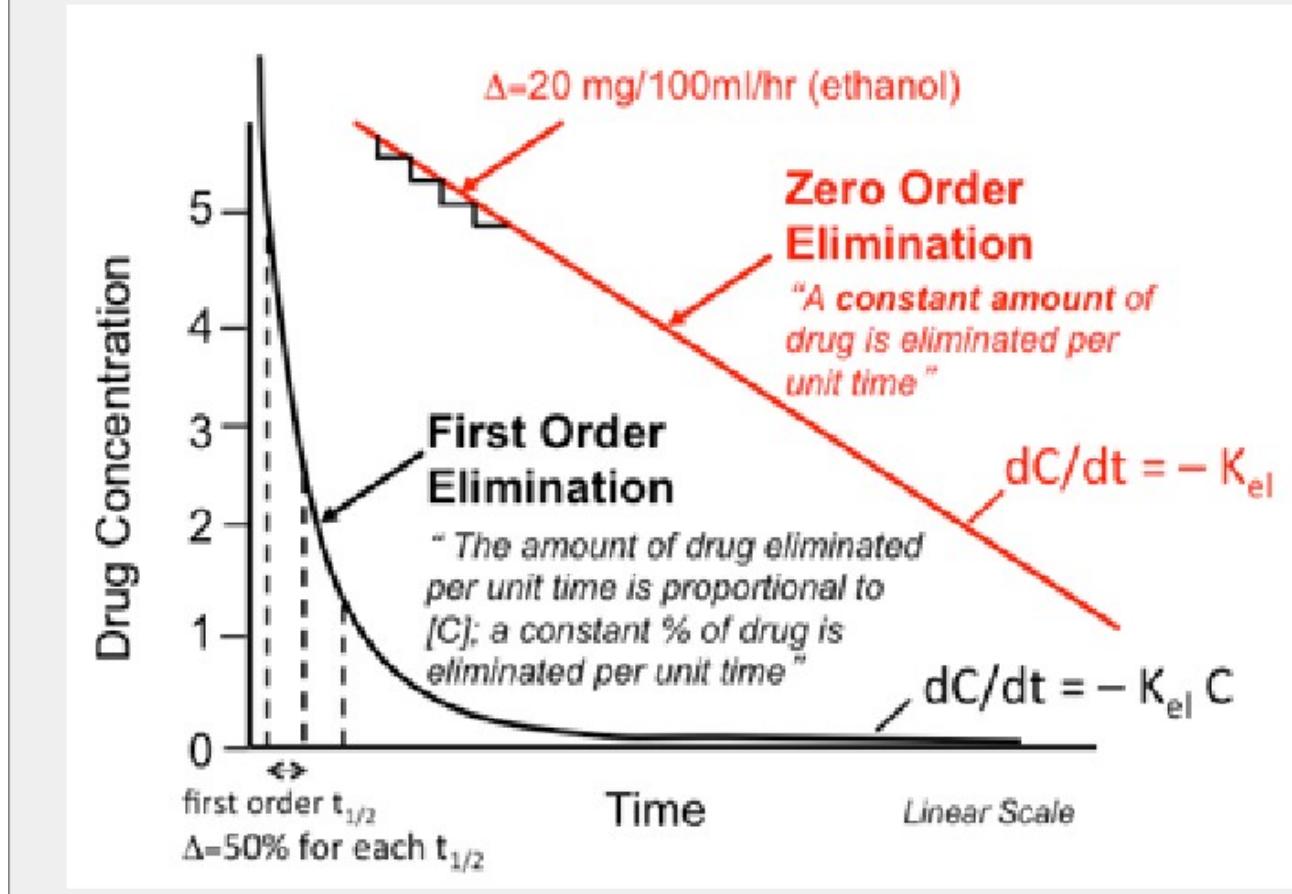
Time-release preparations - Oral

- controlled-release, timed-release, sustained-release
- designed to produce slow, uniform absorption for 8 hours or longer
- better compliance, maintain effect over night, eliminate extreme peaks and troughs

Time-release preparations - Depot

- Depot or reservoir preparations
- parental administration (except IV)
- may be prolonged by using insoluble salts or suspensions in non-aqueous vehicles.

Delivery Kinetics



Region of Action (ROA)

- determined by the physical characteristics of the drug
 - the speed which the drug is absorbed and/ or released
 - the need to bypass hepatic metabolism
 - the need to achieve high conc. at particular sites
-
- No single method of drug administration is ideal for all drugs in all circumstances

Drug Distribution

Dependent upon its route of administration and target area

Tissue is composed of cells which are encompassed within membranes

- Membranes consist of 3 layers
- 2 layers of water-soluble complex lipid molecules (phospholipid)
- a layer of liquid lipid sandwiched within these layers
- Suspended within the layers are large proteins, .

Drug Distribution

- The permeability of a cell membrane, for a specific drug, depends on a ratio of its water to lipid solubility.
- Within the body, drugs may exist as a mixture of two interchangeable forms
 - water (ionized-charged) soluble
 - lipid (non-ionized) soluble
- The concentration of two forms depends on characteristics of the drug molecule

Overview

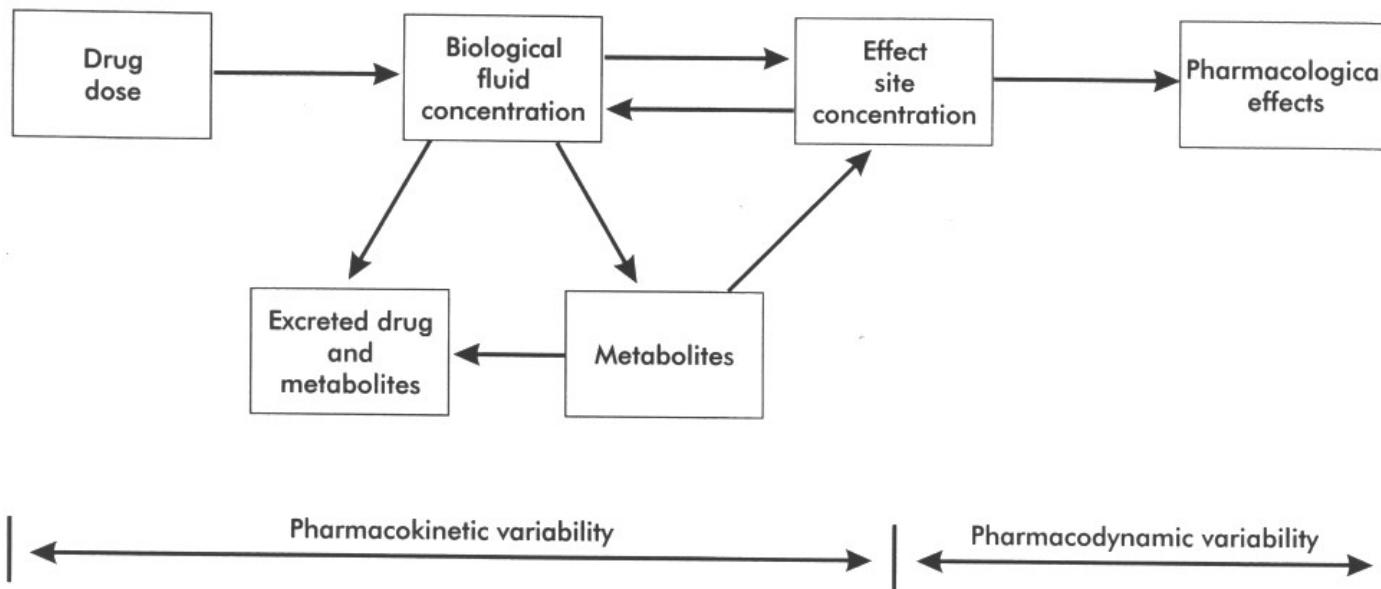


Figure 8-1. Pharmacokinetic and pharmacodynamic variability as determinants of the dose-effect relationship.

Study of drug concentration over time

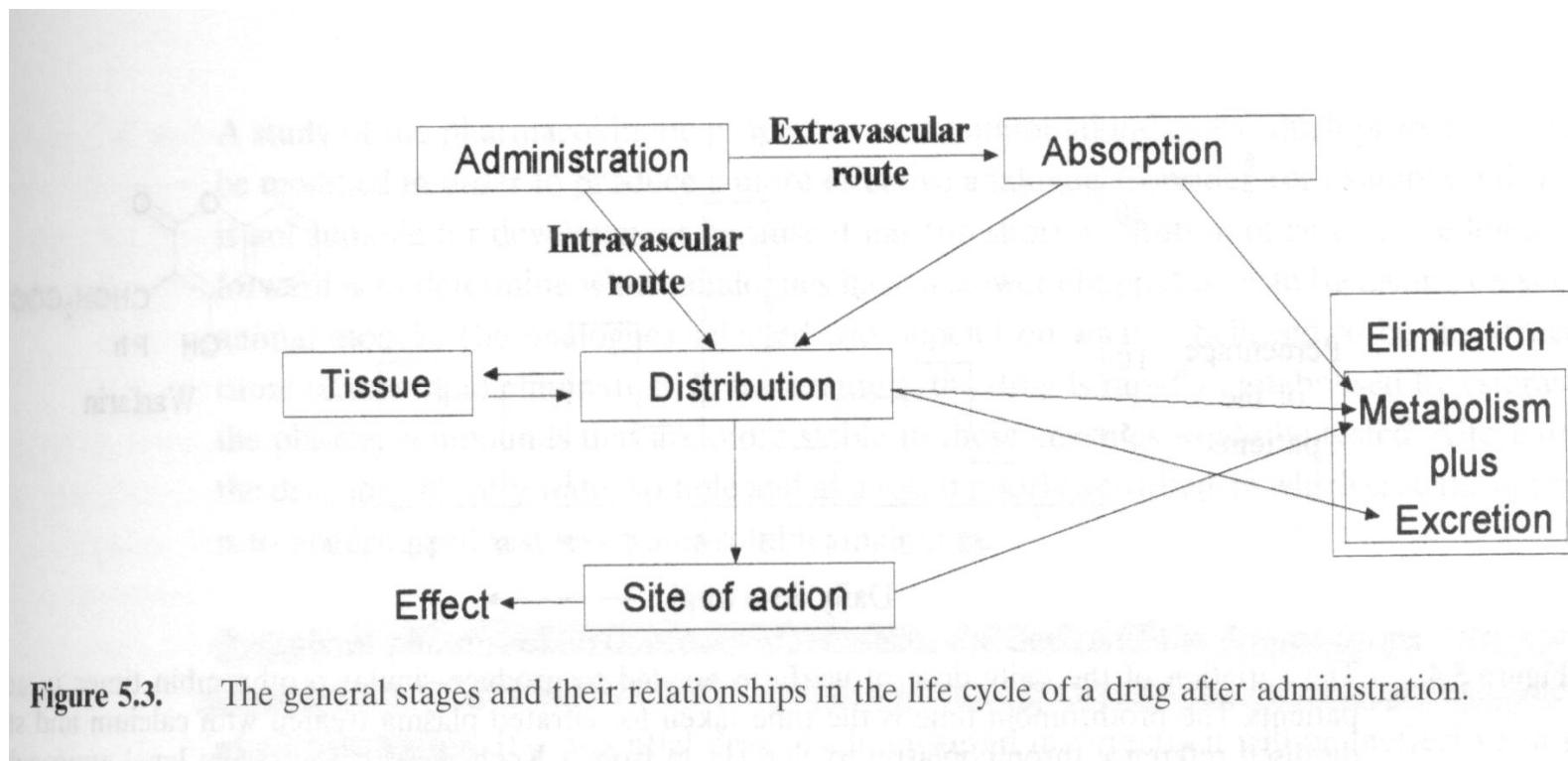


Figure 5.3. The general stages and their relationships in the life cycle of a drug after administration.

Administration Route: Time until effect

Route	Time
intravenous	30-60 seconds
intraosseous	30-60 seconds
endotracheal	2-3 minutes
inhalation	2-3 minutes
sublingual	3-5 minutes
intramuscular	10-20 minutes
subcutaneous	15-30 minutes
rectal	5-30 minutes
ingestion	30-90 minutes
transdermal (topical)	variable (minutes to hours)

Drug Size vs delivery method

- Absorption
 - depends on the route of administration
 - Hydrophobic vs non
 - More lipophilic, more metabolized by gut wall
 - Water soluble are cleared more slowly from the body via kidneys
- Drug distribution
 - depends on how soluble the drug molecule is in fat
 - extent to which the drug binds to blood proteins (albumin)
- Drug elimination
 - Liver
 - Kidney
 - Used up
 - Fat deposits

Reminder on sizes...

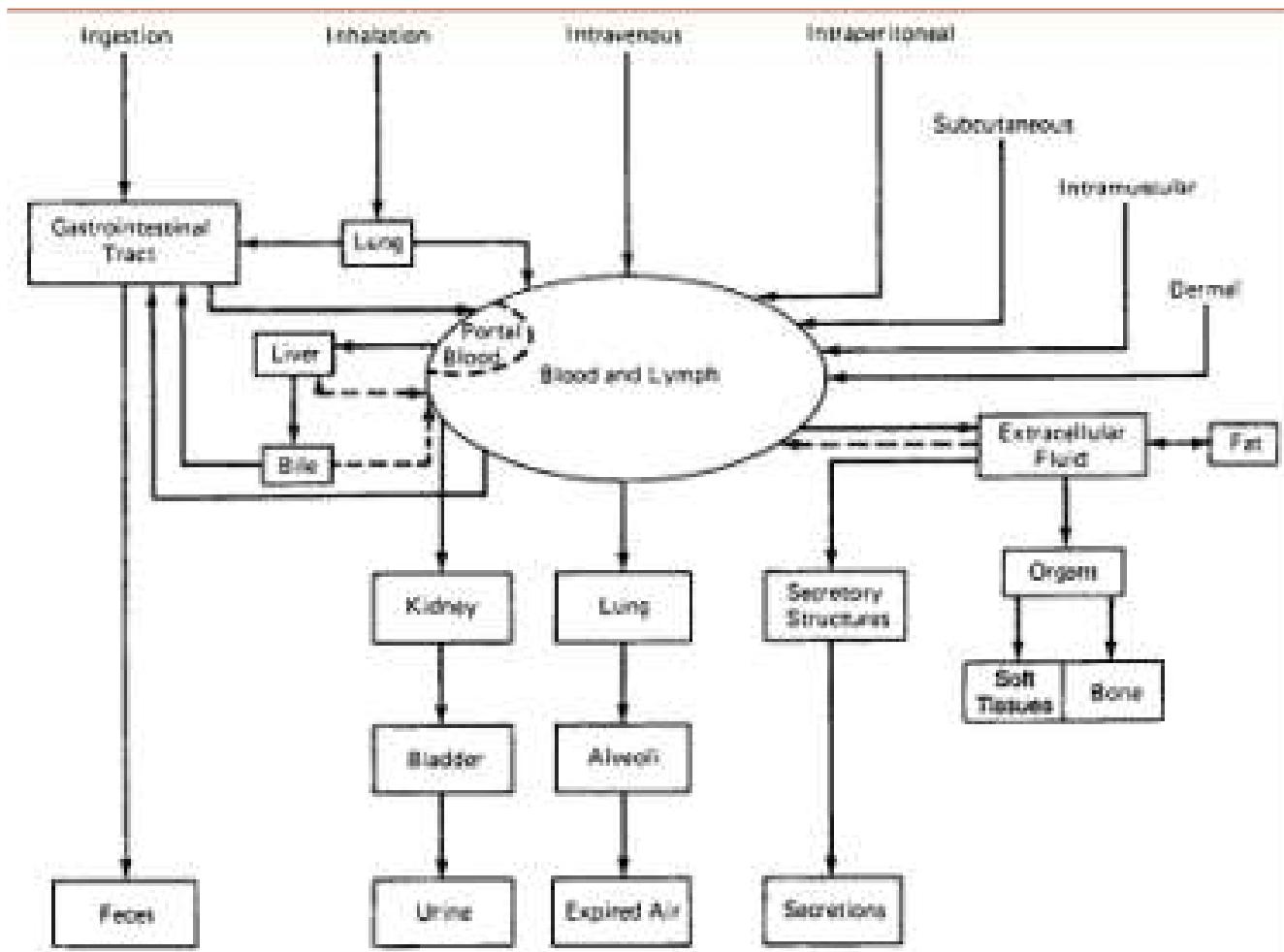
Cell Membranes: Small pores, 8 angstroms, permit small molecules such as alcohol and water to pass through.

Walls of Capillaries: Pores between the cells are larger than most drug molecules, allowing them to pass freely, without lipid solubility being a factor.

Blood/Brain Barrier: This barrier provides a protective environment for the brain. Speed of transport across this barrier is limited by the lipid solubility of the psychoactive molecule.

Placental Barrier: This barrier separates two distinct human beings but is very permeable to lipid soluble drugs.

DEPOSITION OF DRUGS



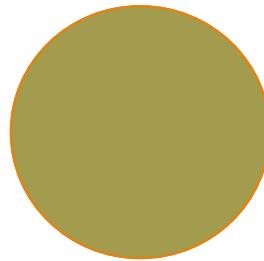
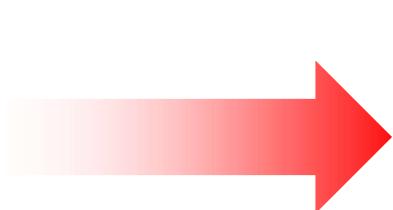
The disposition of chemicals entering the body (from C.D. Klaassen, *Casarett and Doull's Toxicology*, 5th ed., New York: McGraw-Hill, 1996).

Particle Dynamics

Maxwell-Boltzmann Equation

Before we look at diffusion in biological systems, need to examine underlying principles.

The Maxwell-Boltzmann equation governs the motion of small molecules in a gas.



$$E = \frac{mv^2}{2}$$

Maxwell-Boltzmann Equation

From the research of Boltzmann, the energy of a molecule in a gas is known to be proportional to temperature (T) and the Boltzmann constant (k_B):

$$P(\text{m}_i \text{ has energy } E_i) \approx \exp\left(-\frac{E_i}{k_B T}\right) \approx \exp\left(-\frac{mv^2}{2k_B T}\right)$$

Using above relationship it is possible to calculate the normalized distribution for velocity:

$$n(v)dv = 4\pi\left(\frac{m}{2\pi k_B T}\right)^{3/2} v^2 \exp\left(-\frac{mv^2}{2k_B T}\right) dv$$

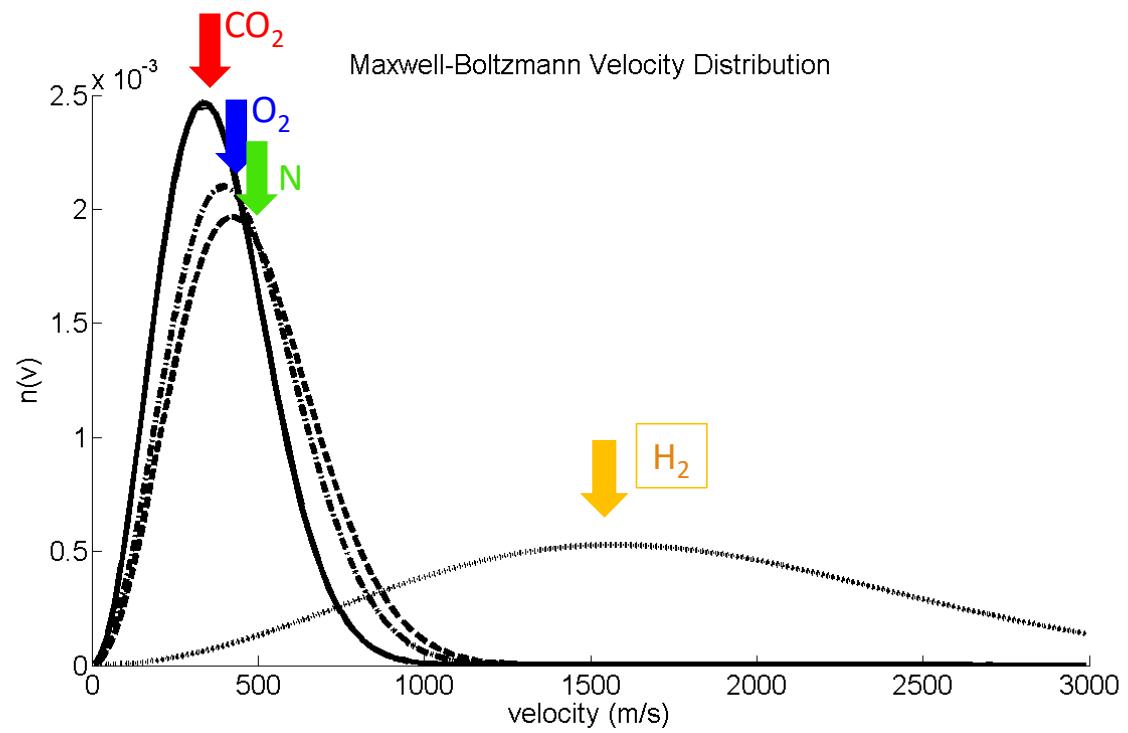
Maxwell-Boltzmann Equation

The average velocities can be calculated:

Root mean square and most probable velocity:

Maxwell-Boltzmann Distributions

The Maxwell-Boltzmann distributions vary heaviest to lightest

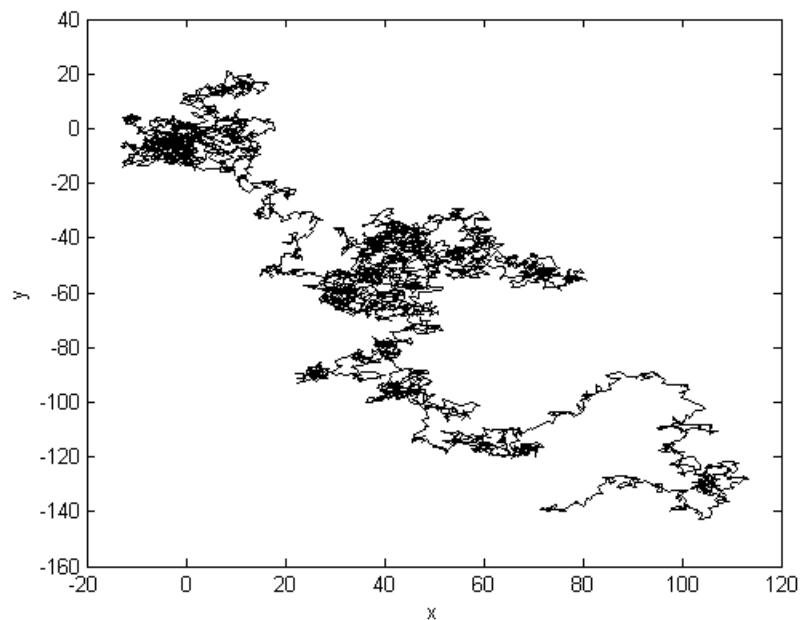


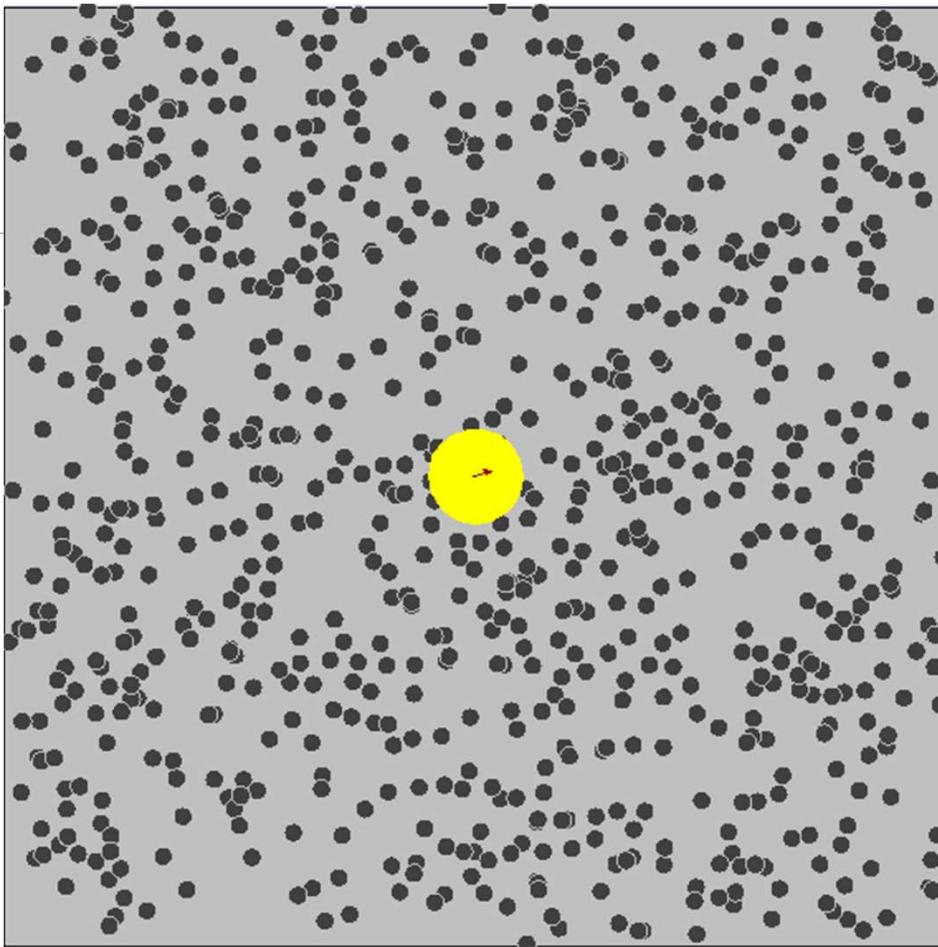
Brownian Motion

Random motion of particles suspended in a medium

Robert Brown noted in 1828 that small particles and organisms appeared to move randomly or appear to jiggle.

In 1905, Albert Einstein finally managed to find an explanation from his work on the photoelectric effect.





Diffusion

Most Brownian motion is too weak to result in diffusion.

However, when particles with very high velocity $v >> v_{rms}$ collide with small molecules, particles or organisms they appear to move in a random pattern $x(t)$.

$$R_{RMS-1D} = \left[x(t)^2 \right]^{\frac{1}{2}} \quad R_{RMS-2D} = \left[x(t)^2 + y(t)^2 \right] \\ = \sqrt{2Dt} \qquad \qquad \qquad = \sqrt{4Dt}$$

$$R_{RMS-3D} = \sqrt{6Dt}$$

Diffusion and Drag

Diffusion is determined using Einstein relation:

η is viscosity and l is the length of the object.

Can also define coefficient of drag according to viscosity:

The force is proportional to speed:

Diffusion Across Energy Membrane

If diffusion requires activation energy (E_a), such as near a barrier. The diffusion is:

$$D = D_0 \exp\left(-\frac{E_a}{k_B T}\right)$$

Diffusion and Perfusion

Blood flow model constructs a paradigm/model for perfusion

Perfusion is the delivery of blood to the capillary bed

In human it is a largely actively driven mechanism driven by the heart

Compartment models simply modeled the movement of pharmacological agents without considering mechanism

Fick's Laws

Fick's Laws were developed in 1855 to describe diffusion in a liquid. Laws are also applicable in gases and solids. They relate diffusion to concentration gradients.

Fick's 1st Law:

J Diffusion Flux/Amount of Substance $\left(\frac{mol}{m^2 s} \right)$

Fick's 2nd Law:

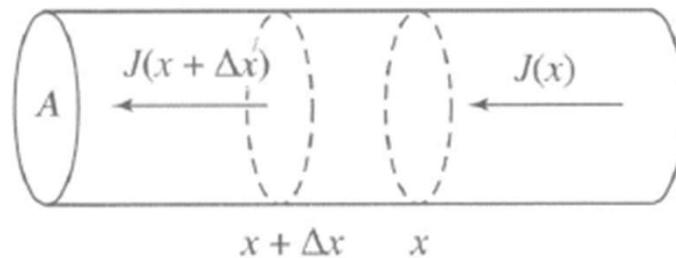
D Diffusion Constant (ie. Rate) $\left(\frac{mol}{s} \right)$

C Concentration $\left(\frac{mol}{m^3} \right)$

Fick's Second Law

Fick's second law arises from conservation of mass. Take Fick's first law and derive with respect to x :

$$\frac{dJ}{dx} = -D \frac{\partial^2 C}{\partial x^2}$$



If material diffuses, then concentration must change

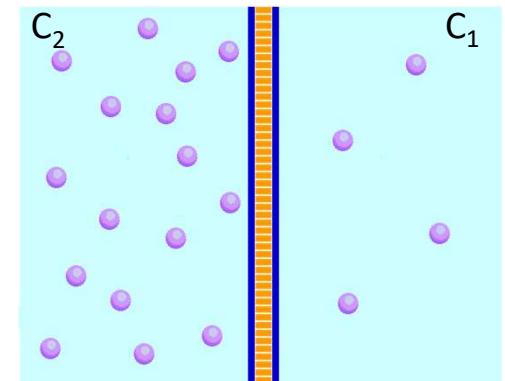
$$\frac{dJ}{dx} = -\frac{\partial C}{\partial t} \quad \text{in 3D} \rightarrow \frac{dC}{dt} = D \nabla^2 C$$

Diffusion Between Compartments

2 compartments are separated by a barrier

The rate of solute diffusion from one compartment to the other is proportional to the difference in concentration of the two compartments, C_1 and C_2

Thus in a short time interval Δt the amount of solute Q that will cross the barrier will be:



<https://year12biologyatsmc.wikispaces.com/Diffusion>

Where K is a constant that depends on the nature of the barrier (e.g. geometry, charge, pores, etc.), and the solute.

Diffusion Between Compartments (Cont.)

Therefore:

where V_1 and V_2 are the volumes of the two compartments, which are assumed constant. In the limit as Δt goes to zero:

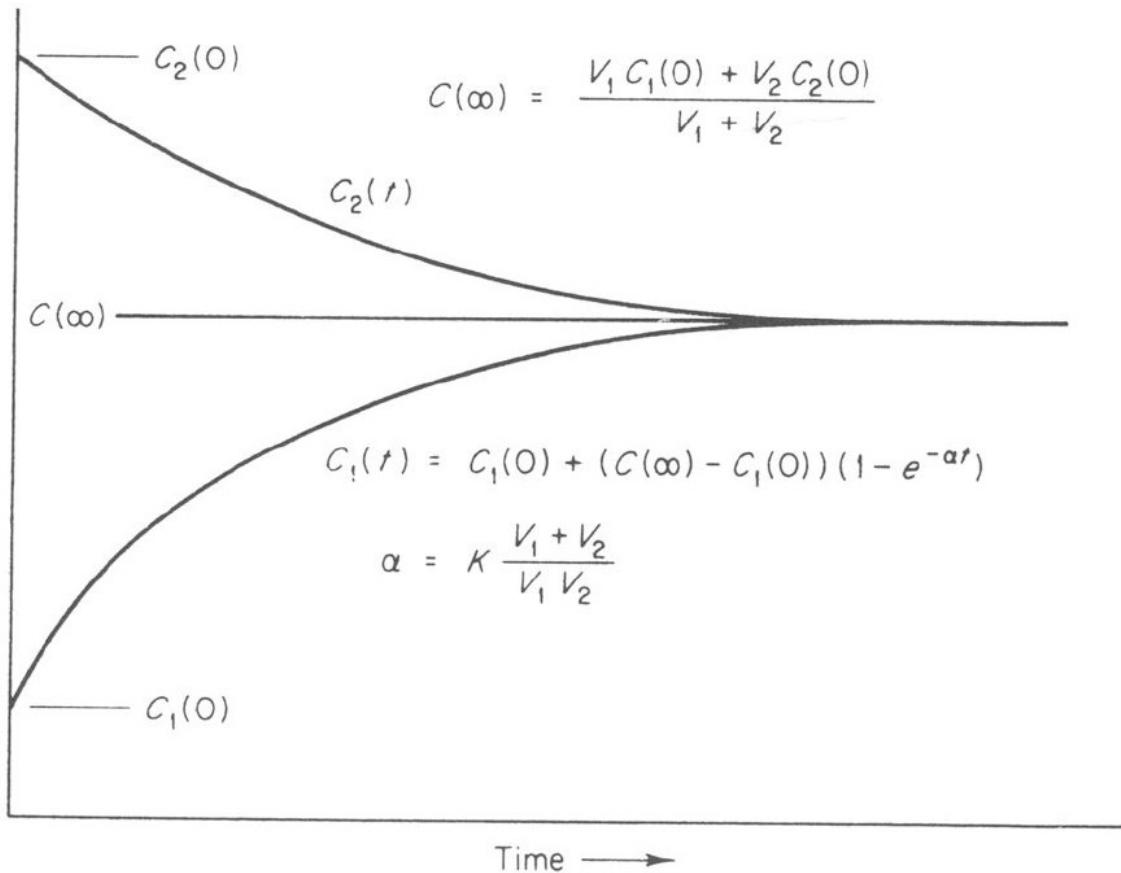
Diffusion Between Compartments (Cont.)

Let the initial concentrations in the two compartments be $C_1(0)$ and $C_2(0)$.

Thus in the initial state, the total amount of solute present is:

At some time= t an equilibrium condition is reached in which the concentrations in the two compartments are equal. i.e. the solute is distributed throughout a volume $V_1 + V_2$

Therefore the concentration at equilibrium is given by:



C(t) graph – 2 compartment

Result is independent of the nature of the diffusion process.

If the initial concentrations and volumes are known, the concentration at infinite time is also known

Diffusion Between Compartments (Cont.)

Reasonable initial educated guess of a model:

$$C_1 = C_1(0) + (1 - e^{-\alpha t}) [C(\infty) - C_1(0)]$$

$$C_2 = C_2(0) + (1 - e^{-\beta t}) [C(\infty) - C_2(0)]$$

The derivatives are:

$$\frac{\partial C_1}{\partial t} = \alpha e^{-\alpha t} [C(\infty) - C_1(0)]$$

$$\frac{\partial C_2}{\partial t} = \beta e^{-\beta t} [C(\infty) - C_2(0)]$$

Diffusion Between Compartments (Cont.)

$$C_1(t) = C_1(0) + (1 - e^{-\alpha t}) [C(\infty) - C_1(0)]$$

$$C_2(t) = C_2(0) + (1 - e^{-\beta t}) [C(\infty) - C_2(0)]$$

From previous slide. So $C_2 - C_1$, with some algebra, equals:

Diffusion Between Compartments (Cont.)

$$C_2(t) - C_1(t) = C(\infty)(e^{-\alpha t} - e^{-\beta t}) + C_2(0)e^{-\beta t} - C_1(0)e^{-\alpha t}$$

If one takes this result and inserts into previous differential equations:

$$\frac{dC_1}{dt} = \alpha e^{-\alpha t}(C(\infty) - C_1(0))$$

$$\frac{dC_2}{dt} = \beta e^{-\beta t}(C(\infty) - C_2(0))$$

Diffusion Between Compartments (Cont.)

SI 58

$$C_2(t) - C_1(t) = e^{-\alpha t}(C_2(0) - C_1(0))$$

SI 53

$$V_2 \frac{dC_2}{dt} = K(C_1 - C_2) = -V_1 \frac{dC_1}{dt}$$

Therefore:

From previous:



Multiply both sides by V_1 ,

$$\frac{dC_1}{dt} = \alpha e^{-\alpha t}(C(\infty) - C_1(0)) \quad V_1 \frac{dC_1}{dt} = \alpha V_1 e^{-\alpha t}(C(\infty) - C_1(0))$$

Diffusion Between Compartments (Cont.)

$$V_1 \frac{dC_1}{dt} = \alpha V_1 e^{-\alpha t} (C(\infty) - C_1(0)) \quad \text{SI 59}$$

Using above solve for α :

$$\alpha = \frac{K}{V_1} \frac{C_2(0) - C_1(0)}{C(\infty) - C_1(0)}$$

This can be simplified by observing that:

$$C(\infty) - C_1(0) = \frac{V_1 C_1(0) + V_2 C_2(0) - (V_1 + V_2) C_1(0)}{V_1 + V_2} = \frac{V_2 (C_2(0) - C_1(0))}{V_1 + V_2}$$

Therefore:

Diffusion Between Compartments (Cont.)

Back to these ugly equations:

$$C_1 = C_1(0) + (1 - e^{-\alpha t}) [C(\infty) - C_1(0)]$$

$$C_2 = C_2(0) + (1 - e^{-\beta t}) [C(\infty) - C_2(0)]$$

put into a nicer form by replacing $C(\infty)$ - $C_1(0)$ by its value from:

$$C(\infty) - C_1(0) = \frac{V_1 C_1(0) + V_2 C_2(0) - (V_1 + V_2) C_1(0)}{V_1 + V_2} = \frac{V_2 (C_2(0) - C_1(0))}{V_1 + V_2}$$

Diffusion Between Compartments (Cont.)

$$C_1 = C_1(0) + (1 - e^{-\alpha t}) [C(\infty) - C_1(0)]$$

$$C_2 = C_2(0) + (1 - e^{-\beta t}) [C(\infty) - C_2(0)]$$

Also, replace $C(\infty)$ - $C_2(0)$ by the analogous form:

$$C_1(t) = C_1(0) + (1 - e^{-\alpha t}) [C_2(0) - C_1(0)] \frac{V_2}{V_1 + V_2}$$

$$C_2(t) = C_2(0) + (1 - e^{-\beta t}) [C_1(0) - C_2(0)] \frac{V_1}{V_1 + V_2}$$

Recap of all this math:

1. Relate the derivatives of concentration of each side to each other
2. Model the equilibrium concentration as a function of the concentration of both sides at time zero
3. Create the model for the equilibrium concentration in terms of the exponentials of each side
4. Link everything together by assume alpha = beta
5. Solve for alpha and use 2) to have alpha in terms of volumes and initial concentrations
6. Model C₁ and C₂ in terms of initial concentrations, and volumes

$$V_2 \frac{dC_2}{dt} = K(C_1 - C_2) = -V_1 \frac{dC_1}{dt}$$

$$C(\infty) = \frac{V_1 C_1(0) + V_2 C_2(0)}{V_1 + V_2}$$

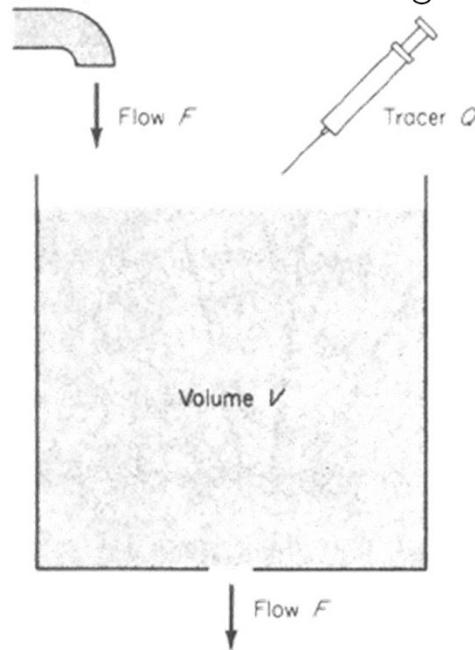
$$C_2(t) - C_1(t) = C(\infty)(e^{-\alpha t} - e^{-\beta t}) + C_2(0)e^{-\beta t} - C_1(0)e^{-\alpha t}$$

$$C_2(t) - C_1(t) = e^{-\alpha t}(C_2(0) - C_1(0))$$

$$\alpha = \frac{K(V_1 + V_2)}{V_1 V_2}$$

Model #1: The basic 1 compartment continuous dilution process

The idealized one-compartment dilution problem consists of a single continuously mixed chamber through which a fluid is flowing at a constant rate.



Say:

$V = 10 \text{ Litres (constant)}$

$F_{\text{in}} = 2 \text{ Litres/min}$

At $t(0)$ 3 Moles of dye is injected

- assume dye is injected instantly
- assume dye is instantly mixed

Therefore at $t=0$, concentration = 0.3 moles/liter

Model 1

- 1) The concentration of dye immediately begins to fall (i.e. being lost in the effluent while water is replacing the lost volume of solution).
- 2) At very long times, which we designate $t(\infty)$, all of the dye has been flushed out.
- 3) To find the concentration of dye at any time following its injection, we must calculate instantaneous description of a continuous process.

Note minus sign indicates
[tracer] is decreasing

Model 1

ΔQ is a quantity, not a concentration:

We're interested in concentration at time=t (i.e. $C(t)$). So, divide both sides of the equation by Δt and take the limit as $\Delta t \rightarrow 0$

Therefore:

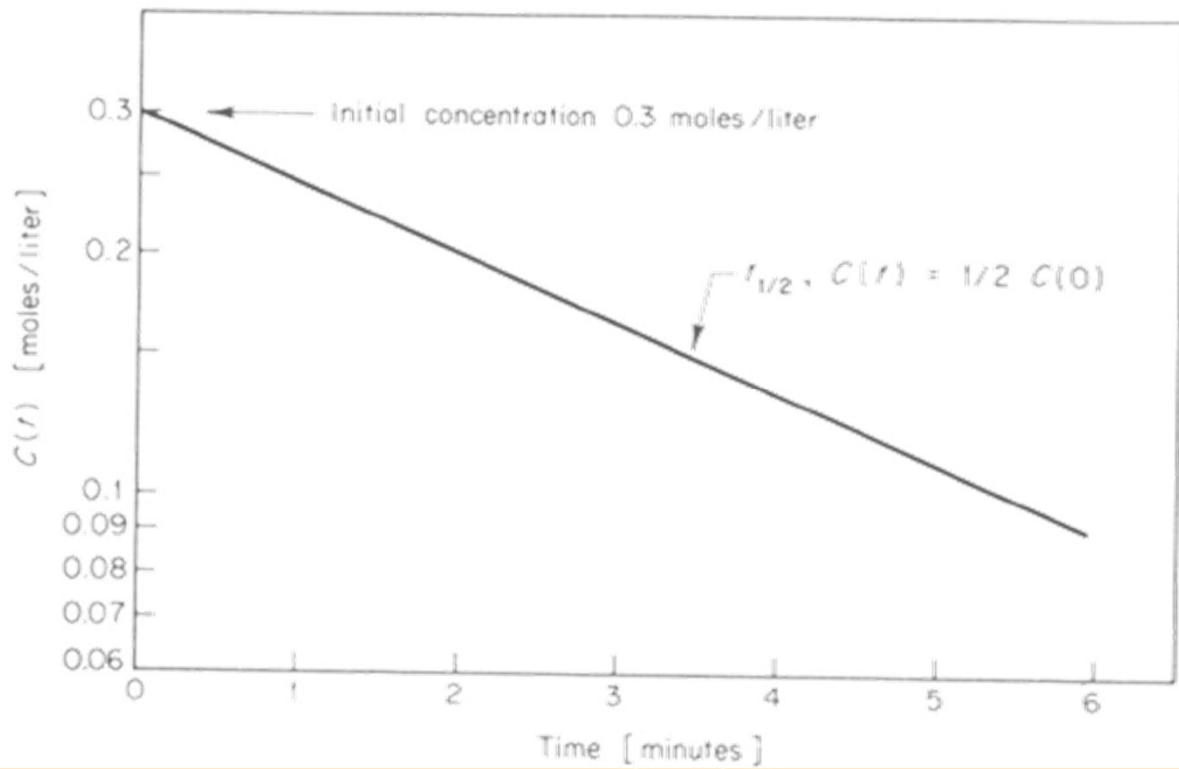
Model 1

Check units:

$$\frac{\text{moles}}{\text{liter}} = \frac{\text{moles}}{\text{liter}}$$

Exponent units:

$$\frac{F}{V} t = \frac{\text{liters}}{\frac{\text{min}}{\text{liters}}} \text{ min}$$



Model #2: One compartment with metabolic turnover

- same mathematical form as dilution process.
- 1) To this pool is added a small quantity $q(0)$ moles of a metabolite.
- 2) The rate of disappearance of the metabolite is analogous to the rate of disappearance of the tracer in the one-compartment dilution process.
- 3) Here the total pool ($Q + q$) is analogous to volume [model 1] and rate of metabolite turnover is analogous to flow [model 1].
- assume addition of the metabolite does not change it's rate of utilization.

Model 2

Since $q(t) \ll Q$ we can make the approximation:

Therefore:



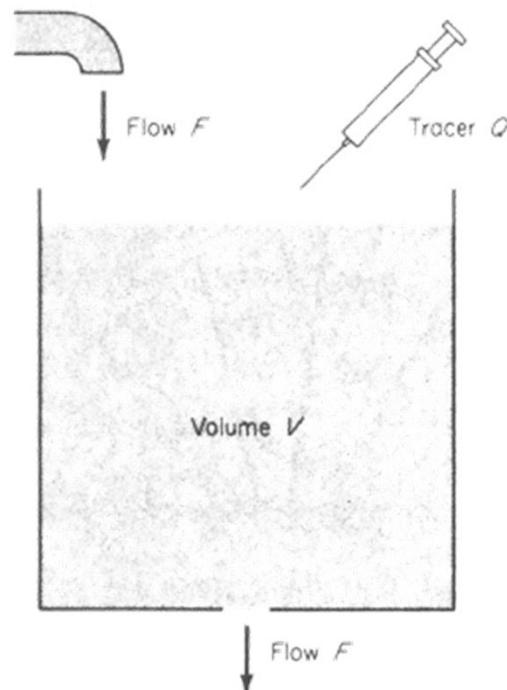
R = metabolism of metabolite
(moles/sec)
Q = moles

Model #3: Radioactive Decay

Again, same mathematical form is given by the decay of a radioactive isotope.

Here, # of atoms that decay per unit time is proportional to the number of remaining radioactive atoms:

Model #4: One compartment with Constant Injection



Constant Injection
rate = R moles/sec

As previously, amount of tracer lost from the tank in a small unit of time Δt is:

Model 4

- but at the same time tracer is entering the tank at a rate R
- Thus, in the interval Δt the net change in quantity of tracer in the tank is given by:

- with same logic as previously (i.e. Q is moles and want a concentration, so divide by volume):

Model 4

- divide by Δt and take the limit as $\Delta t \rightarrow 0$

$$\frac{\Delta C}{\Delta t} = \frac{R - FC(t)}{V} \quad \longrightarrow$$

Units Check:

$$\frac{\text{moles}}{\text{liter sec}} = \frac{\text{moles}}{\text{sec liters}} - \frac{\text{liter}}{\text{sec liters}} \frac{\text{moles}}{\text{liter}}$$

Solution:



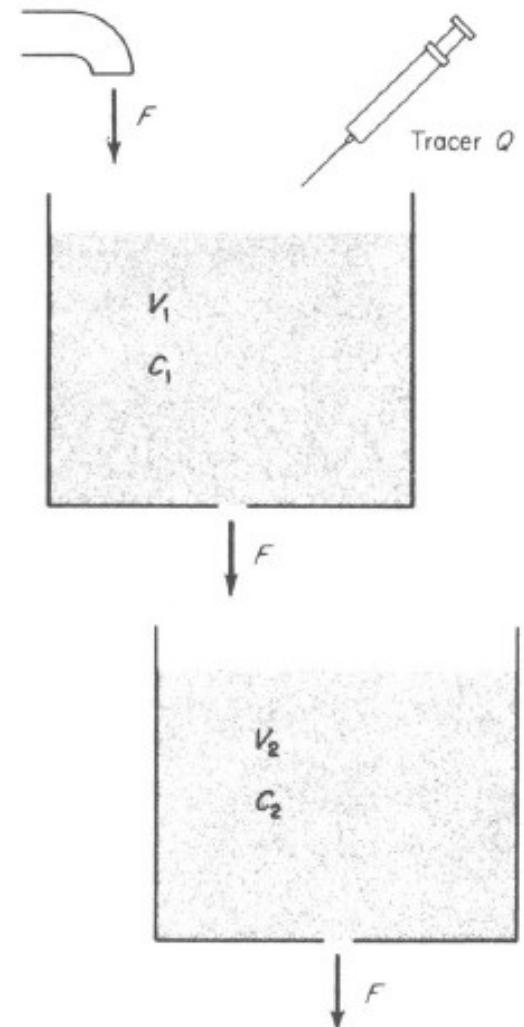
Model 5

- 1) First compartment is identical to Model 1.

$$C_1(t) = C_1(0)e^{-(F/V_1) \cdot t}$$

- 2) Second compartment (V_2), is filled by the effluent from the first compartment

- 3) V_2 is a flow-through-type compartment, losing solution at a rate F .



Model 5

- 1) Need to calculate concentration (C_2) in compartment 2 as a function of time
- 2) find total amount of tracer that enters and leaves the compartment in a small interval of time Δt .
- 3) Amount leaving will be the rate of flow F multiplied by the concentration in 2nd compartment (C_2) multiplied by a small time interval Δt .
- 4) Amount entering during that time will be flow rate, $F \times \Delta t \times C_1$.

Model 5

Net change in tracer in compartment 2 is:

this is converted to a change of concentration
by dividing by the volume, V_2 :

Again divide by Δt and take the limit as $\Delta t \rightarrow 0$:

$$\Delta Q_2 = -\Delta t C_2(t)F + \Delta t C_1(t)F$$

$$\Delta C_2 = \frac{F}{V_2} C_1(t) \Delta t - C_2(t) \frac{F}{V_2} \Delta t$$

$$\frac{dC_2}{dt} = \frac{F}{V_2} [C_1(t) - C_2(t)]$$

$$= \frac{F}{V_2} [C_1(0)e^{-(F/V_1)t} - C_2(t)]$$

Solution:

$$C_2(t) = \frac{V_1 C_1(0)}{V_1 - V_2} [e^{-(F/V_1)t} - e^{-(F/V_2)t}]$$

Model 5

This is a complicated problem!

- usually wish to visualize it by plotting $C_2(t)$ as a function of time. BUT
- confronted with the fact that V_1 V_2 C_1 and F can all take on wide ranges of values in practical problems
- making plots for wide ranges of four different parameters is prohibitive.
- therefore examine function to see if it is possible to reduce number of parameters

$$C_2(t) = \frac{V_1 C_1(0)}{V_1 - V_2} \left[e^{-(F/V_1) \cdot t} - e^{-(F/V_2) \cdot t} \right]$$

Model 5

F*t?

$$C_2(t) = \frac{V_1 C_1(0)}{V_1 - V_2} \left[e^{-(F/V_1) \cdot t} - e^{-(F/V_2) \cdot t} \right]$$

Exponents?

Model 5

$$C_2(t) = \frac{V_1 C_1(0)}{V_1 - V_2} [e^{-p} - e^{-Kp}]$$

- It would be nice to express the leading fraction in terms of the same dimensionless quantities.
- easily done by dividing through by V_1 and thus we find :

Where: $K = \frac{V_1}{V_2}$

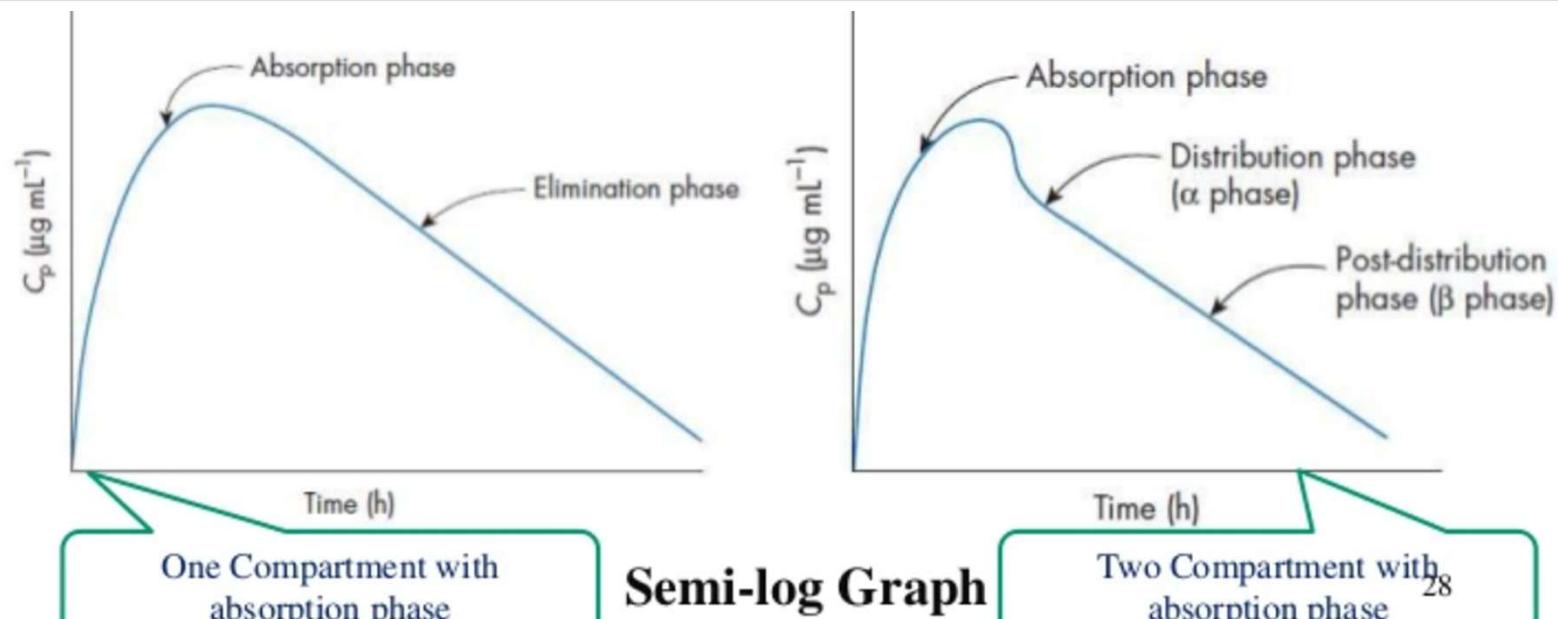
- This function, which is a ratio of concentrations and is therefore dimensionless, is expressed in terms of only two dimensionless variables p and K .

Model 5

$$\frac{C_2(t)}{C_1(0)} = \frac{1}{1 - K^{-1}} \left[e^{-p} - e^{-Kp} \right] \quad \text{Where: } K = \frac{V_1}{V_2}$$

Thus a problem that initially had 4 independent variables is reduced to one of only 2 independent dimensionless variables!

Model 5



Anas Bahnassi PhD 2011

$$V_1 > V_2$$

$$V_1 < V_2$$

Getting to more complicated Models

Homogeneous and Inhomogeneous Equations

When the dependent variable appears exactly once in every term of the equation, the equation is said to be homogeneous:

$$\frac{d^2C(t)}{dt^2} + t \frac{dC(t)}{dt} + 4C(t) = 0 \quad \sin \omega t \frac{dC(t)}{dt} + C(t) \cos \omega t = 0$$

If there are terms that do not contain the dependent variable,
the equations are called inhomogeneous:

$$\frac{dC(t)}{dt} + C(t) = 4 \quad \frac{d^2C(t)}{dt^2} + \frac{dC(t)}{dt} + t = 0$$

Homogeneous and Inhomogeneous Equations

Linear homogeneous equations with constant coefficients have a solution that can be expressed as a sum of exponentials

The number of such exponential terms required for a solution is the same as the order of the differential equation:

$$\frac{d^2C(t)}{dt^2} + B \frac{dC(t)}{dt} + DC(t) = 0$$

This is a 2nd order linear homogeneous differential equation with solution:

where the constants A1 and A2 are arbitrary and α_1 and α_2 are solutions of the quadratic algebraic equation:

Homogeneous and Inhomogeneous Equations

Now consider a similar differential equation but with the addition of an inhomogeneous term:

$$\frac{d^2C(t)}{dt^2} + B \frac{dC(t)}{dt} + DC(t) = 4$$

The solution is given by:

The solution of an inhomogeneous linear equation will always consist of the solution to the homogeneous equation plus the addition of a constant or a function of the independent variable.

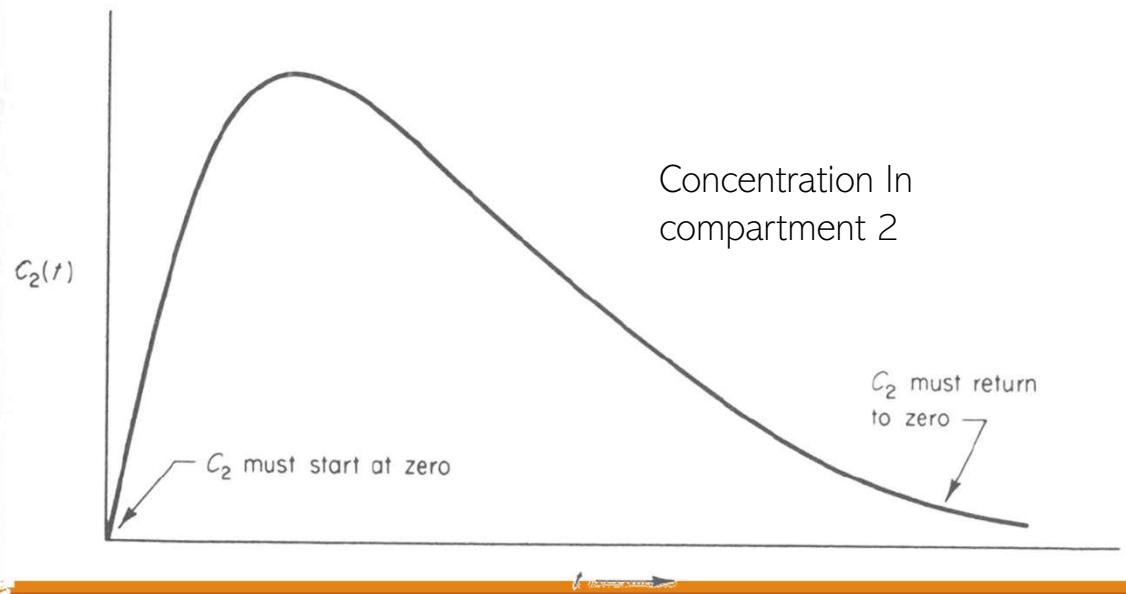
The arbitrary constants A1 and A2 cannot be determined from the differential equation and must be determined by other means.

Let's go back to the 2 compartment series dilution:

$$\frac{dC_2}{dt} = \frac{F}{V_2} [C_1(t) - C_2(t)] = \frac{F}{V_2} [e^{-(F/V_1)t} - e^{-(F/V_2)t}]$$

This is an example of an inhomogeneous linear differential equation

- cannot be a single exponential, since a single exponential cannot both start and end at 0
- sum of 2 exponentials?



Let's go back to the 2 compartment series dilution:

$$C_2(t) = A_1 \exp(-\alpha_1 t) + A_2 \exp(-\alpha_2 t)$$

2 conditions must be true:

- 1) Both exponentials must have negative exponents, since a positive exponent would imply a concentration that becomes infinite at infinite time.
- 2) The constants A_1 and A_2 must be equal to each other but of opposite signs, so that their sum will be zero at zero time.

Thus, this must be of the form:

Differentiate:

$$\frac{dC_2}{dt} = A \left[\exp(-\alpha_1 t) - \exp(-\alpha_2 t) \right]$$

From previous:

$$\frac{dC_2}{dt} = \frac{F}{V_2} [C_1(t) - C_2(t)] = \frac{F}{V_2} [C_1(0)e^{-(F/V_1)t} - C_2(t)]$$

$$A\alpha_2 e^{-\alpha_2 t} - A\alpha_1 e^{-\alpha_1 t} = \frac{F}{V_2} C_1(0) e^{-(F/V_1)t} - \frac{F}{V_2} A e^{-\alpha_1 t} - \frac{F}{V_2} A e^{-\alpha_2 t}$$

Or:

$$A \left(\frac{F}{V_2} - \alpha_1 \right) e^{-\alpha_1 t} + A \left(\alpha_2 - \frac{F}{V_2} \right) e^{-\alpha_2 t} = \frac{F}{V_2} C_1(0) e^{-(F/V_1)t}$$

2 compartment model cont

$$\alpha_1 = \frac{F}{V_2}, \quad \alpha_2 = \frac{F}{V_1}$$

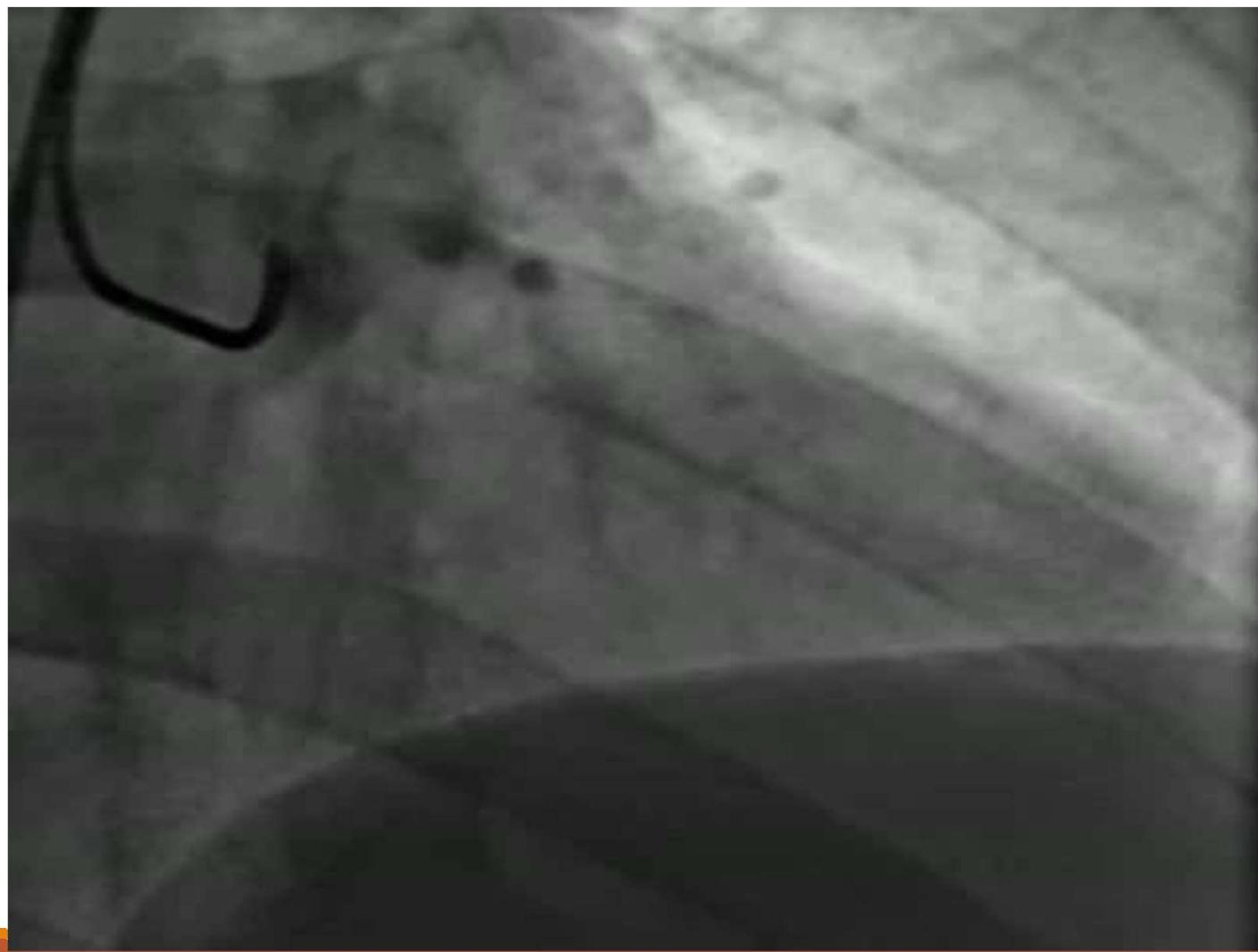
Based on mathematics of exponentials either α_1 , or α_2 must equal F/V_2 and the other α must be F/V_1

$$A = \frac{(F/V_2)C_1(0)}{(F/V_1) - (F/V_2)}$$

Therefore our model can be derived 2 ways

$$C_2(t) = \frac{V_1}{V_1 - V_2} C_1(0) (e^{-(F/V_1)t} - e^{-(F/V_2)t})$$

Imaging and Pharmacokinetic Modeling of Tracers



GE MEDICAL SYSTEMS
SIGNA HDx SJMROCO
Ex: 3197
Se: 1000
Im: 1+C
COL OCor A 48.6
DFOV 14.0cm

DT:1.00
Ph:1/17

ET:1

R
S
A

SAL

St Josephs Hosp
tricks
M47Y
AW1188584572.301.1177100917
Apr 18 2007
02:08:45 PM
Mag = 1.00

FL:
ROT:

3D/TRICKS/30
TR:8.1
TE:3.1/Fr
EC:1 /1 31.2kHz

HRWRIST
FOV:14x9.8/W
1.4thk/-0.7ov
408/01:56
320X192/0.75 NEX
MP/Z2

IPR

WW: 4845 WL: 3520

Modeling Imaging Tracers

1). Dynamic susceptibility weighted MRI (dscMRI)

- See decreased signal intensity

2). Dynamic contrast enhanced MRI (dceMRI)

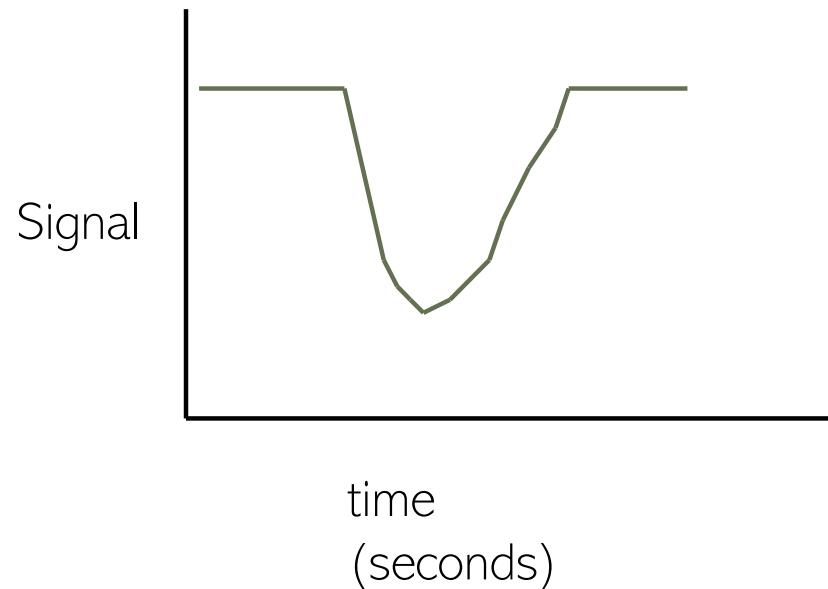
- see increased signal intensity

Using MRI these are almost exclusively done with Gadolinium (Gd) contrast agents

APPLICATIONS:

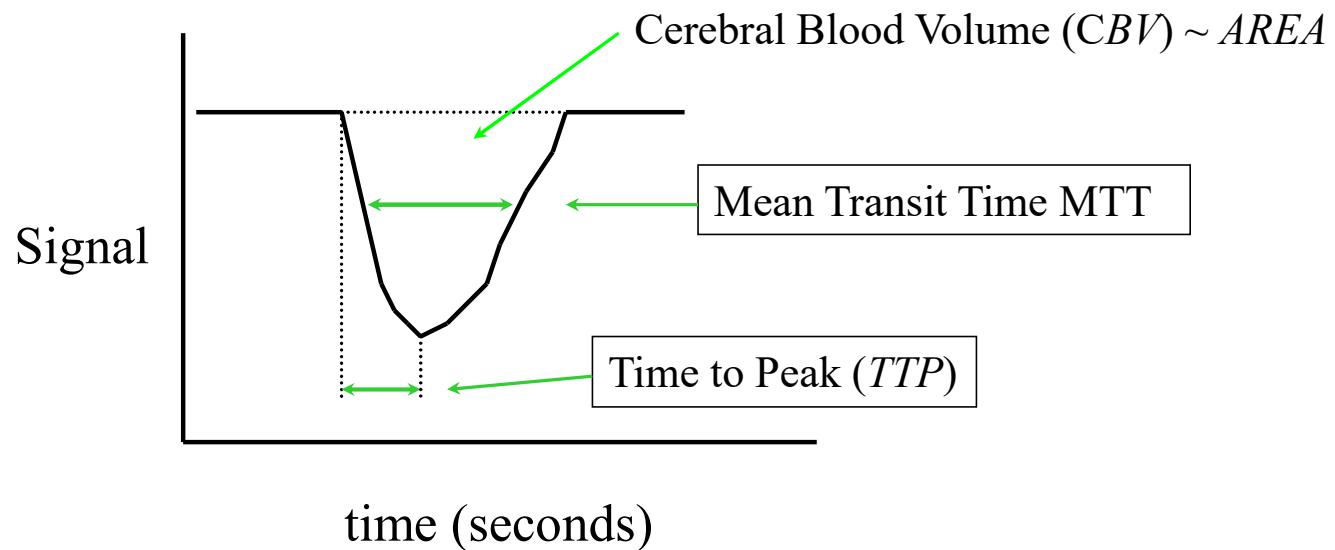
- stroke assessment
- tumour vascularity
- many other possibilities

MRI Model #1: dscMRI

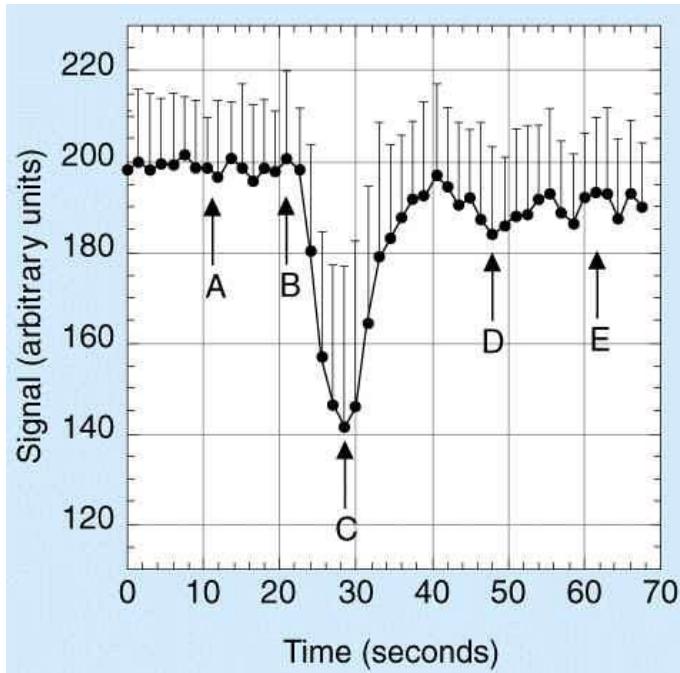


MRI Model #1: dscMRI

- Looking at 'raw data' of signal vs. time
- Want to model blood flow and blood volume
- This approach is most often for brain.
- Cerebral blood flow (CBF) and cerebral blood volume (CBV)



Sample Brain Data:



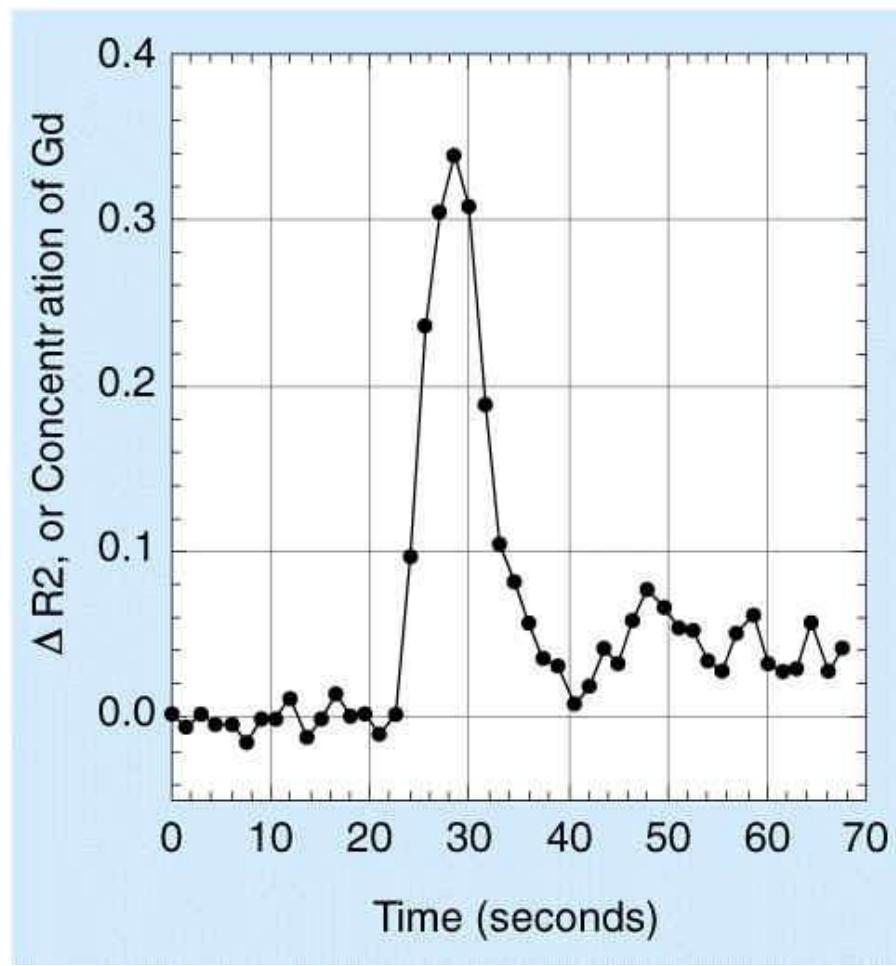
A = baseline

B = bolus time of arrival

C = Time to Peak (TTP)

D = second pass

E = third pass



$$\Delta R_2 = \frac{-\ln(S / S_o)}{TE}$$

$$C_T(t) = \Delta R_2(\text{tissue})$$

$$C_A(t) = \Delta R_2(\text{artery})$$

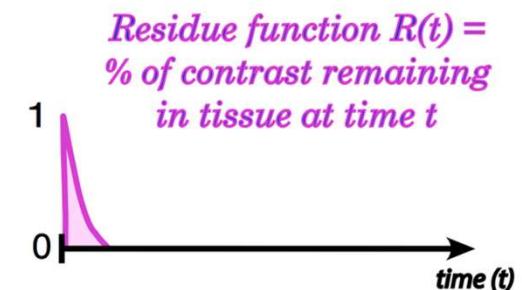
$$CBV = \frac{\int_0^{\infty} C_T(t) dt}{\int_0^{\infty} C_A(t) dt}$$

Note: TE = MRI parameter called “echo time”; set by MRI operator

Arterial Input Function (AIF):

The linear system here is the tissue capillary bed

- 1). Tissue response to a Dirac delta function $\delta(t)$ at $t=0$.
- 2). After injection, see dispersion of the bolus within tissue and a range of particle transit times.
- 3). The fraction of injected particles remaining at time= t after impulse injection is the residue (dimensionless function $R(t)$).
- 4). Immediately after injection $R(t)$ is maximal with value $R(0) = 1$, then $R(t)$ decreases

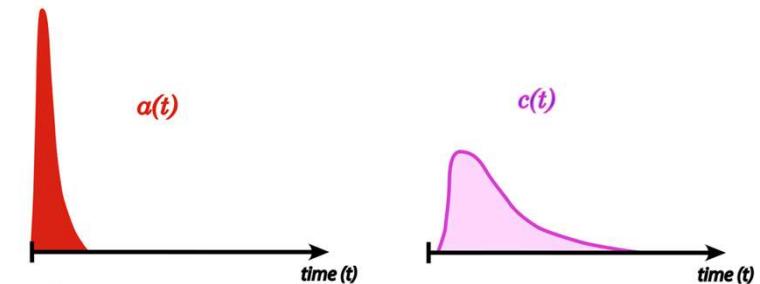


Arterial Input Function (AIF):

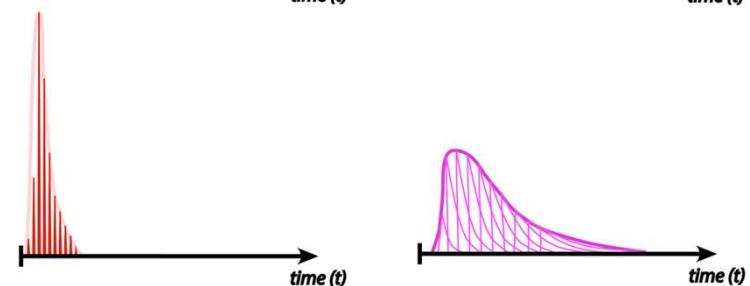
- Width of $R(t)$ reflects the distribution of particle transit times through a tissue.
 - area under $R(t)$ reflects the average time a particle spends transitioning through a tissue vascular bed (mean transit time, MTT):
-
- Instantaneous bolus is only hypothetical. The arterial input function, $CA(t)$, is thus always temporally dispersed.
 - broad $CA(t)$ can be represented as a set of Dirac delta functions at different time delays (τ), each producing an independent response.

Model vs Real life AIF

a real-life arterial input function (AIF)
with corresponding resultant tissue
concentration curve $C_T(t)$



Modeling the tissue response by a set of
Dirac delta functions



CBF (FT) requires AIF!

Blood Flow (CBF)

- require AIF (arterial input function)

$$C_T(t) = F_T \cdot C_A(t) \otimes R(t) = F_T \int_0^t C_A(\tau) R(t-\tau) d\tau$$

$C_T(t)$ = [Gd] in tissue at time = t

$C_A(t)$ = arterial [Gd] concentration at time = t

$R(t)$ = vascular residue function

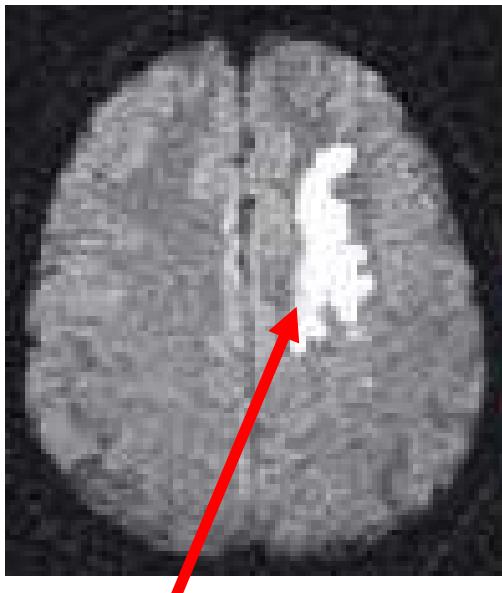
F_T = tissue blood flow

NOTE: Some versions of this equation contain a constant (ρ/h) preceding FT, where ρ = tissue density (g/mL) and h = fraction accounting for difference in hematocrit between capillaries and larger vessels.

CBF Model Assumptions...

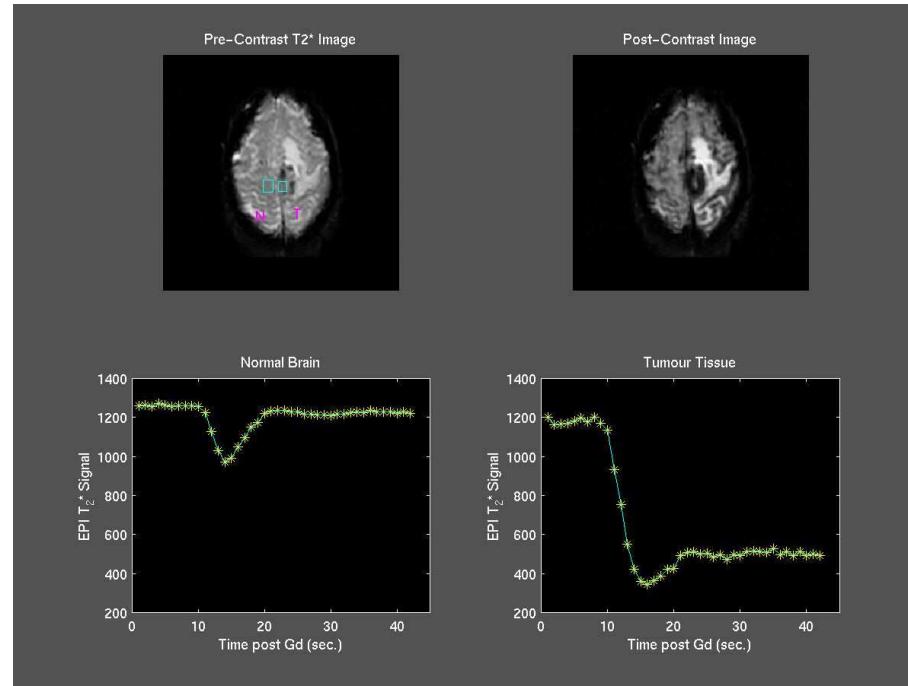
- 1) Cerebral blood flow (CBF) and cerebral blood volume (CBV) remain constant during the measurement period.
- 2) All injected tracer molecules eventually leave the system (i.e. there is nothing that stays around in the tissue for great lengths of time, lost within the tissue).
- 3) The system has a linear response to inputs

MRI Diffusion-Perfusion mismatch: Acute ischemic stroke

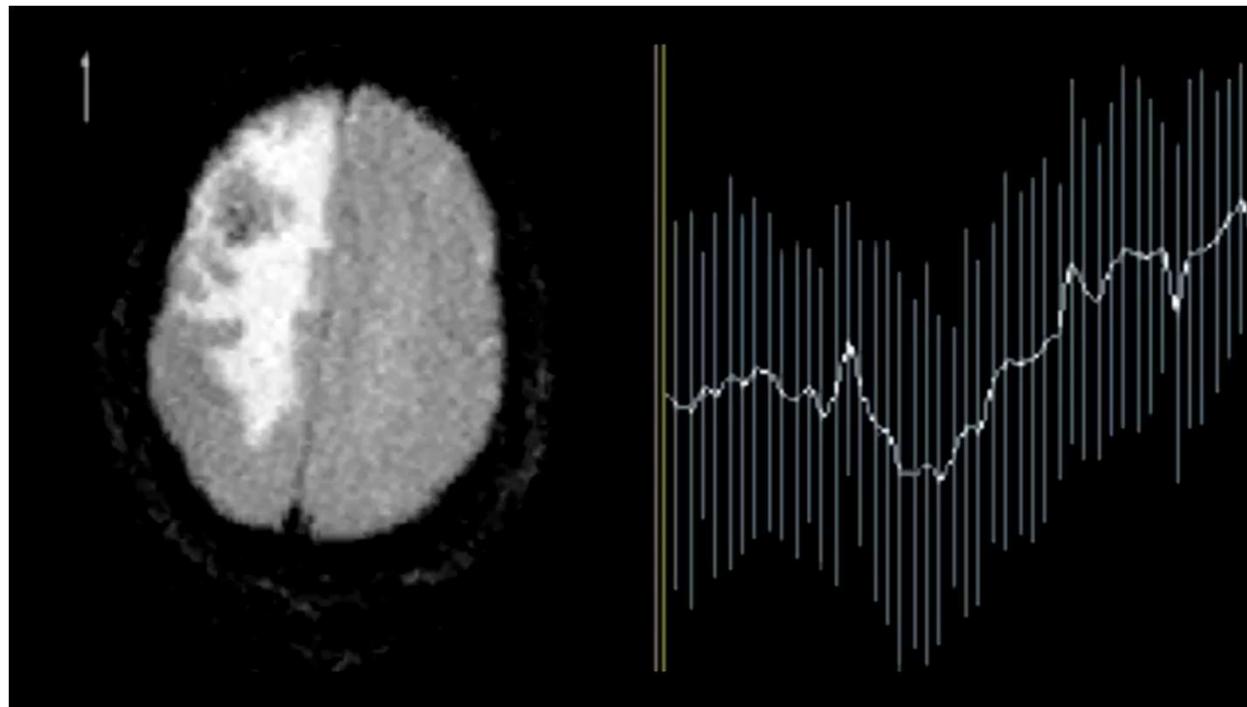


Diffusion (DWI) abnormality < perfusion (CBF) abnormality
= ischemic penumbra (risk of infarction)

Brain Tumour: Example of FAILED model!

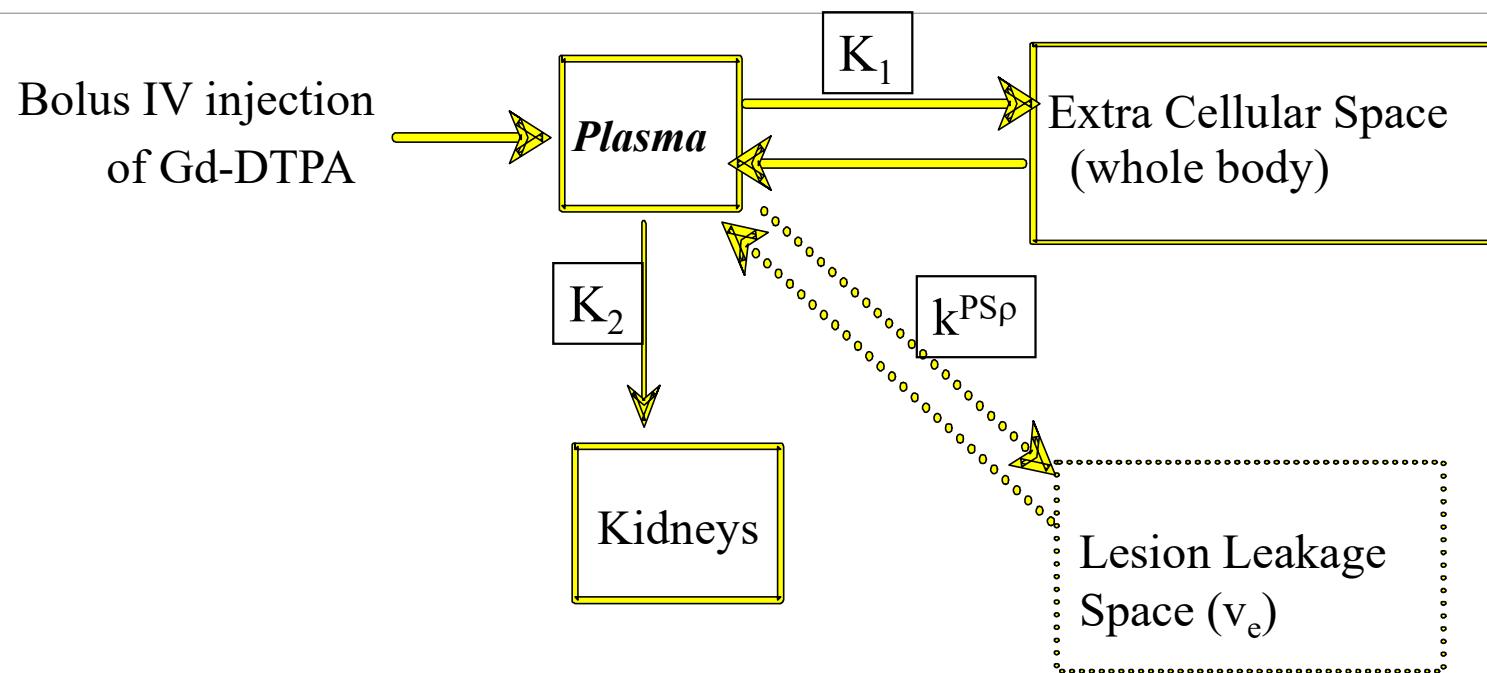


BBB Breakdown: FAILED model!



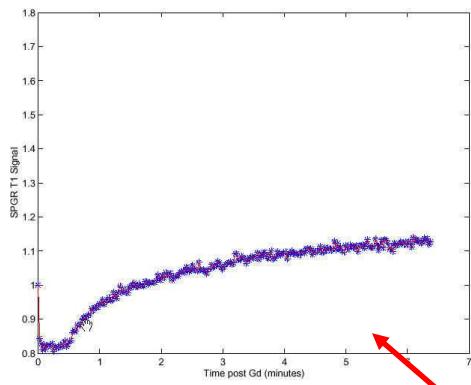
Tissue $\Delta R_2 \pm \text{std}$

Pharmacokinetic Model:

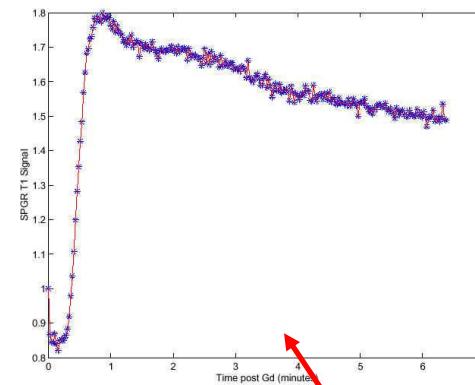


MRI Model #2: dceMRI

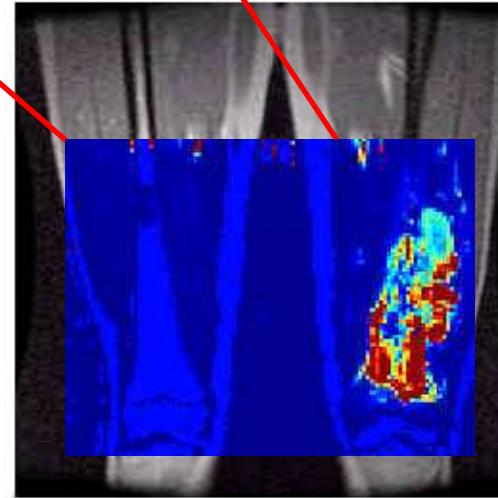




Initial SPGR image



SPGR at 50 seconds



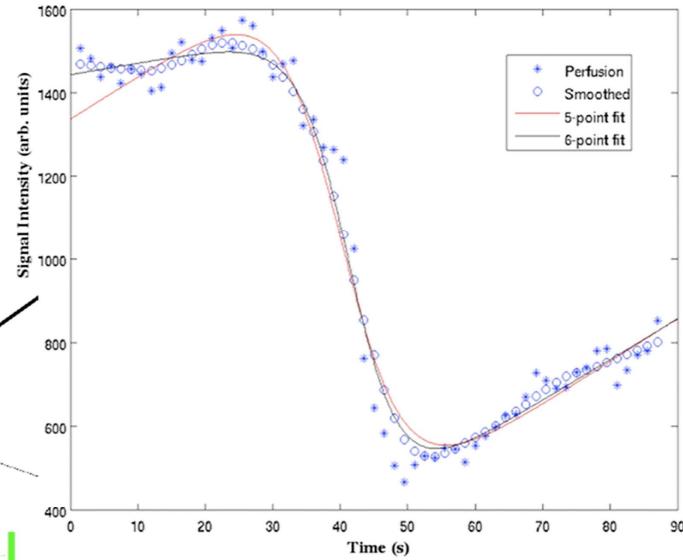
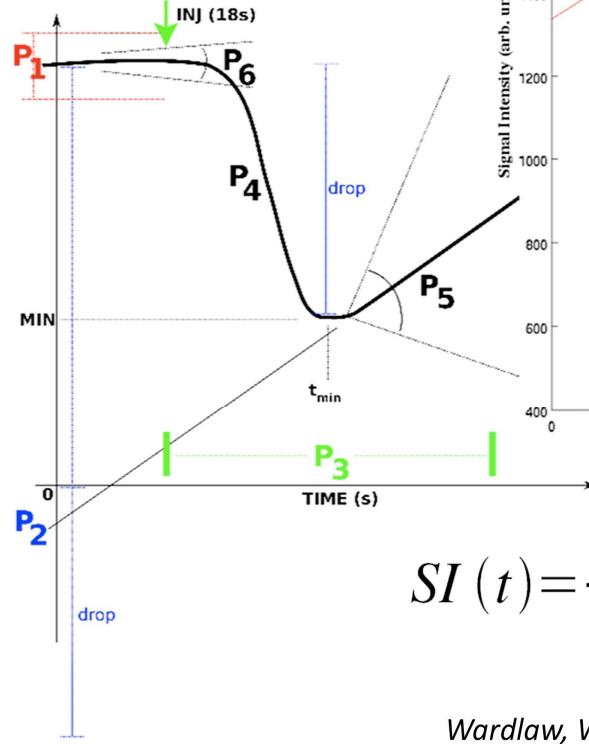
Caveats of Dynamic Contrast Enhanced MRI

Need an IV injection of contrast agent (typically Gadolinium based). A potential problem with patients who have reduced GFR and shouldn't be injected with this

Mathematical modeling requires fast injection (3-5cc per second). This is a problem for children, chemotherapy patients, and others.

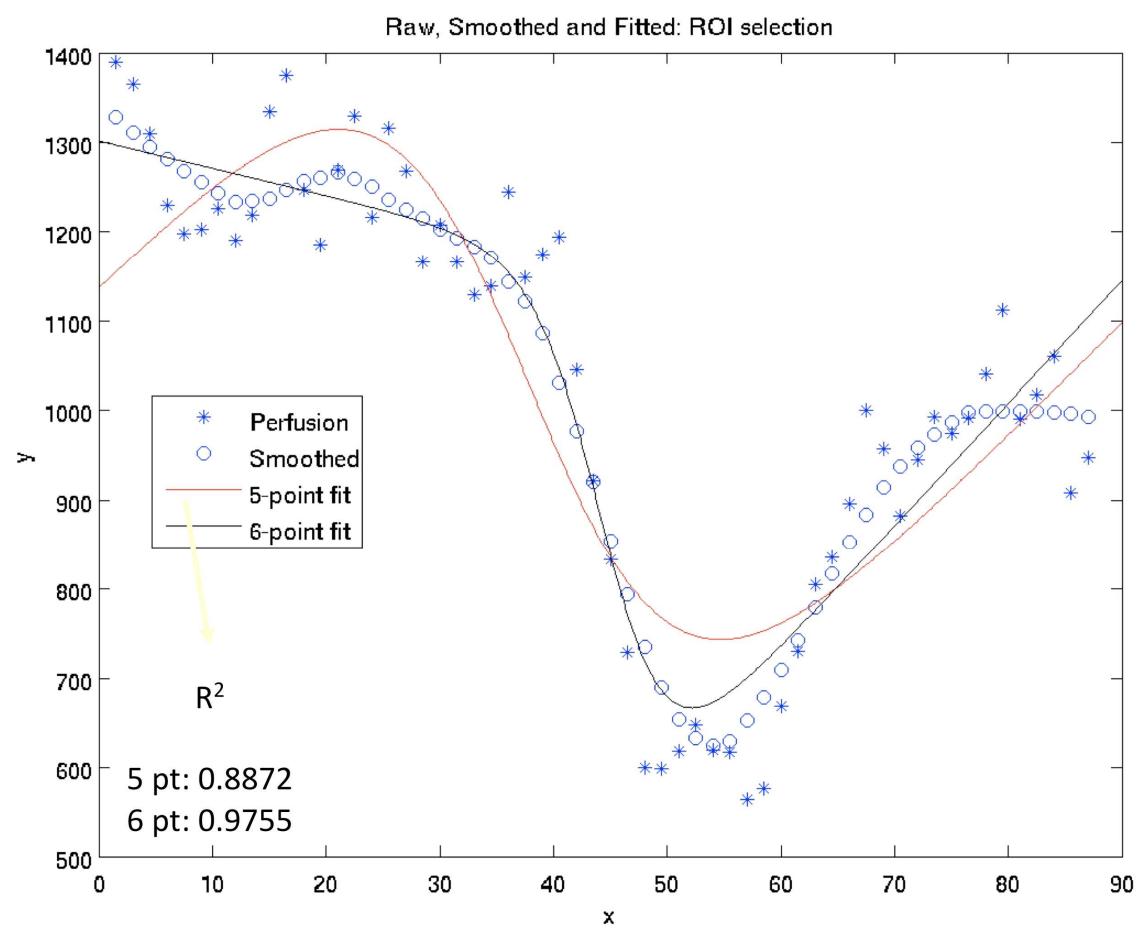
What is the appropriate mathematical model? What should be the 'initial conditions' of the model?

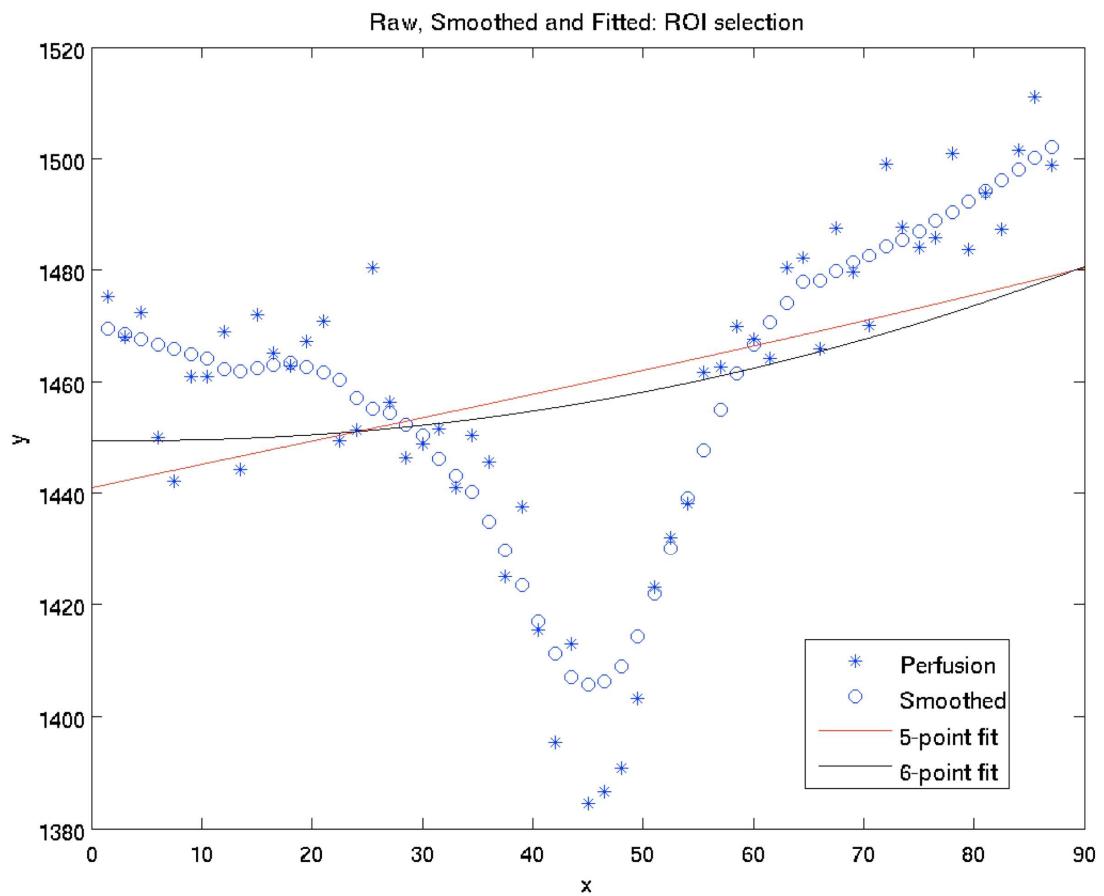
T2* Logistic Model



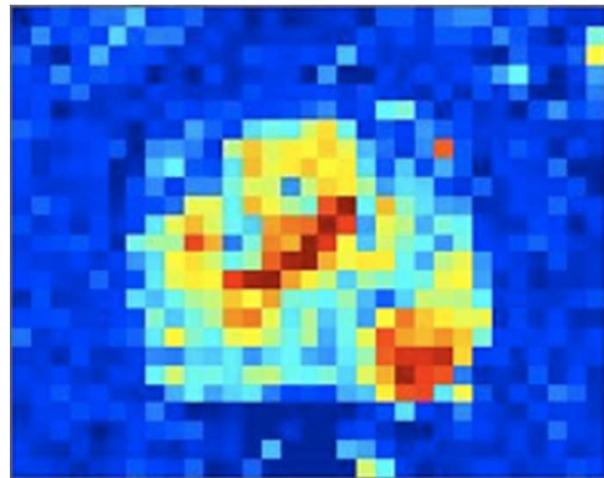
$$SI(t) = \frac{P_1 + P_6 t}{1 + e^{P_4(t - P_3)}} + (P_2 + P_5 t)$$

Wardlaw, Wong, Noseworthy (2009) Submitted



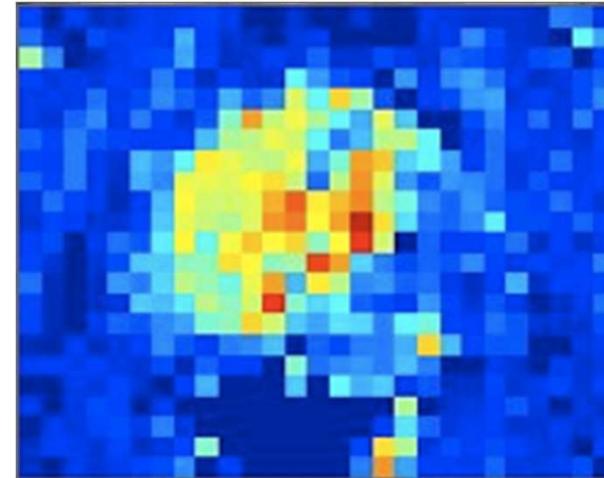


dceMRI of Prostate Cancer



Perfusion:

- Appropriate Model
- hypervasculär peripheral zone



Perfusion:

- inappropriate model
- now what ?

Tensor Math

Tensors are a mathematical construct that allows for the simple representation of vectors and directional fields. In some sense, tensors can be thought of as matrices.

For example a scalar for temperature can be a real number.

	Normal Notation	Tensor Notation
Scalar:	a	a

Vectors are directional values.

	Normal Notation	Tensor Notation
Vector:	$v_x i + v_y j + v_z k$	\vec{v}

When are Tensors Necessary?

Consider Fick's law in elementary form:

$$J = -D \frac{\partial C}{\partial x}$$

- flux, J , of a diffusing species flows opposite the gradient in concentration
- magnitude of the flux is proportional to the steepness of the gradient.
- adequate for diffusion across a slab or membrane

$$\vec{J} = -D \vec{\nabla} C$$

$$\vec{\nabla} = (\partial / \partial x, \partial / \partial y, \partial / \partial z)$$

Diffusion in 3d

$$J_x = -D \frac{\partial C}{\partial x}$$

$$J_y = -D \frac{\partial C}{\partial y}$$

$$J_z = -D \frac{\partial C}{\partial z}$$

This can be used to describe a great number of situations as long as the material in question is isotropic.

$$J_x = -D_{xx} \frac{\partial C}{\partial x} - D_{xy} \frac{\partial C}{\partial y} - D_{xz} \frac{\partial C}{\partial z}$$

$$J_y = -D_{yx} \frac{\partial C}{\partial x} - D_{yy} \frac{\partial C}{\partial y} - D_{yz} \frac{\partial C}{\partial z}.$$

$$J_z = -D_{zx} \frac{\partial C}{\partial x} - D_{zy} \frac{\partial C}{\partial y} - D_{zz} \frac{\partial C}{\partial z}$$

$$\begin{bmatrix} J_x \\ J_y \\ J_z \end{bmatrix} = - \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix} \begin{bmatrix} \partial C / \partial x \\ \partial C / \partial y \\ \partial C / \partial z \end{bmatrix}$$



Dan Fleisch

Author of *A Student's Guide to Vectors and Tensors*

<https://youtu.be/f5liqUk0ZTw/>

Lecture 6

TAYLOR DEVET MASC.

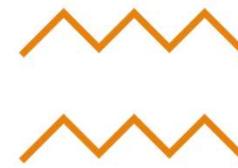
PHD. CANDIDATE BIOLOGICAL AND BIOMEDICAL ENGINEERING
MCGILL UNIVERSITY

SHRINERS HOSPITAL FOR CHILDREN

Todays Aims...



PCA



ICA

Too much data, no idea where to start?

```
>> load cities
```

9 different indicators of the quality of life in 329 U.S. cities.

For each category, a higher rating is better.

represent a multivariate data table as smaller set of variables (summary indices)

	Climate	Housing	Health	Crime	Transportation	Education	Arts	Recreation	Economics
Abilene, TX	521	6200	237	923	4031	2757	996	1405	7633
Akron, OH	575	8138	1656	886	4883	2438	5564	2632	4350
Albany, GA	468	7339	618	970	2531	2560	237	859	5250
Albany-Troy, NY	476	7908	1431	610	6883	3399	4655	1617	5864
Albuquerque, NM	659	8393	1853	1483	6558	3026	4496	2612	5727

Principal Component Analysis

PCA is a linear transformation used to transform one set of variables into another set of variables.

PCA is used to provide information on the true **dimensionality** of a data set

PCA tells you if the data set can be transformed into a fewer number of variables that still contain **most** of the essential information.

PCA also implements that transformation.

PCA

Used in complex systems such as neuroscience, imaging photometry, meteorology, oceanography

- the number of variables to measure can occasionally be huge, and knowledge of which variables to measure is not apparent
- at times underlying relationships can often be quite simple but are obscured.

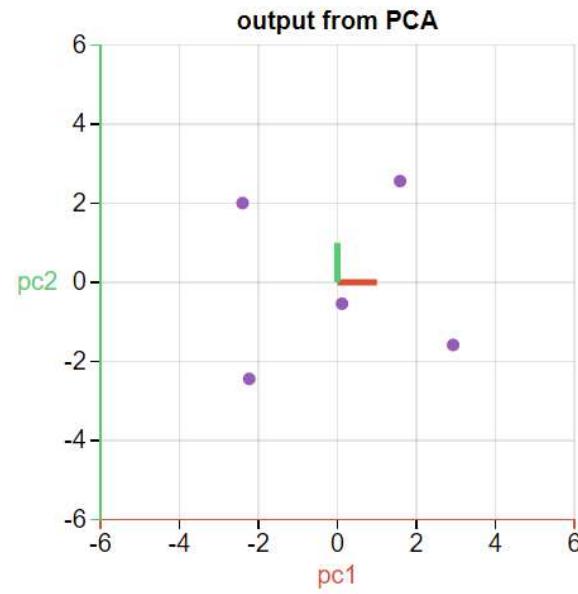
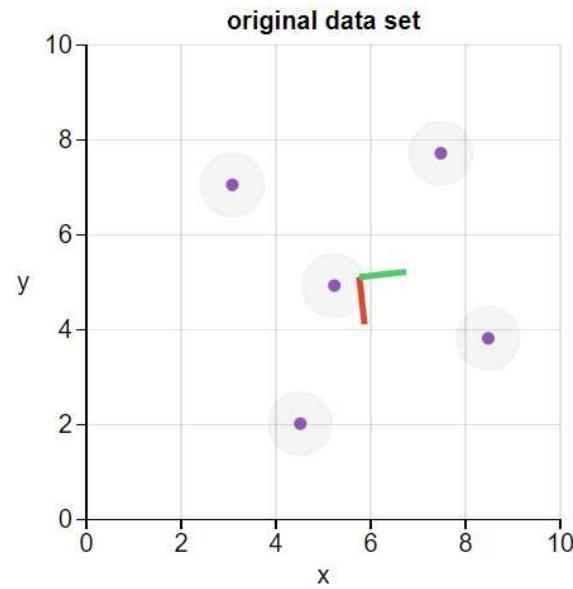
PCA

Model free multivariate statistical approach

PCA rotates (transforms) a data set until all the variables are uncorrelated.

- Uncorrelated variables provide no information on one another.
- summarize data variability into as few spatial components ("eigenvalues or eigenimages") as possible.
- 1st eigenvalue represents the largest source of variance, the second the largest source of residual variance orthogonal to first, etc.
- groups of multivariate data often subtly reflect the same driving process/behavior of the system.
- simplify the problem by replacing a group of variables with a single new variable.

Interactive Example



<https://setosa.io/ev/principal-component-analysis/>

Principal Component Analysis (PCA)

Linear decomposition of very correlated data to a new coordinate system along maximal variance

$$X = TP^T$$

- Each principal component (p_i) is vector in direction of largest variance in data and orthogonal to all other components
- Projection of each variable onto principal component produces loading vector (t_i)

Usually not all components needed due to correlation

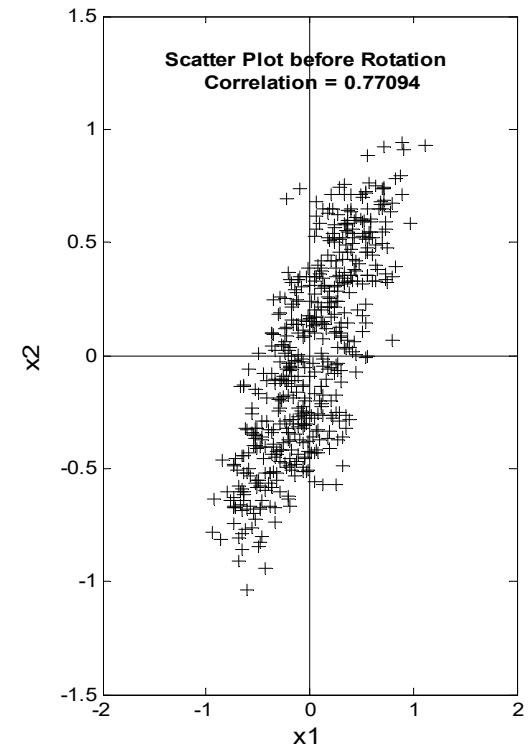
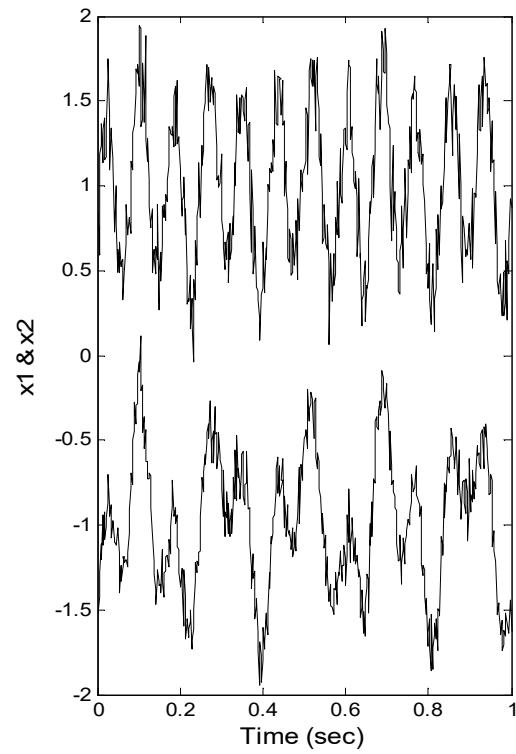
- Allows to create statistical model (\hat{X}) of data with minimal assumptions

$$\begin{aligned} X &= t_1 p_1^T + t_2 p_2^T + \cdots + t_K p_K^T \\ &= t_1 p_1^T + t_2 p_2^T + \cdots + t_A p_A^T + E \quad A < K \\ &= \hat{X} + E \end{aligned} \qquad \text{\scriptsize K is number of data sources}$$

PCA Example

E.g. 2-variable data set made from the mixtures of two sine waves. Each mixture contains different amplitudes of the two sinusoids and noise.

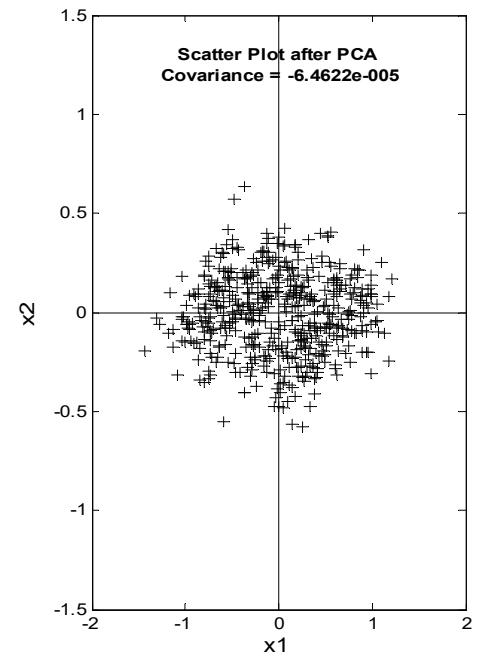
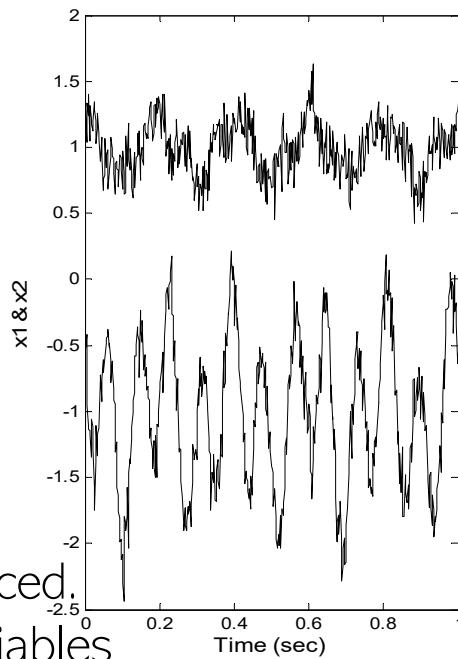
- There is a high degree of correlation between x_1 and x_2 .
- Knowing the value of x_1 gives you a range of possible x_2 values.



PCA Example

After rotation the two variables are no longer correlated.

- However the new variables are still mixtures of the two sources (just a different mixing).
- The new variables are **not more meaningful** than the original variables.
- Moreover this data set can not be reduced.
In this case, you really do need two variables to represent the two sinusoids.



Uncorrelated Does Not Mean Independent

If two (or more) variables are statistically independent

- they will also be uncorrelated

If two (or more) variables are uncorrelated

- they may not be statistically independent

This is suggested by the plots for two-variable data set plotted as time and scatter plots in the next slide.

The two signals are uncorrelated, but they are highly related and not independent.

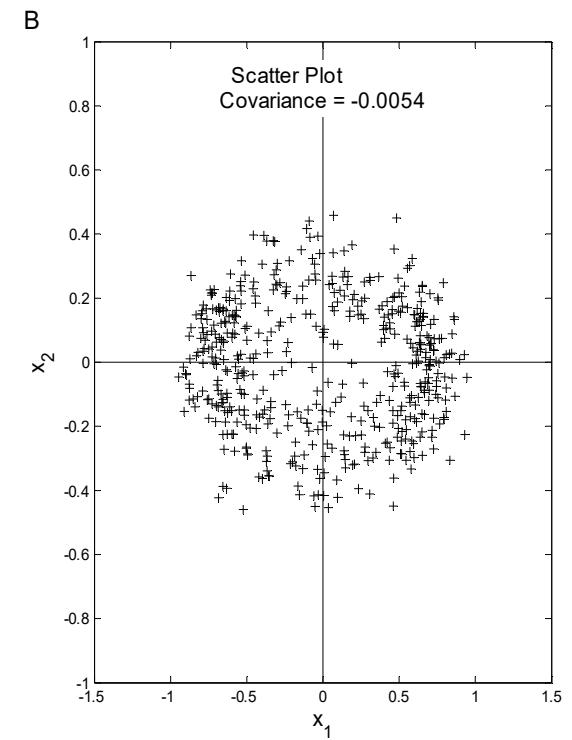
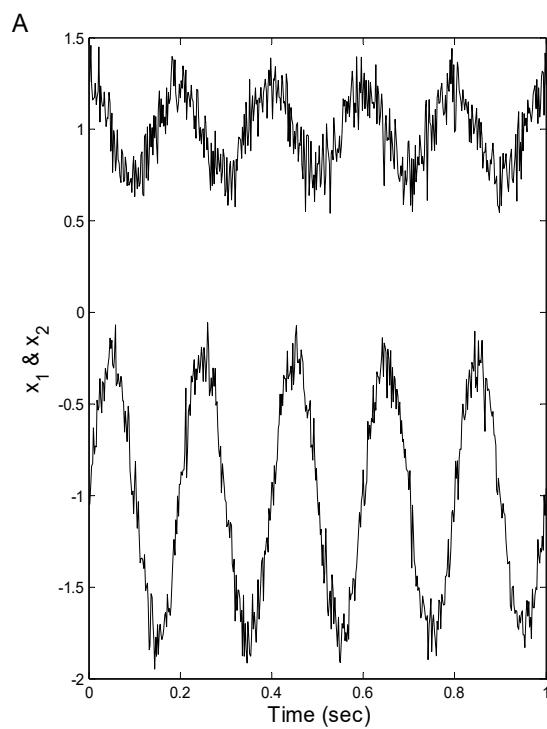
Example

Time (A) and scatter plot (B) of two variables that are uncorrelated, but not independent.

In fact, the two variables are highly dependent.

There is clearly a relationship (nonlinear) between them.

In fact, they were generated by a single equation, that of a circle, with noise added.



Spring Example

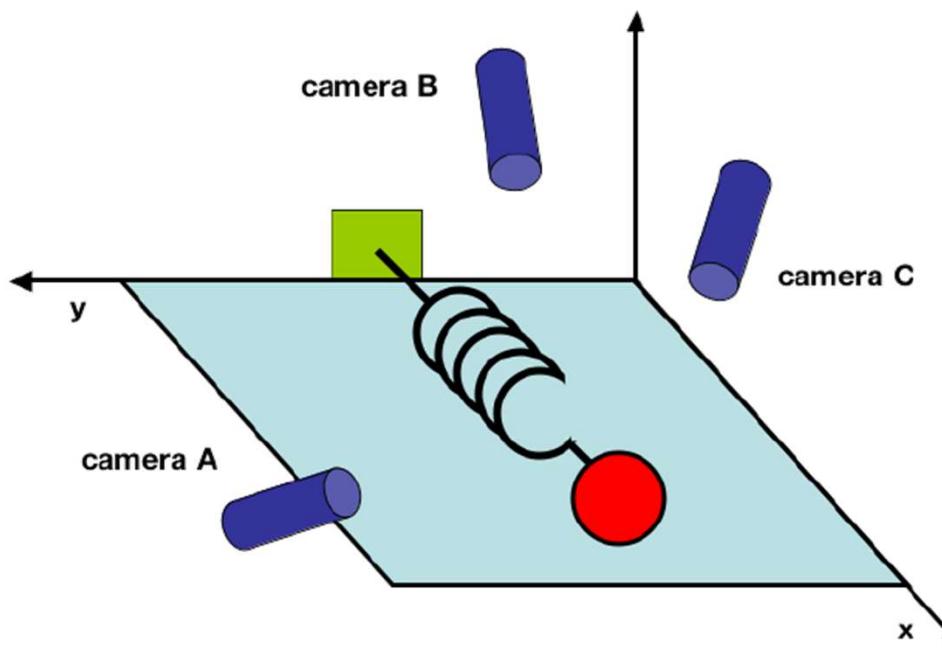
e.g. simple toy problem from physics: motion of an ideal spring.

- system consists of a ball of mass m attached to a massless, friction-less spring.

The ball is released a small distance away from equilibrium (i.e. the spring is stretched).

Because the spring is “ideal,” it oscillates indefinitely along the x-axis about its equilibrium at a set frequency.

Spring Example



Each camera is at some arbitrary angle sampling images at 120 Hz (i.e. camera records an image indicating the two dimensional position of the ball, a projection)

Spring Example

- record for several minutes. The big question remains: how do we get from this data set to a simple equation of x ?
- also what about noise?
- The goal of PCA is to compute the most meaningful basis to re-express a noisy data set.
- hopefully this new basis will filter out the noise and reveal hidden structure.

Spring Example

- For each time point a camera records ball position
- at each time point we have 6 variables measured
- so with 120Hz sampling we have 72000 values in 10 minutes

$$\vec{X} = \begin{bmatrix} x_A \\ y_A \\ x_B \\ y_B \\ x_C \\ y_C \end{bmatrix}$$

- 3 cameras
- 2D info (x and y)

- Each sample X is an m -dimensional vector
- m is the number of measurement types.

Spring Example

From linear algebra we know that all measurement vectors form a linear combination of this set of unit length basis vectors. What is this orthonormal basis?

Pretend we gathered our toy example data above, but only looked at camera A. What is an orthonormal basis for (x_A, y_A) ?

In the two dimensional case, $\{(1, 0), (0, 1)\}$ can be recast as individual row vectors to make a 2×2 identity matrix, I

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

Spring Example

generalize this to the m-dimensional case by constructing an $m \times m$ identity matrixx

each row is an orthonormal basis vector b_i with m components

All of the camera data has been recorded in this basis and thus it can be trivially expressed as a linear combination of $\{b_i\}$.

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_m \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \mathbf{I}$$

PCA: Linearity

PCA: Is there another basis, which is a linear combination of the original basis, that best re-expresses the data?

One strict assumption: linearity. This vastly simplifies the problem by

- (1) restricting the set of potential bases
- (2) formalizing the implicit assumption of continuity in a data set

In the toy example X is an $m \times n$ matrix where $m = 6$ and $n = 72000$.

Let Y be another $m \times n$ matrix related by a linear transformation P .

X is the original recorded data set and Y is a re-representation of that data set.

$$PX = Y$$

- \mathbf{p}_i are the *rows* of \mathbf{P}
- \mathbf{x}_i are the *columns* of \mathbf{X} (or individual \vec{X})
- \mathbf{y}_i are the *columns* of \mathbf{Y} .

PCA: Linearity

This equation represents a change of basis and can be interpreted a number of ways:

$$P\mathbf{X} = \mathbf{Y}$$

1. \mathbf{P} is a matrix that transforms \mathbf{X} into \mathbf{Y} .
2. Geometrically, \mathbf{P} is a rotation and a stretch which again transforms \mathbf{X} into \mathbf{Y} .
3. The rows of \mathbf{P} , $\{\mathbf{p}_1, \dots, \mathbf{p}_m\}$, are a set of new basis vectors for expressing the *columns* of \mathbf{X} .

PCA: Linearity

By assuming linearity the problem reduces to finding the appropriate change of basis

But, need to deal with:

- 1) noise
- 2) rotation
- 3) redundancy

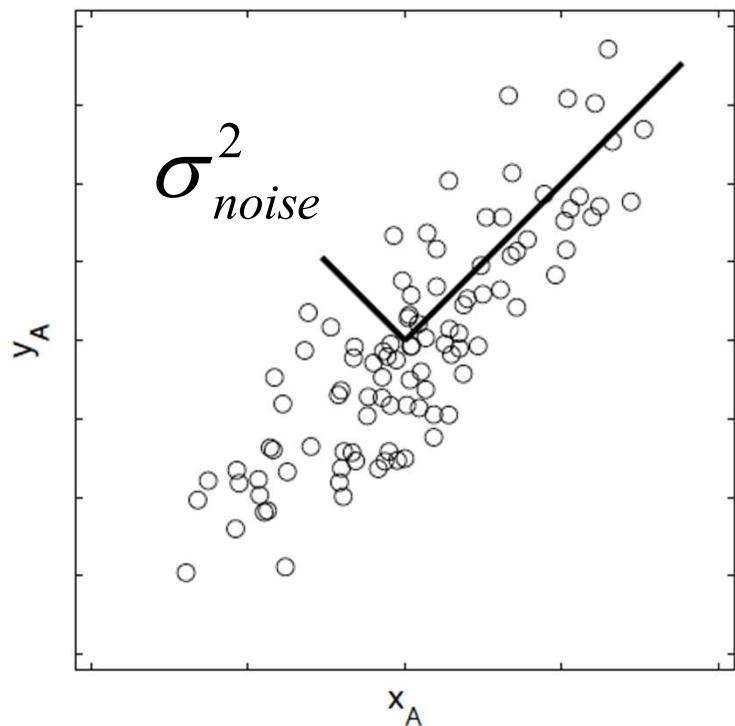
Noise and Rotation

Measurement noise must be low or else no information about the system can be extracted! This is irrespective of the analysis technique

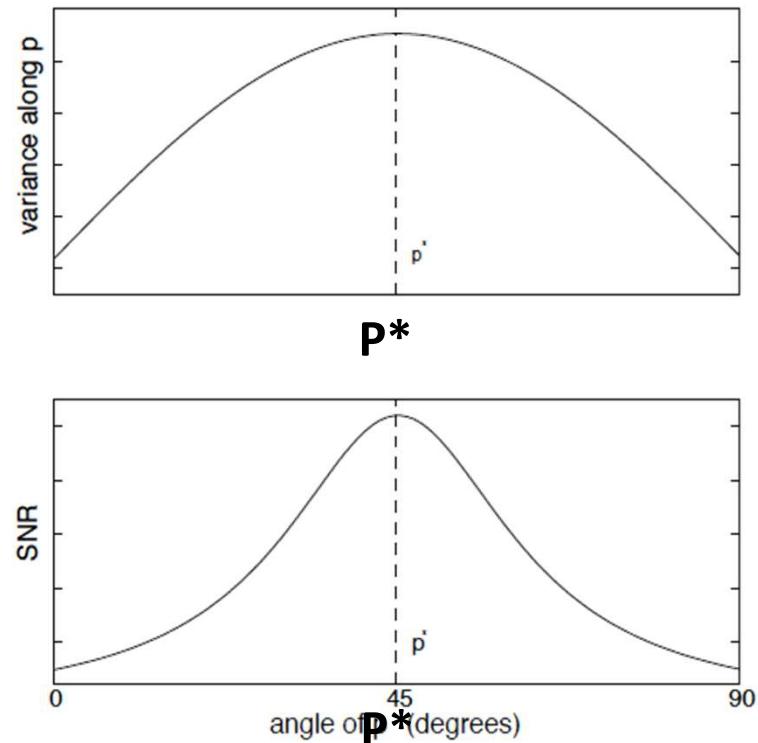
$$SNR = \frac{\sigma_{signal}^2}{\sigma_{noise}^2}$$

- high SNR ($\gg 1$) indicates high precision data, while low SNR indicates noise contaminated data

Revisiting the Spring



σ_{signal}^2



\mathbf{p}^*

angle of \mathbf{p}^* (degrees)

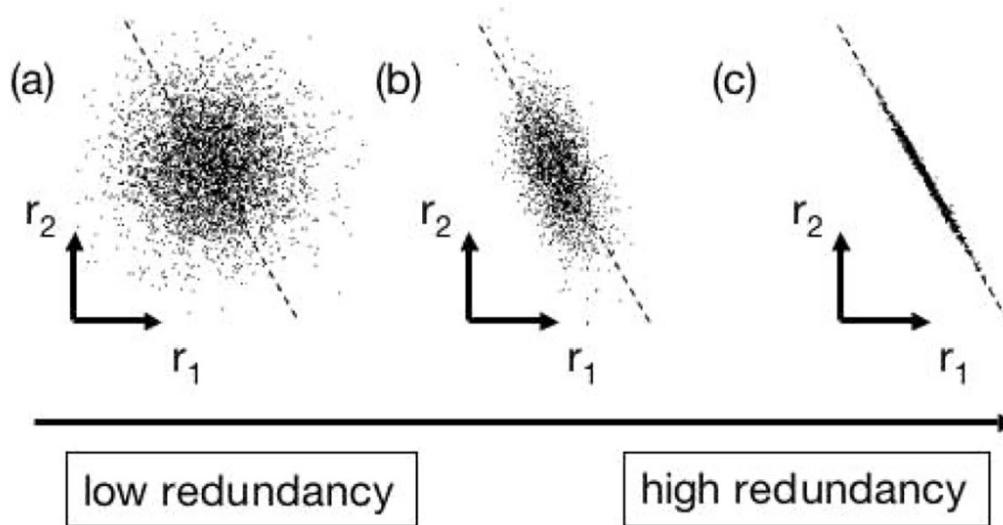
2D PCA

- Quantitatively assume that directions with largest variances in our measurement vector space contain the dynamics of interest (i.e. direction along the long axis of the cloud).
- Maximizing the variance (and by assumption the SNR) corresponds to finding the appropriate rotation of the naive basis (i.e. direction of P^*)
- In this simple 2-dimensional case P^* falls along the direction of the best-fit line for the data cloud. Thus, rotating the naive basis to lie parallel to P^* would reveal the direction of motion of the spring.

What about multiple dimensions??

Redundancy

- an additional potential confounding factor. i.e. multiple sensors recording the same information



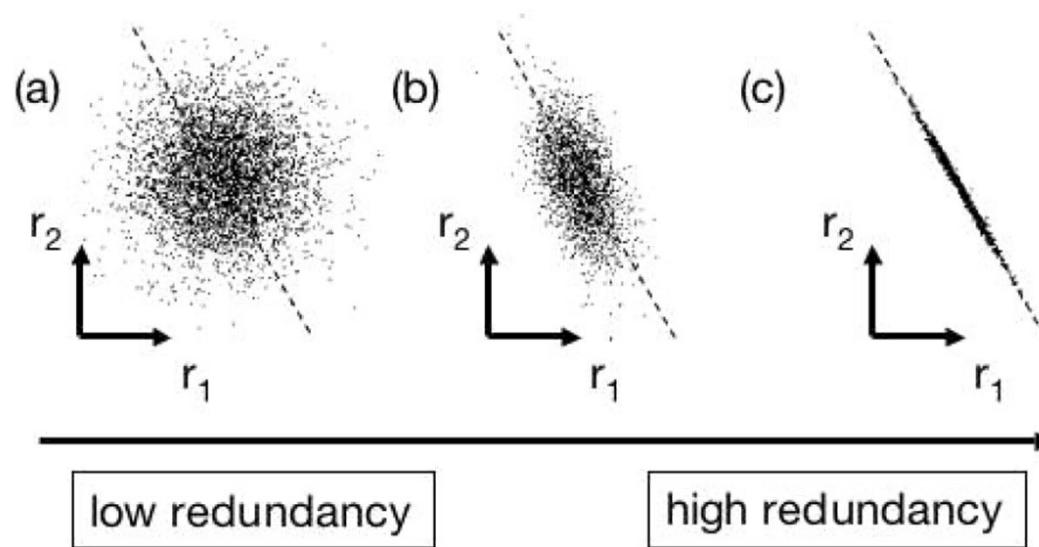
In (c) it would have been more meaningful to just have recorded a single variable, not both. Because one can calculate r_1 from r_2 (or vice versa) using the best fit line.

Recording only 1 expresses data more concisely and reduces number of sensor recordings (dimensional reduction).

Redundancy

simple to identify redundancy in 2 variable cases: find slope of the best-fit line and judge the quality of the fit.

BUT: How can this be generalized to higher dimensions?



Measuring Redundancy

Consider two sets of measurements with zero means, where in both cases the subscript denotes the sample number:

Variances:

$$A = \{a_1, a_2, \dots, a_n\}$$

$$\sigma_A^2 = \langle a_i a_i \rangle_i$$

$$B = \{b_1, b_2, \dots, b_n\}$$

$$\sigma_B^2 = \langle b_i b_i \rangle_i$$

$$\text{covariance of } A \text{ and } B \equiv \sigma_{AB}^2 = \langle a_i b_i \rangle_i$$

- covAB measures the degree of linearity between 2 variables.
- If covAB is large there is high redundancy.

Measuring Redundancy

$$\sigma_{AB}^2 \geq 0$$

$\sigma_{AB}^2 = 0$ iff A and B are totally uncorrelated

$$\sigma_{AB}^2 = \sigma_A^2 = \sigma_B^2 \quad (if A = B)$$

Variance

$$\sigma_{AB}^2 \geq 0$$

$\sigma_{AB}^2 = 0$ iff A and B are totally uncorrelated

$$\sigma_{AB}^2 = \sigma_A^2 = \sigma_B^2 \quad (\text{if } A = B)$$

- can equivalently convert A and B into corresponding row vectors:

$$\mathbf{a} = [a_1 \ a_2 \ \dots \ a_n]$$

$$\mathbf{b} = [b_1 \ b_2 \ \dots \ b_n]$$

now express the covariance as a dot product:

$$\sigma_{\mathbf{ab}}^2 \equiv \frac{1}{n-1} \mathbf{ab}^T$$

(note: $(n-1)^{-1}$ is used for normalization)

Combining vectors

Now, we are not stuck with only 2 vectors in real life. So we can arbitrarily call them:

$$\mathbf{x}_1 \equiv \mathbf{a}$$

$$\mathbf{x}_2 \equiv \mathbf{b}$$

Where additional indexed row vectors:

$$\mathbf{x}_3, \dots, \mathbf{x}_m$$

Define new $m \times n$ matrix, X :

- rows of X correspond to all measurements of a particular type (x_i).

Each column of X corresponds to a set of measurements from one particular trial

$$X = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_m \end{bmatrix}$$

Covariance Matrix

- matrix $\mathbf{X}\mathbf{X}^T$ computes ij^{th} element of \mathbf{C}_x
- ij^{th} element of \mathbf{C}_x is dot product between the vector of the i^{th} measurement type with the vector of the j^{th} measurement type

properties of \mathbf{C}_x :

- 1) \mathbf{C}_x is a square symmetric $m \times m$ matrix
- 2) Diagonal terms of \mathbf{C}_x are the variance of particular measurement types (large=interesting; small=noise)
- 3) Off-diagonal terms of \mathbf{C}_x are the covariance between measurement types (large=high redundancy)

$$\mathbf{C}_x \equiv \frac{1}{n - 1} \mathbf{X}\mathbf{X}^T$$

Goals of PCA

- (1) minimize redundancy, measured by covariance
- (2) maximize signal, measured by variance.

$$PX = Y$$

By definition covariances must be non-negative, thus the minimal covariance is zero.

Therefore, the “optimized” covariance matrix of \mathbf{CY} would have all off-diagonal terms equal to zero (i.e. \mathbf{CY} must be diagonal). So, need to diagonalize \mathbf{CY} .

- The principal components of \mathbf{X} are the eigenvectors of \mathbf{XX}^T ; or the rows of \mathbf{P} .
- The i^{th} diagonal value of \mathbf{CY} is the variance of \mathbf{X} along \mathbf{p}_i .

Performing PCA

1. Organize a data set as an $m \times n$ matrix, where m is the number of measurement types and n is the number of trials
2. Subtract the mean for each measurement type or row x_i
3. Calculate the eigenvectors of the covariance matrix.

A Multivariate Example

In multivariate analysis, multiple variables are analyzed together and often represented as a single vector variable that includes the different variables :

$$\mathbf{x} = [x_m(1), x_m(2), \dots, x_m(N)]^T \text{ for } 1 \leq m \leq M$$

The 'T' stands for transposed.

The variable X is composed of M variables (rows) each containing N observations (columns). In signal analysis, the observations are time samples, ($t = 1, \dots, N$)

$$X = \begin{bmatrix} x_1[1] & x_1[2] & x_1[3] & \dots & x_1[N] \\ x_2[1] & x_2[2] & x_2[3] & & x_2[N] \\ x_3[1] & x_3[2] & \ddots & & x_3[N] \\ \vdots & \vdots & & \ddots & \vdots \\ x_M[1] & x_M[2] & x_M[3] & \dots & x_M[N] \end{bmatrix}$$

Multivariate Analysis (cont)

In signal processing the observations are time samples while in image processing they are pixels.

Multivariate data, as represented by X above, can also be considered to reside in M -dimensional space, where each spatial dimension contains one set of observations.

Multivariate analysis takes into account relationships between and within the multiple variables.

(For example, the covariance matrix includes information about the relationship between variables as well as about the variables themselves.)

Multivariate Analysis (continued)

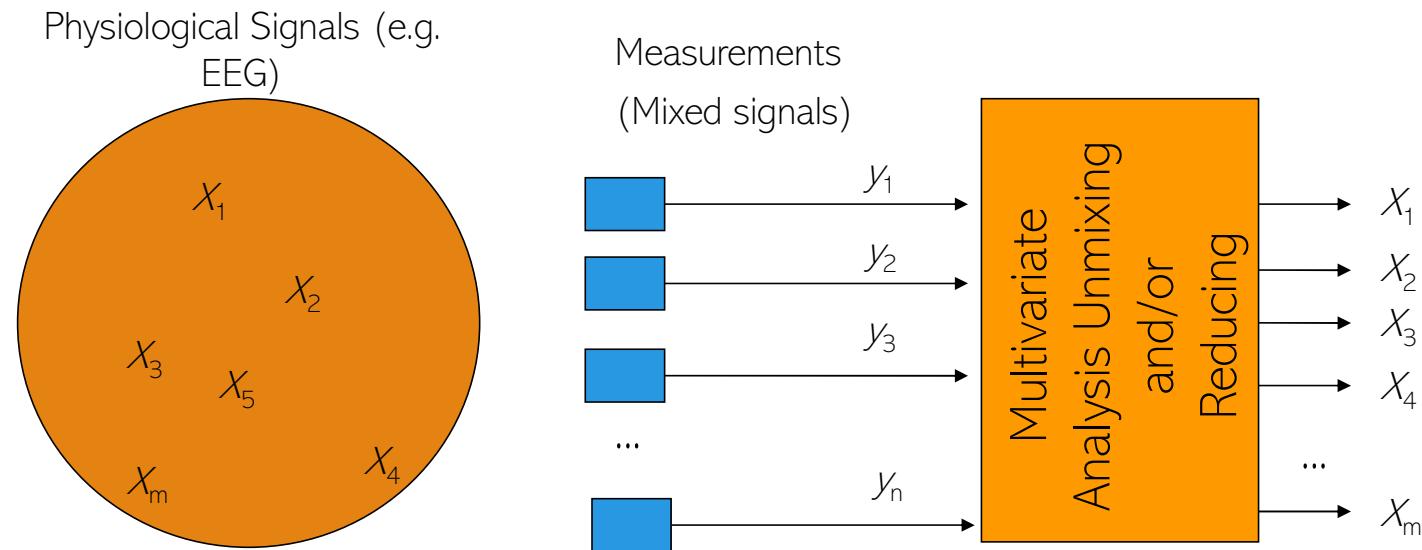
A major concern of multivariate analysis is to find transformations of the data that make the data set **smaller or easier** to understand.

1) Can the relevant information contained in a multi-dimensional variable be expressed using fewer variables? Are variables redundant?

2) Can the data be transformed to be more meaningful?

If so, the more meaningful variables are described as hidden, or latent, in the original data (they are the **latent variables**).

Multivariate Applications in Behavioral Physiology



The y 's (measured) are mixtures of the x 's.

You have the y 's, but you want the x 's.

Usually $n > m$.

Multivariate Data Transformations

In transformations that reduce the dimensionality of a multi-variable data set, the idea is to transform one set of variables into a new set where some of the new variables have values that are much smaller than the rest.

The data transformation used to produce the new set of variables is often a linear function.

A linear transformation can be represented mathematically as:

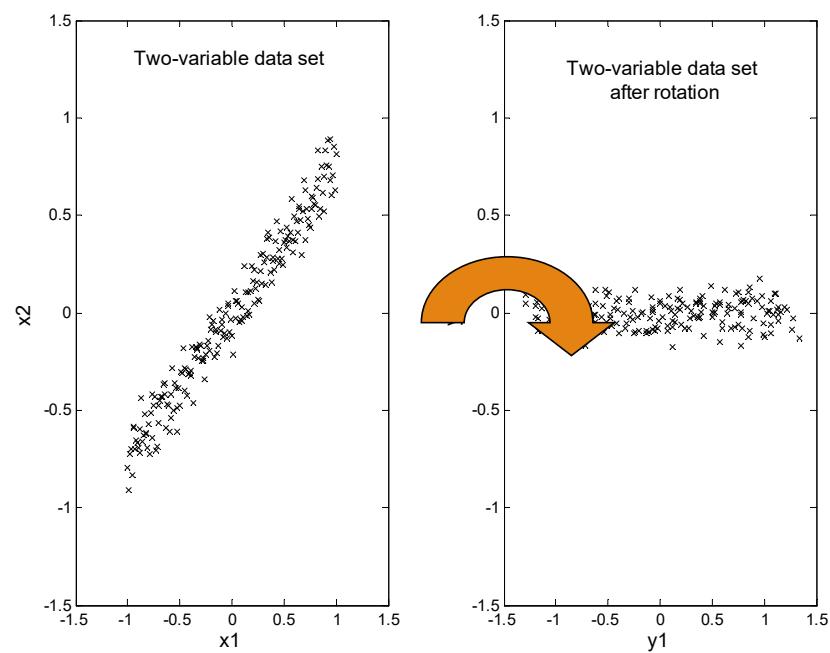
$$\begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_M(t) \end{bmatrix} = W \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_M(t) \end{bmatrix} \quad \text{or: } y_i(t) = \sum_{j=1}^M w_{ij} x_j(t) \quad i = 1, \dots, N$$

where $W (= w_{ij})$ is a matrix that defines the transformation.

Linear Transformations

A linear transformation can be interpreted as a **rotation** of the data set.

- The rotated data set still contains two variables, but the variance of one of the variables is quite small compared to the other.
- Perhaps this new variable (y_2) just represents noise and could be eliminated



Determining PCA using Singular Value Decomposition

The easiest way to implement PCA is using singular value decomposition (SVD).

Singular value decomposition factors a data matrix, X , into the product of three matrices:

- 1) A diagonal matrix, S , containing the square root of the eigenvalues
- 2) A principal components matrix, U ,
- 3) and its orthonormal version, V

$$X = USV^T$$

The symbols and the details of SVD vary depending on the source; here the definitions are based on MATLAB

SVD (continued)

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$$

where \mathbf{X} is the $m \times n$ data matrix. This matrix is decomposed into:

\mathbf{U} = $m \times m$ orthonormal matrix;

\mathbf{S} = $m \times n$ diagonal matrix;

\mathbf{V} = $n \times n$ matrix containing the principle components

**these are based on MATLABS SVD routines, can be different depending on the source

SVD (continued)

U provides a transformation matrix that will produce a rotated data set which has zero covariance.

S is the covariance matrix of the new (rotated) data set.

It is a diagonal matrix where the diagonals are the variances of the new data set

Squaring the diagonals produces the eigenvalues denoted: $\lambda_1, \lambda_2, \dots, \lambda_n$ which are ordered by magnitude: $\lambda_1 > \lambda_2 > \dots > \lambda_n$.

V has columns that are the characteristic vectors, or eigenvectors u_1, u_2, \dots, u_n .

When scaled by their respective variances (the square root of the eigenvalues) they become the principle components, the new data set.

Data Reduction: The Scree Plot

The eigenvalues are related to the [variances](#) of the principle components.

They are a measure of the associated importance of each principal component

The eigenvalues are in order of magnitude:

- 1) The first principal component accounts for the maximum variance possible in a single component.
- 2) The second component accounts for the maximum of the remaining variance for a single component
- :
- N) The last principal component accounts for the smallest amount of variance.

Data Rotation

Many multivariate operations involve data rotation. From basic trigonometry, it is easy to show that, in two dimensions, rotation of a data point (x_1, x_2) can be achieved by multiplying the data points by the sines and cosines of the rotation angle:

$$\begin{aligned}y_1 &= x_1 \cos(\theta) + x_2 \sin(\theta) \\y_2 &= -x_1 \sin(\theta) + x_2 \cos(\theta)\end{aligned}$$

where θ is the angle through which the data set is rotated in radians. Using matrix notation, this operation can be done by multiplying the data matrix by a ‘rotation’ matrix:

Principle Component Scaling

It is common to normalize the principal components by the variances so that different components can be compared.

A number of other normalizing schemes exist.

Here we multiply the eigenvector by the square root of its eigenvalue (i.e., the variance)

This gives rise to principal components that have the same value as a manually rotated data array.

Data Reduction using PCA

The eigenvalues describe how much of the variance is accounted for by the associated principal component and if a component is really necessary.

Eigenvalues are ordered by size; that is:

$$\lambda_1 > \lambda_2 > \lambda_3 \dots > \lambda_M.$$

If an eigenvalue is zero or 'close to' zero, then its associated principal component contributes little to the data and can be eliminated. This component accounts for only a small amount of the variance in the data.

This tells us the effective dimension of the data set.

How do you decide if an eigenvalue is small enough so that its associated component can be removed from the data set?

Data Reduction (cont)

There are two popular methods for determining eigenvalue “thresholds”.

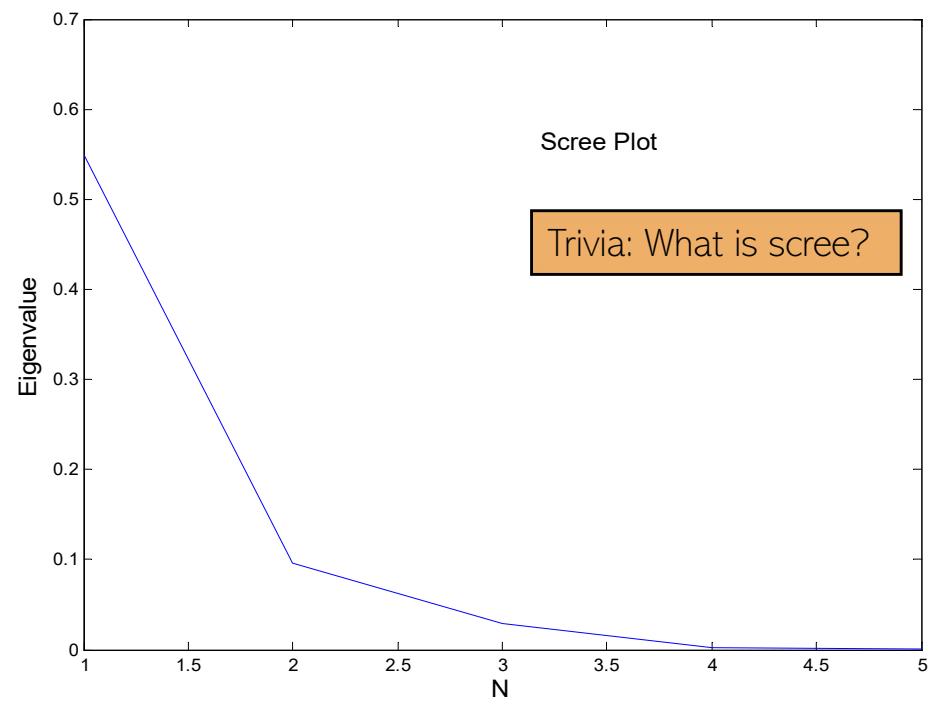
- 1) Take the sum of all eigenvectors (which must account for all the variance), then delete those eigenvalues that fall below some percentage of that sum (e.g. 95% or whatever you want).

- 2) Scree Plot: plot of eigenvalues in order and look for breakpoints in the slope of this curve. Eigenvalues representing noise should not change much in value and will plot as a flatter slope when plotted sequentially.

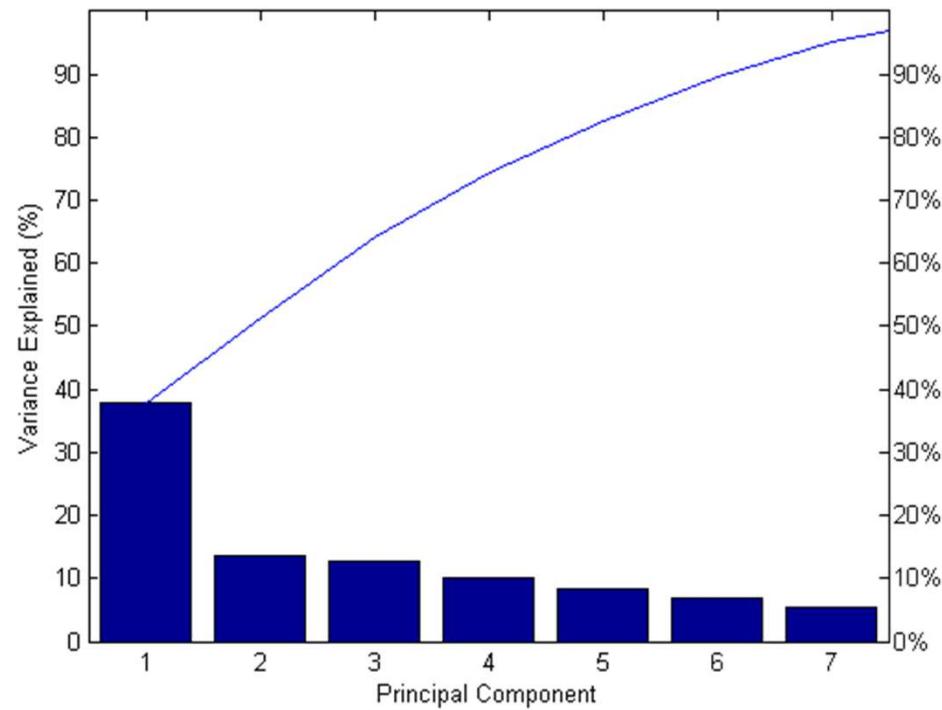
The Scree Plot

Eigenvalues are in order of large to small.

The actual dimension of the data set is taken where the Scree plot becomes more-or-less flat.



Scree Plot pt 2

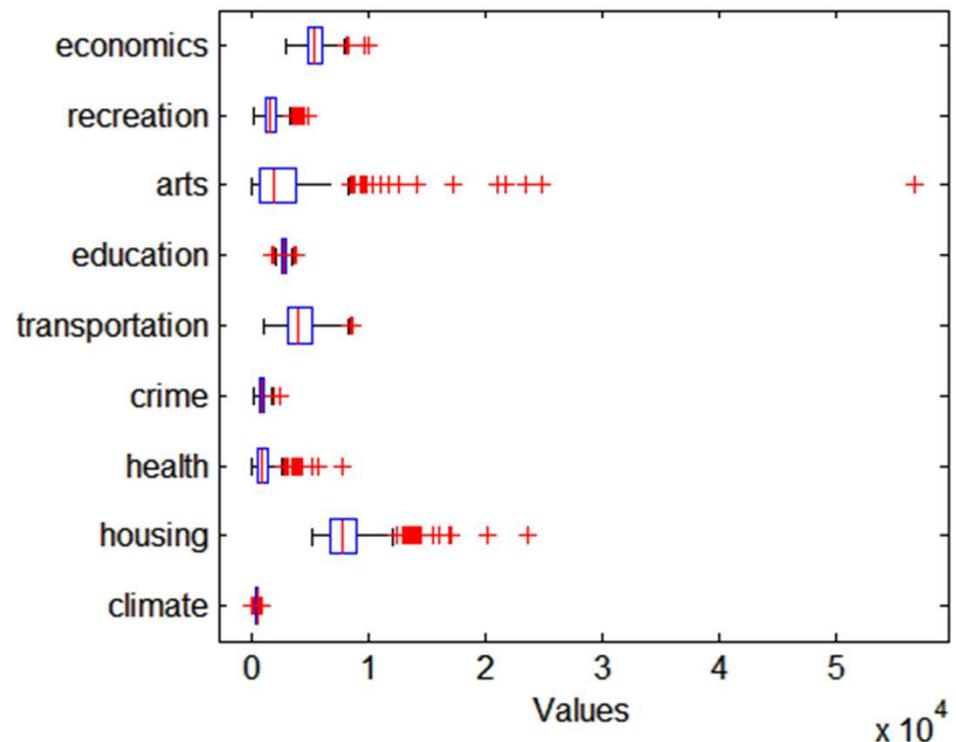


Example: Quality of Life in U.S. Cities

Matlab Sample Data: Quality of Life in U.S. Cities

Load **cities**

example shows how to perform a weighted principal components analysis and interpret the results. Here there are 9 components



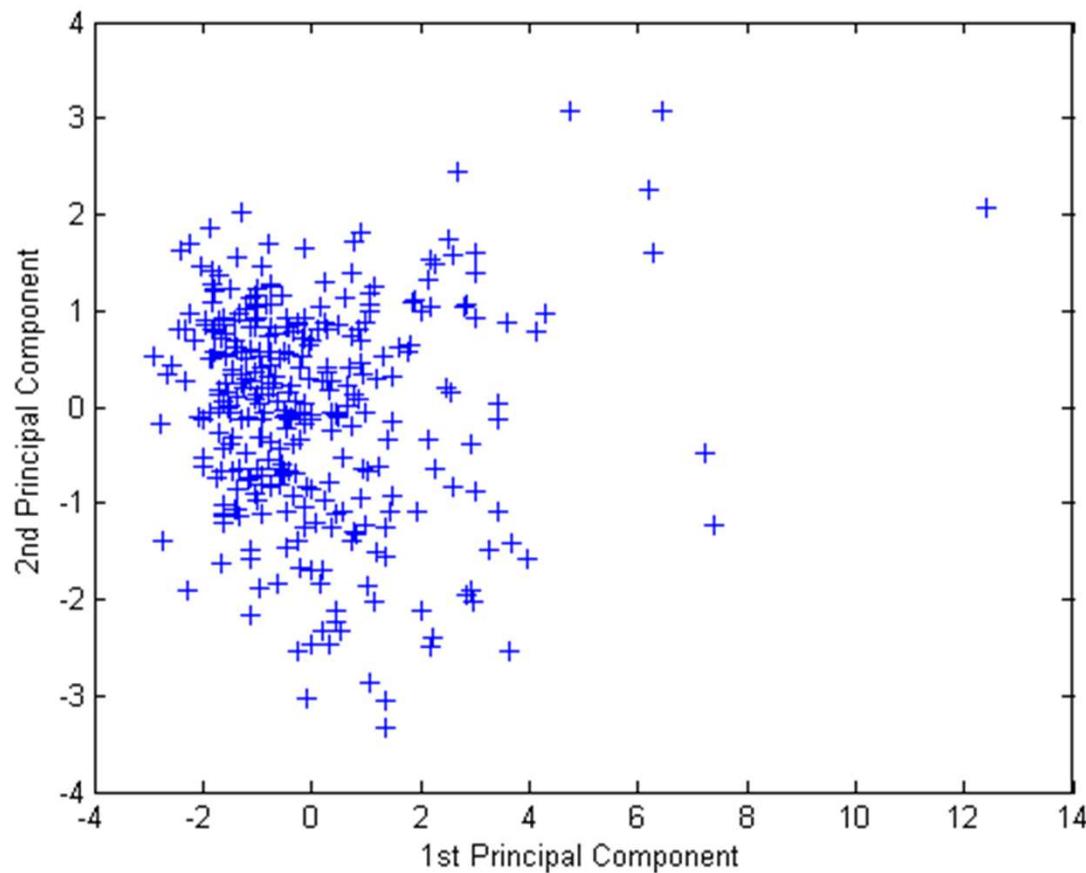
U.S. Cities Example

- Checking pairwise correlation between the variables shows relatively strong correlation among some variables (up to 0.85).
- PCA constructs independent new variables which are linear combinations of the original variables.

NOTE:

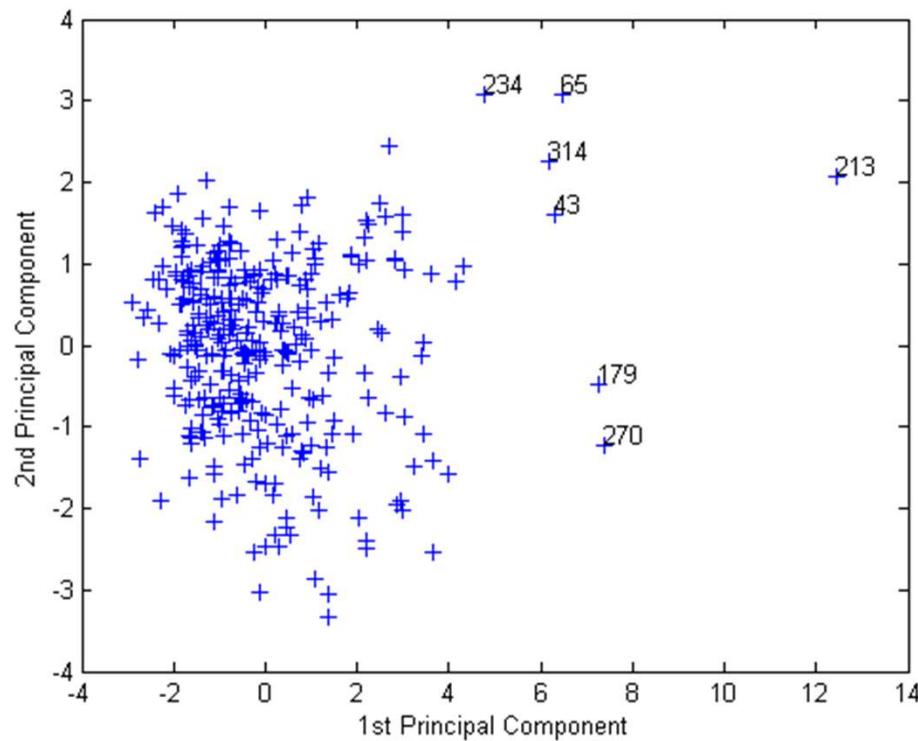
- when all variables have the same units, it is appropriate to compute principal components from 'raw data'.
- When the variables are in different units or the difference in the variance of different columns is substantial (like this example), scaling of the data or use of weights is often preferable.

This plot shows the centered and scaled ratings data projected onto the first two principal components. PCA computes the scores to have mean zero.
→ each + is a particular city



Note outliers. Graphically identify using **gname**

Boston, Chicago, Los Angeles, Long Beach, New York, Philadelphia, San Francisco, Washington DC. These labeled cities appear more extreme than the remainder of the data

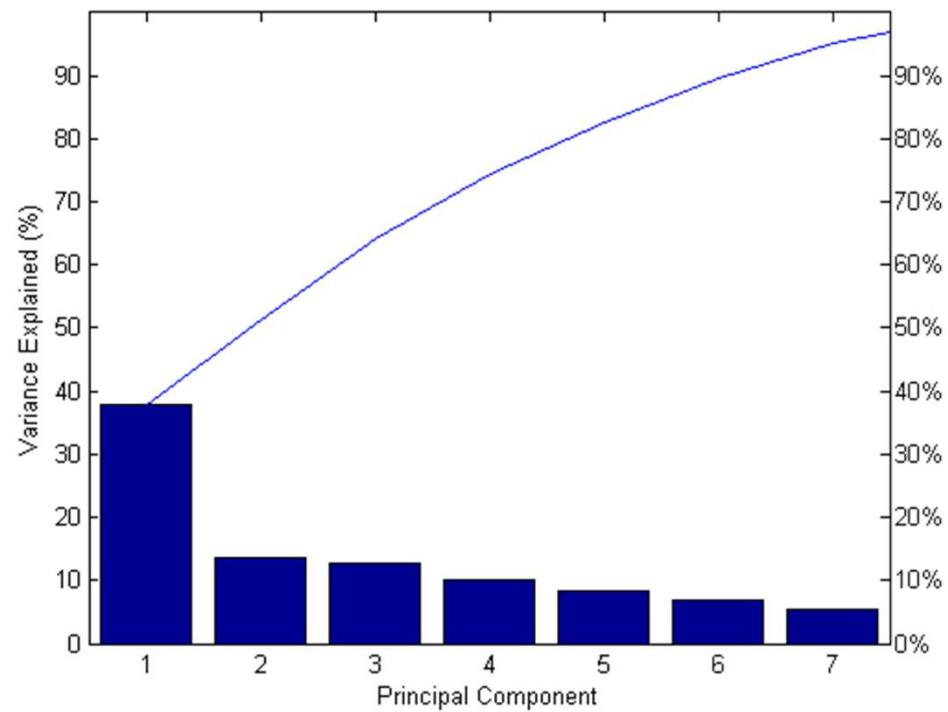


Scree Plot

What are the sources of variance (in order of importance)?

- 95% of the variance is explained by the first 7 components

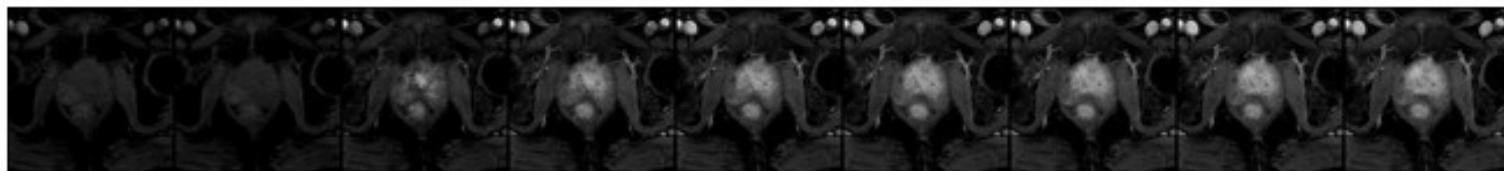
Can have 9 principle components total (i.e. # of measured 9 classes)



PCA of 4D Image Data

By “folding” spatial dimensions into matrix, X, it is possible to analyze “temporal signatures” of 4D MRI data

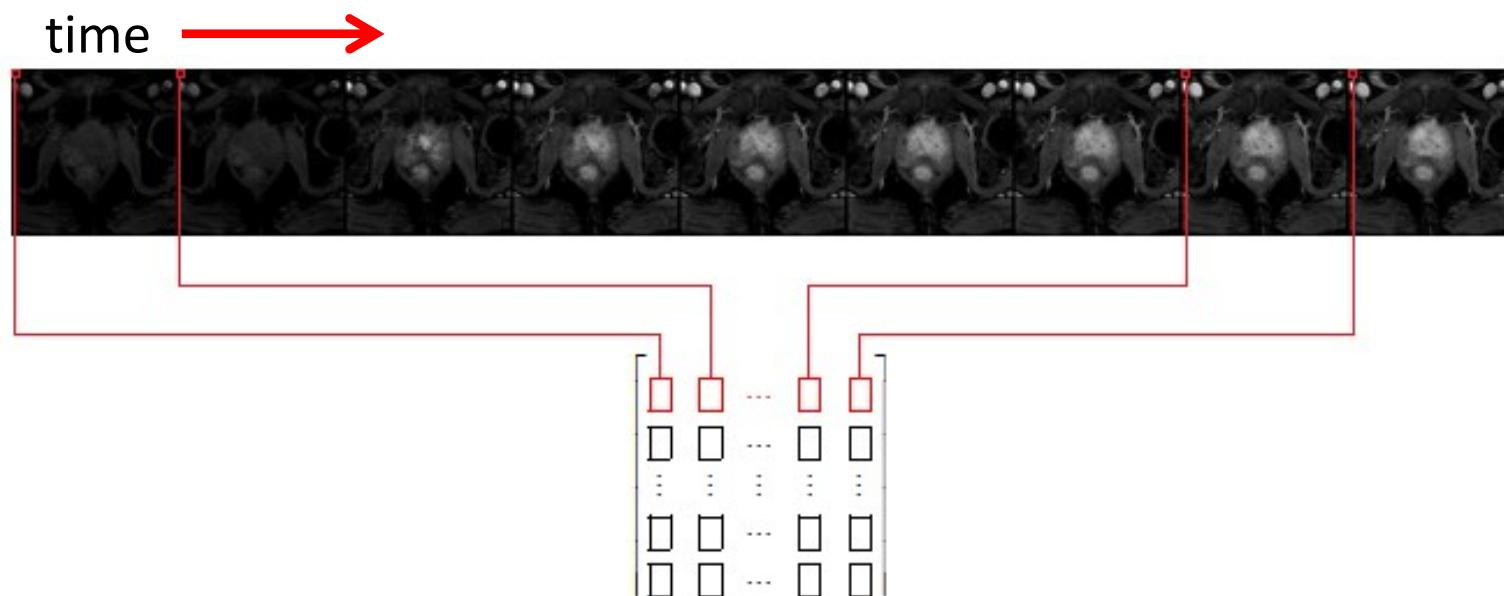
time →



$$\begin{bmatrix} \square & \square & \cdots & \square & \square \\ \square & \square & \cdots & \square & \square \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \square & \square & \cdots & \square & \square \\ \square & \square & \cdots & \square & \square \end{bmatrix}$$

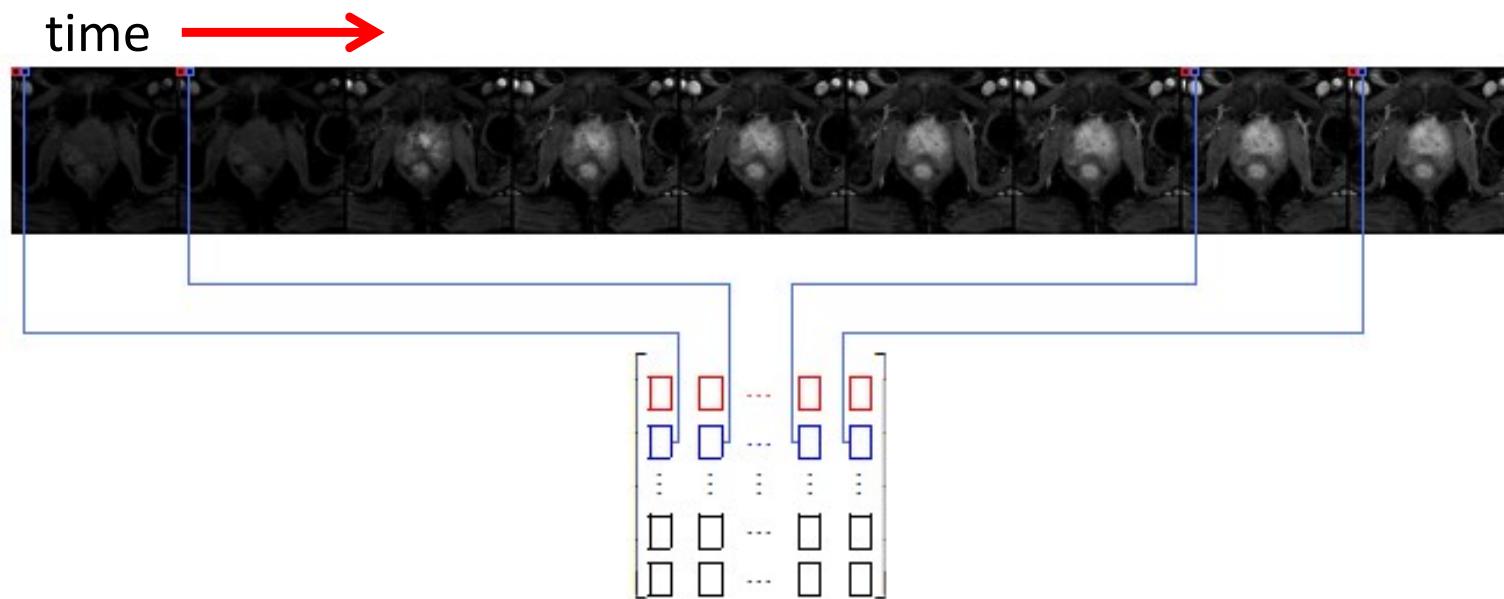
PCA of 4D Image Data

By “folding” spatial dimensions into matrix, X, it is possible to analyze “temporal signatures” of 4D MRI data



PCA of 4D Image Data

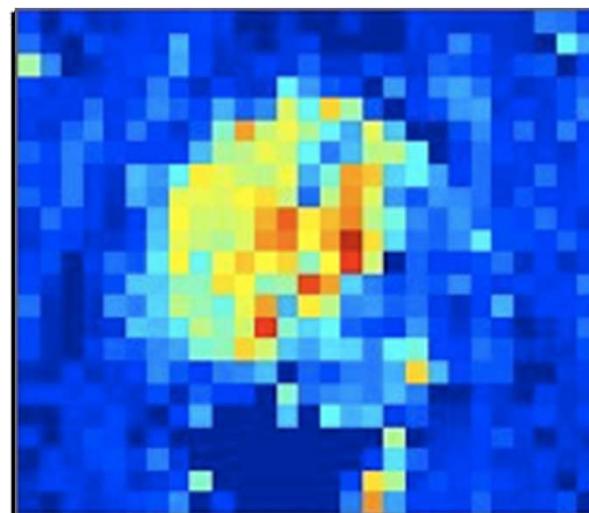
By “folding” spatial dimensions into matrix, X, it is possible to analyze “temporal signatures” of 4D MRI data



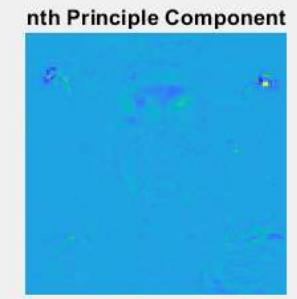
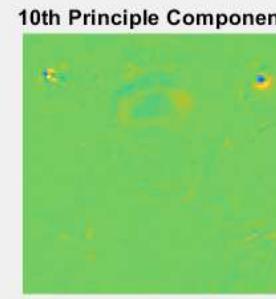
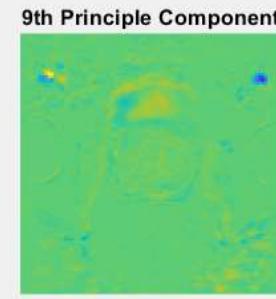
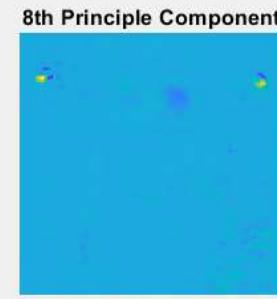
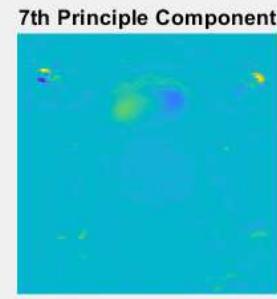
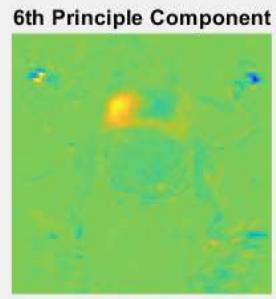
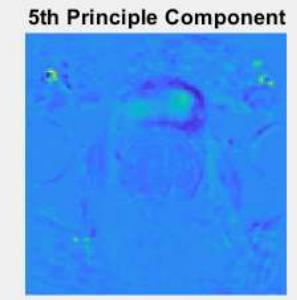
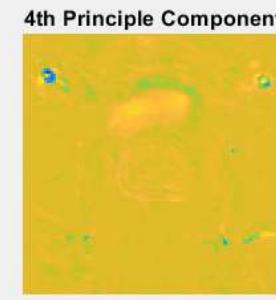
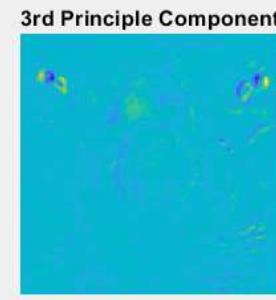
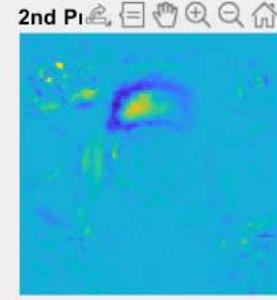
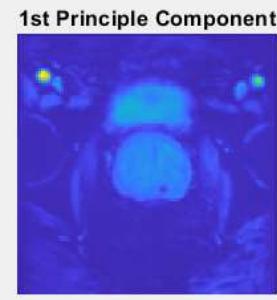
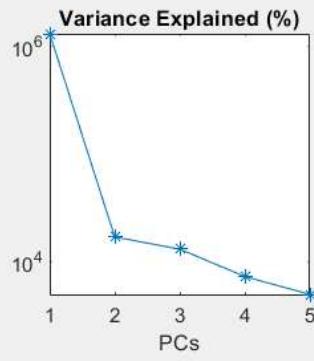
PCA of 4D Image Data

PCA generates a new set of variables: principal components.

All the principal components are orthogonal to each other so there is no redundant information.

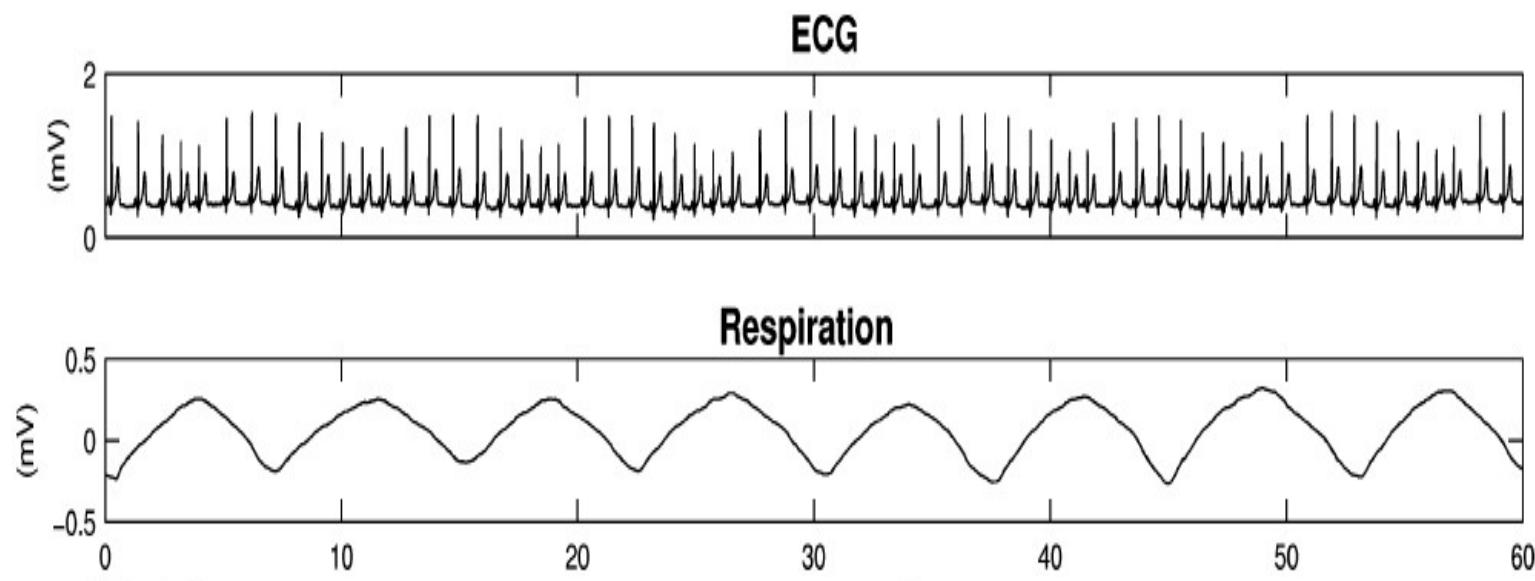


Static T2-weighted prostate image (left) and corresponding overall degree of enhancement coded green; rate of wash-in / wash-out coded red (right). *Bruwer, MacGregor, Noseworthy (2008) J. Chemometrics 22:708-716.*



Principal Component Analysis as a Tool for Analyzing Beat-to-Beat Changes in ECG Features: Application to ECG-Derived Respiration

Philip Langley*, Emma J. Bowers, and Alan Murray



Independent Component Analysis



Star Trek (i.e. the Original Series). Season 1, Episode #20, "Court Martial" Aired: Feb. 2nd, 1967

Independent Component Analysis

Independent Component Analysis seeks to transform the original data set into a number of independent variables. The motivation is primarily to uncover more meaningful variables, not to reduce the dimensions of the data set.

When data set reduction is also desired it is usually accomplished by preprocessing the data set using PCA.

One of the more dramatic applications of Independent Component Analysis (ICA) is found in the ‘cocktail party problem.’ In this situation, multiple people are speaking simultaneously in the same room.

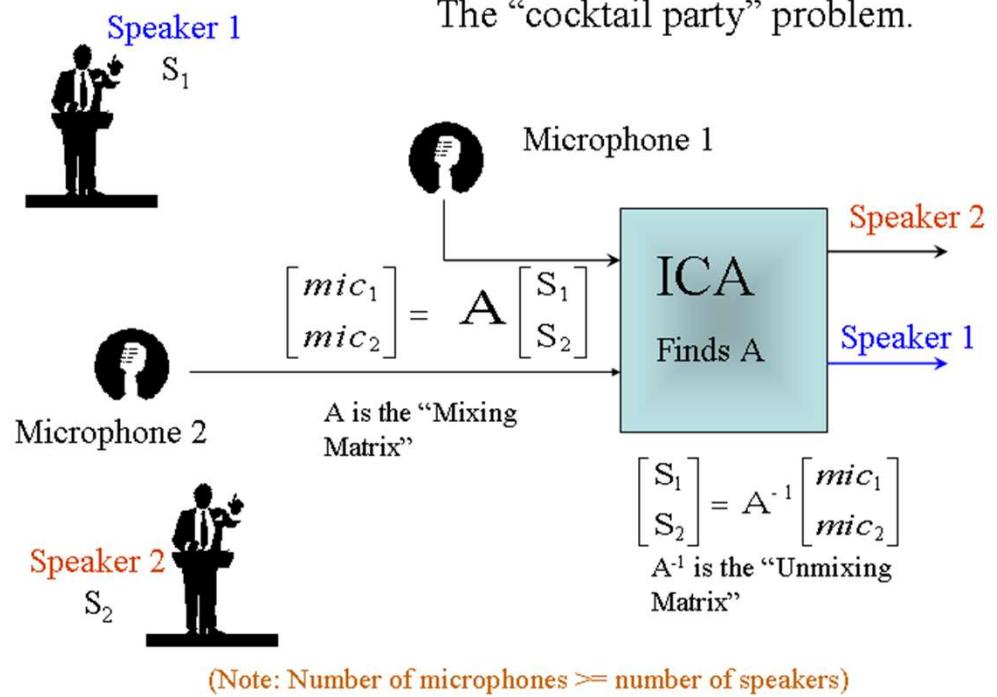
Independent Component Analysis (ICA)

- a method for extracting useful information from data.
- reveals the driving forces which underlie a set of observed phenomena. e.g. firing of a set of neurons, mobile phone signals, brain images (fMRI), stock prices, voices, etc.
- a large set of signals are measured, and it is known that each measured signal depends on several distinct underlying factors i.e. each measured signal is essentially a mixture of underlying factors.

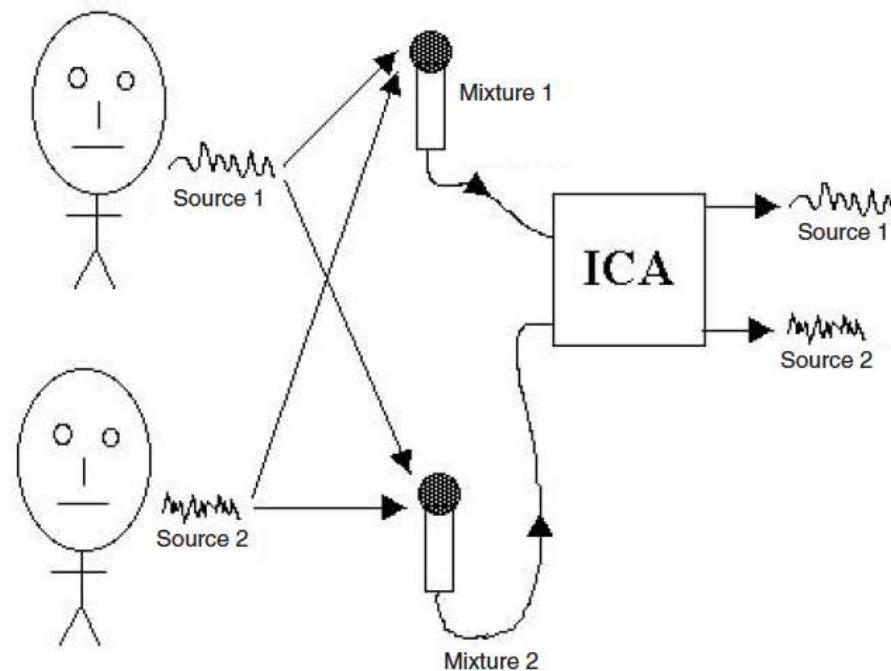
The Cocktail Party Problem

ICA “unmixes” the signals in each microphone to recover the speech of each speaker.

No information is needed about either the placement of speakers or microphones nor the content of the speeches.



ICA



ICA

If each voice signal is examined at a fine time scale then it becomes apparent that the amplitude of one voice at any given point in time is unrelated to the amplitude of the other voice at that time.

The reason that the amplitudes of the two voices are unrelated is that they are generated by two unrelated physical processes (i.e., by two different people).

If we know that the voices are unrelated then one key strategy for separating voice mixtures into their constituent voice components is to look for unrelated time-varying signals within these mixtures

ICA

ICA is based on a generative model: how the measured signals are produced.

The model assumes that the measured signals are the product of instantaneous linear combinations of the independent sources. Such a model can be stated mathematically as:

$$x_i(t) = a_{i1} s_1(t) + a_{i2} s_2(t) + \dots + a_{iN} s_N(t) \quad \text{for } i = 1, \dots, N$$

where $x(t)$ are the signals obtained from the sources, $s(t)$.

You have $x(t)$, but you want $s(t)$.

Note that this is a series of equations for the N different signal variables, $x_i(t)$, $i = 1, 2, \dots, N$

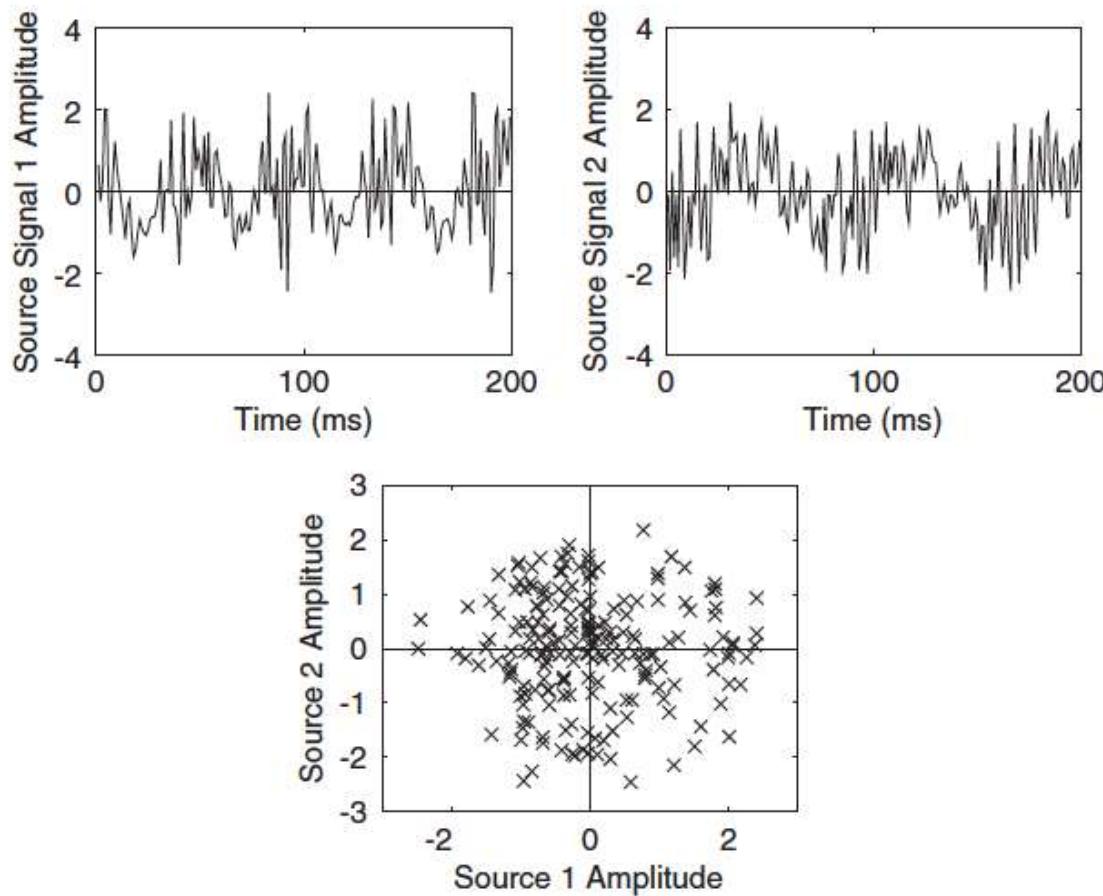
Independent Component Analysis Equations

In matrix form,
this equation
becomes:

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_N(t) \end{bmatrix} = A \begin{bmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_N(t) \end{bmatrix} \quad \text{or } \mathbf{x} = \mathbf{As}$$

Measured Hidden

where \mathbf{s} is a vector composed of all the source signals, \mathbf{A} is the mixing matrix composed of the constant elements a_{ij} , and \mathbf{x} is a vector of the measured signals.



ICA

ICA separates a set of signal mixtures into a corresponding set of statistically independent component signals or source signals .

These mixtures can be sounds, electrical signals, e.g., electroencephalographic (EEG) signals, or images (e.g., faces, fMRI data).

The defining feature of the extracted signals is that each extracted signal is statistically independent of all the other extracted signals

How Independent Component Analysis Works

A form of Blind Source Separation

- ICA is based on the simple assumption that if different signals are from different physical processes (e.g., different people speaking) then those signals are statistically independent.
- ICA takes advantage of the fact that this assumption can be reversed, leading to a new assumption:
 - i.e. if statistically independent signals can be extracted from signal mixtures then these extracted signals must be from different physical processes.

The result is separation of signal mixtures into statistically independent signals.

Set of source signals (s):
- 5 people speaking



Set of Signal
Mixtures (x)



Estimated extracted independent
components, each an estimate of 1
of the original (y)



Mixing Process (A)
- speaker-microphone distances



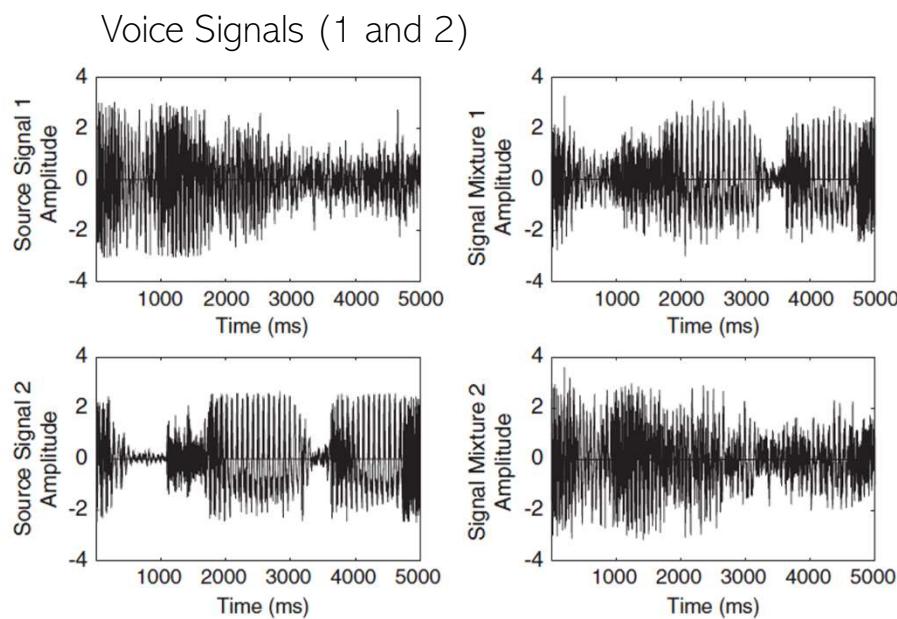
Un-mixing Process (W)

Mixing Signals

When two speech source signals are mixed to make two signal mixtures 3 effects follow:

- 1) Independence
- 2) Normality
- 3) Complexity

Each of these effects can be used as a basis for unmixing

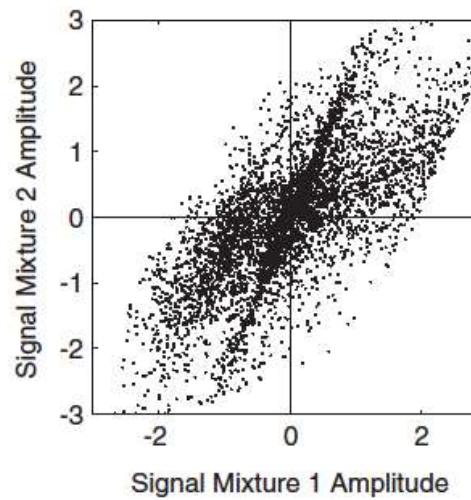
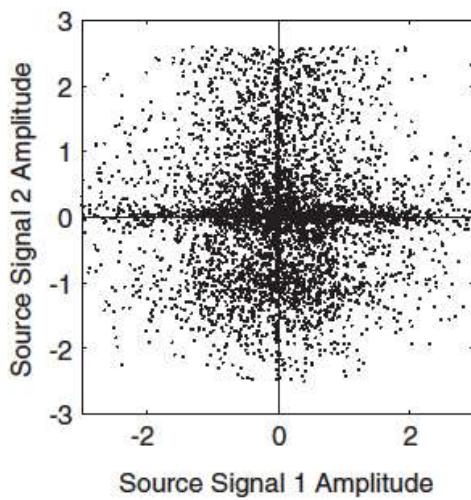


Mixtures (e.g.
different
microphone
placements)

Independence:

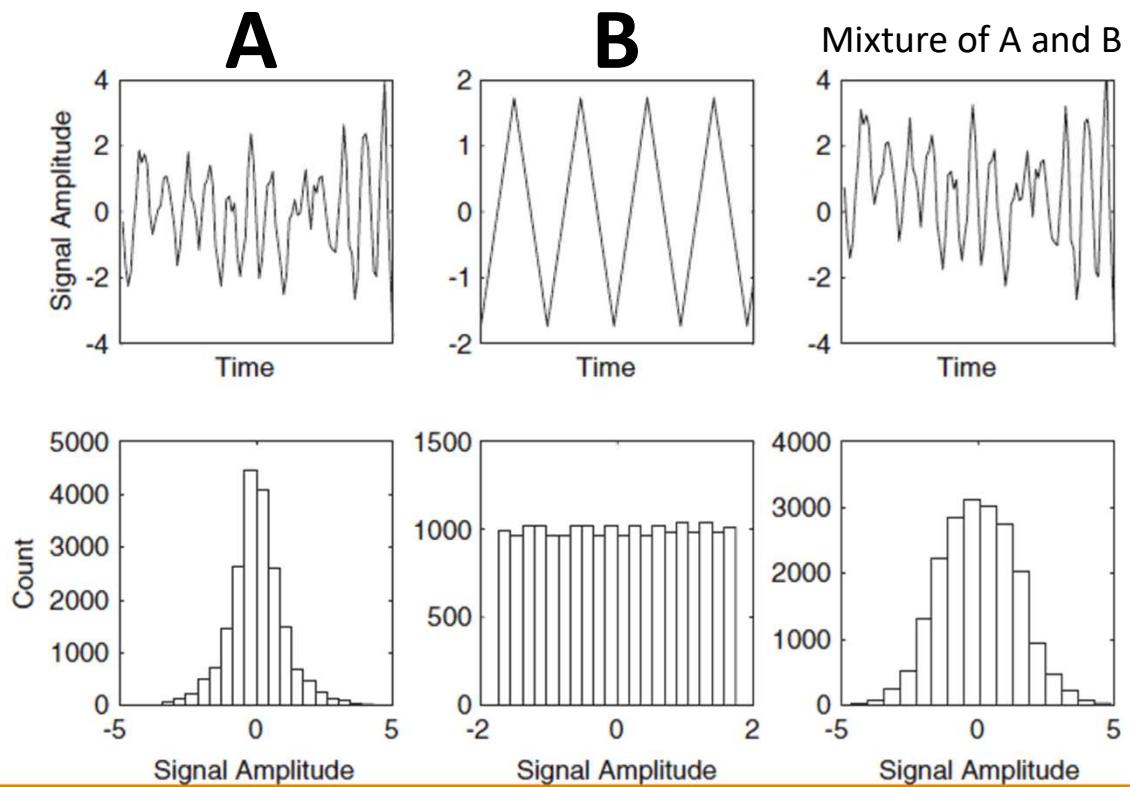
Whereas speech source signals are statistically independent, their signal mixtures are not.

- because each source signal is shared between both mixtures such that the resultant commonality between signal mixtures ensures that they cannot be independent.



Normality:

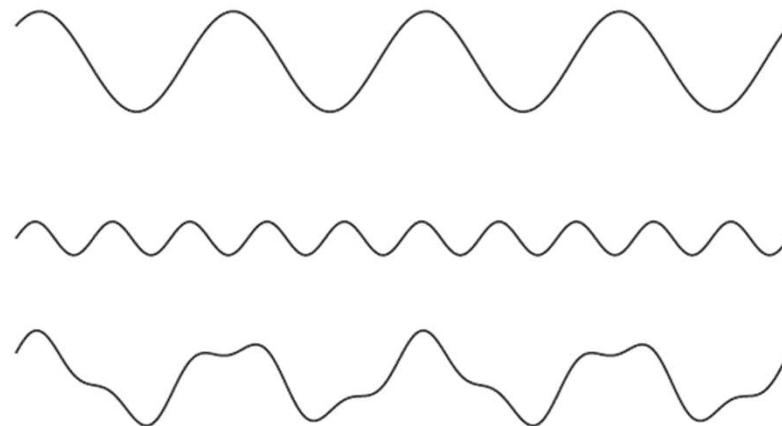
- If the values in a speech source signal are plotted as a histogram then a Gaussian shape emerges.
- Compare to non-speech histogram of a sawtooth signal yields flat histogram.



Complexity

The temporal complexity of any mixture is greater than (or equal to) that of its simplest (i.e., least complex) constituent source signal.

- This ensures that extracting the least complex signal from a set of signal mixtures yields a source signal.
- This complexity conjecture can be used as a basis for blind source separation.



Requirements for ICA

If source signals have some property X, and signal mixtures do not, then given a set of signal mixtures we should attempt to extract signals with “as much X as possible”.

Thus, replacing X with independence, normality and complexity yields principles of unmixing:

1) Independence

- If source signals are independent and signal mixtures are not then extracting independent signals from a set of signal mixtures should recover the required source signals

2) Normality

- If source signals are non-Gaussian and signal mixtures are not then extracting signals with non-Gaussian behaviour from a set of signal mixtures should recover the required signals.

3) Complexity

- If source signals have low complexity structure and signal mixtures do not then extracting signals with low complexity from a set of signal mixtures should recover the required signals.

For ICA:

- mixing process is specified in terms of a set of constants: mixing coefficients
- if these are known then they can be used to derive a set of unmixing coefficients, which can be used to extract source signals from signal mixtures.

$$s_1 = (s_1^1, s_1^2, \dots, s_1^N)$$

$$s_2 = (s_2^1, s_2^2, \dots, s_2^N)$$

Two time varying signals, where s is a time-varying signal which takes amplitudes s_1, s_2, \dots, s_N

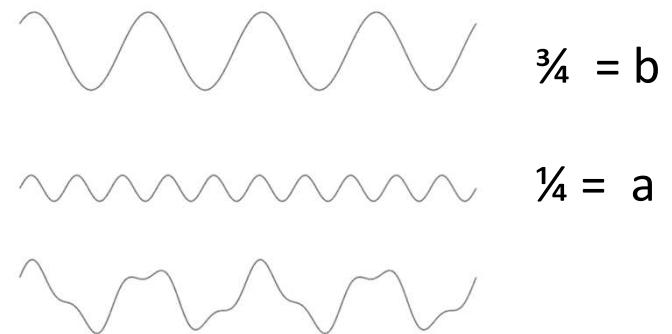
Mixing Signals

The different distance of each source from a microphone ensures that each source contributes a different amount to the mic's output x_1 .

As an example let's say the microphone-source distances are such that 1/4 of the mixture is from source s_1 and 3/4 are from source s_2 .

the mixture x_1 can be specified as a weighted sum of the 2 source signals.

So the mixture amplitude $x_1 t$ at a given time t is the weighted sum of the source signals $s_1 t$ and $s_2 t$ at that time:



Mixing Signals

$$(x_1^1, x_1^2, \dots, x_1^N) = a \times (s_1^1, s_1^2, \dots, s_1^N) + b \times (s_2^1, s_2^2, \dots, s_2^N)$$

Or in more compact form:

$$x_1 = as_1 + bs_2$$

What if another microphone is added?

$$x_2 = cs_1 + ds_2$$

the pair of signal mixtures (x_1, x_2) is analogous to the pair of source signals

Therefore (x_1, x_2) can be represented as a vector variable

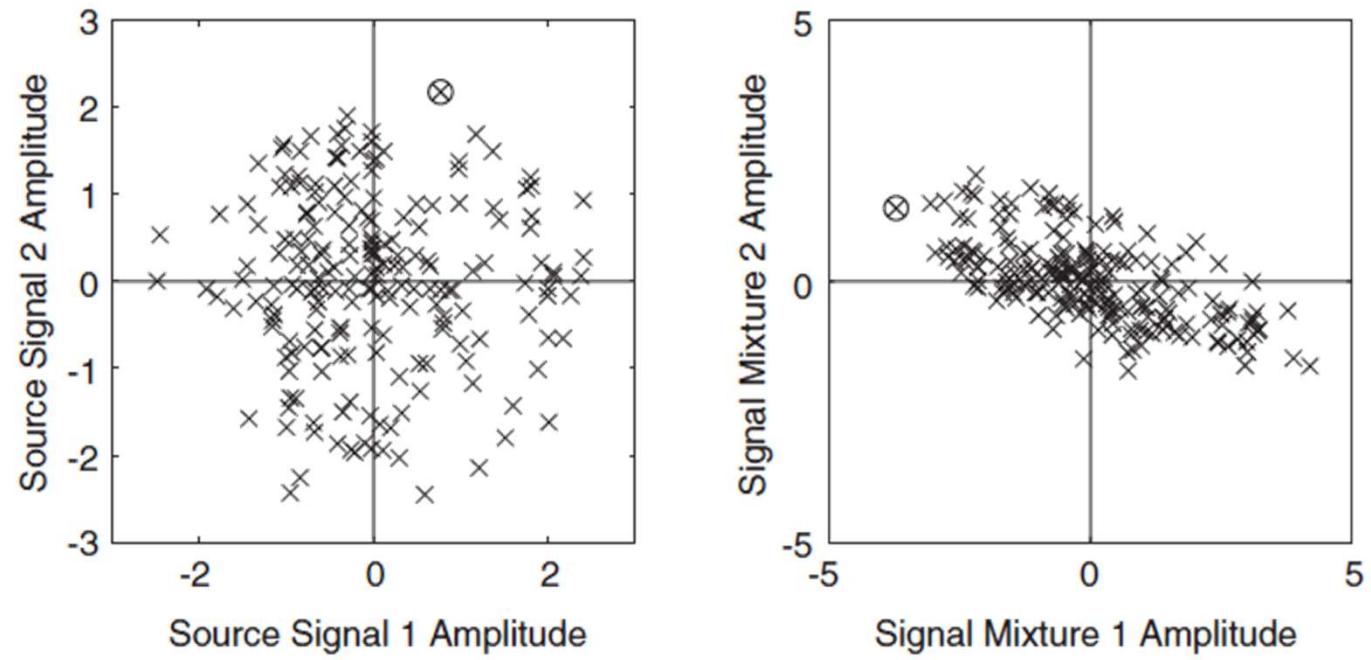
Mixing Signals

The mixing process, represented by the four mixing coefficients (a, b, c, d), transforms one vector variable s to another vector variable x

each source signal data point:

at a given time, t is transformed to a corresponding signal mixture data point:

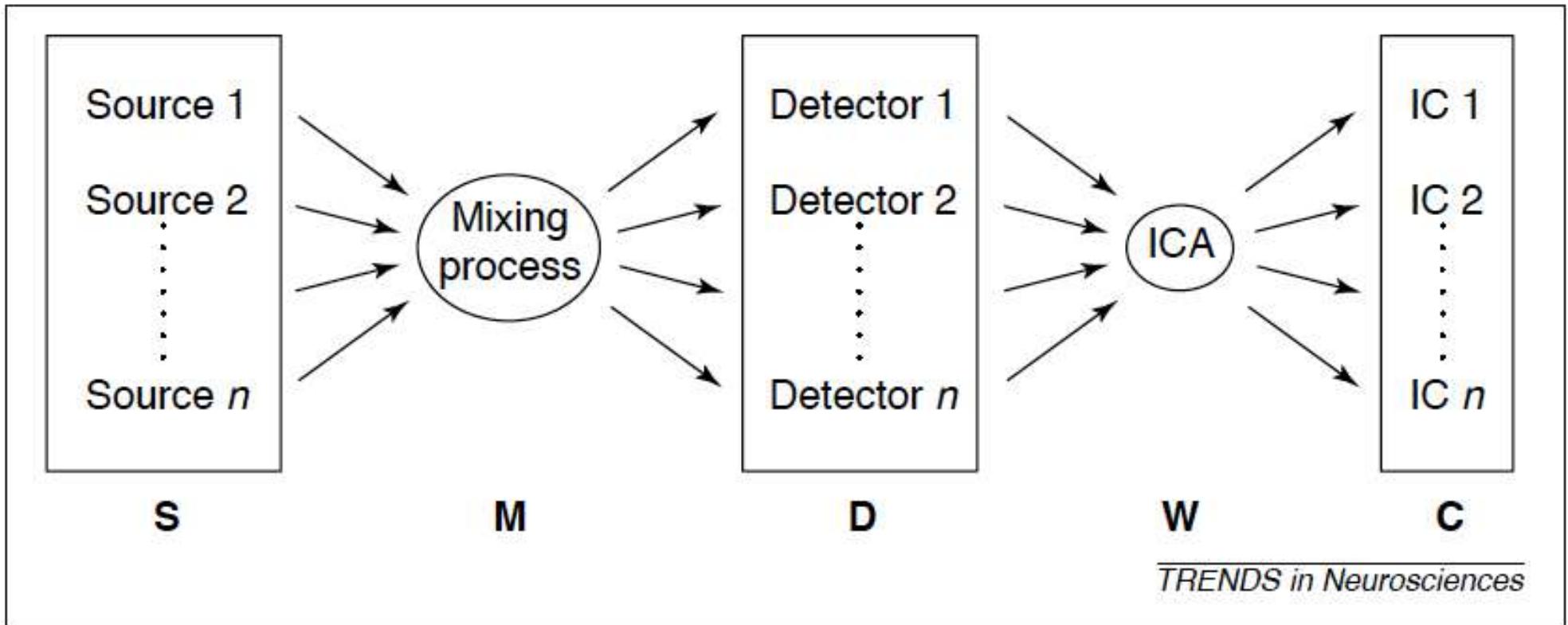
denoted as



Independent component analysis at the neural cocktail party

Glen D. Brown, Satoshi Yamada and Terrence J. Sejnowski

(Added paper to website for download)



TRENDS in Neurosciences

Example from the Article

Each sequence is a measured variable

And each column is a different time point

C in this case is the XOR of A and B

0 Covariance...

PCA?

Box 1. Minimizing mutual information

Consider the following three sequences of ones and zeros, labelled A, B and C:

- A** 111011101001000000101111000001011001...
- B** 00100110110001011011001001111111010...
- C** 110010000101010110011101011110100011...

Example from the Article

ICA?

There are 8 possible combinations of 3 binary digits (2^3)

A 111011101001000000101111000001011001...
B 001001101100010110110010011111110101...
C 110010000101010110011101011110100011...

X 000100010110111110100001111101001101...
Y 0011011110101010011000101000010111001...

This is called a 3rd order redundancy relationship, non linear

ICA can minimizes redundancy in data, no matter the order

Normally data with mutual information has covariance AND higher order dependencies

Example from the Article

Signals from each detector make up D

Each column of D is t

ICA will create a square matrix W($m=n=\#$ of detectors) so that

$$\begin{matrix} W & | & D \end{matrix} = \begin{matrix} C \end{matrix}$$

C are the independent components and they are forced to be as independent as possible

W is the unmixing matrix, each row is an unmixing function

Mixing Matrix

ICA techniques are used to solve for the unmixing matrix, A^{-1} , from which the independent components, s , can be obtained:

$$s = A^{-1}x$$

where A^{-1} is the unmixing matrix.

If you know the mixing matrix, A finding the unmixing matrix is easy. But usually you do not know A .

If the measured signals, x , are related to the underlying source signals, s , by a linear transformation (i.e., a rotation and scaling operation), then some inverse transformation (rotation/scaling) should recover the original signals.

Finding the Unmixing Matrix: A^{-1}

To estimate the mixing matrix, ICA needs to make basic two assumptions:

1. The source variables, s , are truly independent.
2. They are non-Gaussian.

Both conditions are usually met when the sources are real signals.

A third restriction is that the mixing matrix must be square: the number of sources should equal the number of measured signals.

ICA Limitations

ICA can only be applied to non-Gaussian signals because it relies on higher-order statistics to separate the variables.

- Higher-order statistics (i.e. moments and related measures) of Gaussian signals are zero.

Since ICA has available only the measured variables there are some limits to what ICA can do:

- ICA cannot determine the variances, hence the energies or amplitudes, of the actual sources.
- Unlike PCA, the order of the components cannot be established. (Logical if amplitudes cannot be determined.)

ICA Algorithms

The unmixing matrix is determined by optimizing some “objective function” related to independence of data.

There a large number of ICA algorithms that have been developed. Many are available as downloadable MATLAB files.

All approaches use optimization. They differ in:

1. The specific **objective function** that is optimized .
2. The **optimization** method that is used.

Optimization Criteria

One of the most intuitive approaches uses an objective function that is related to the non-Gaussianity of the data set.

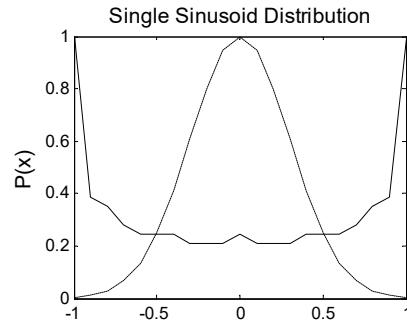
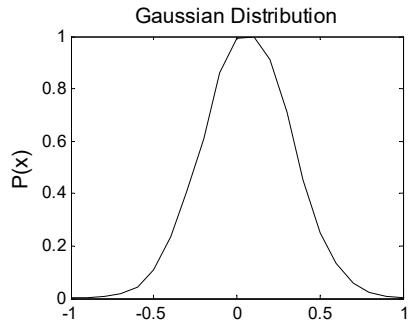
This approach takes advantage of the fact that mixtures tend to be more Gaussian than the distribution of independent sources.

Mixtures of non-Gaussian sources will be more Gaussian than the unmixed sources

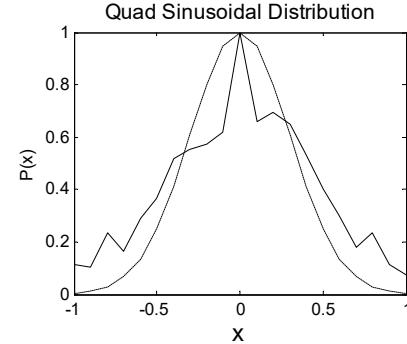
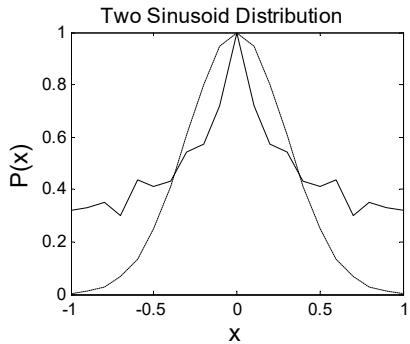
Central Limit Theorem

the sum of k independent, identically distributed random variables converges to a Gaussian distribution as k becomes large

regardless of the distribution of the individual variables.



The distribution of a single sinusoid is markedly non-Gaussian.

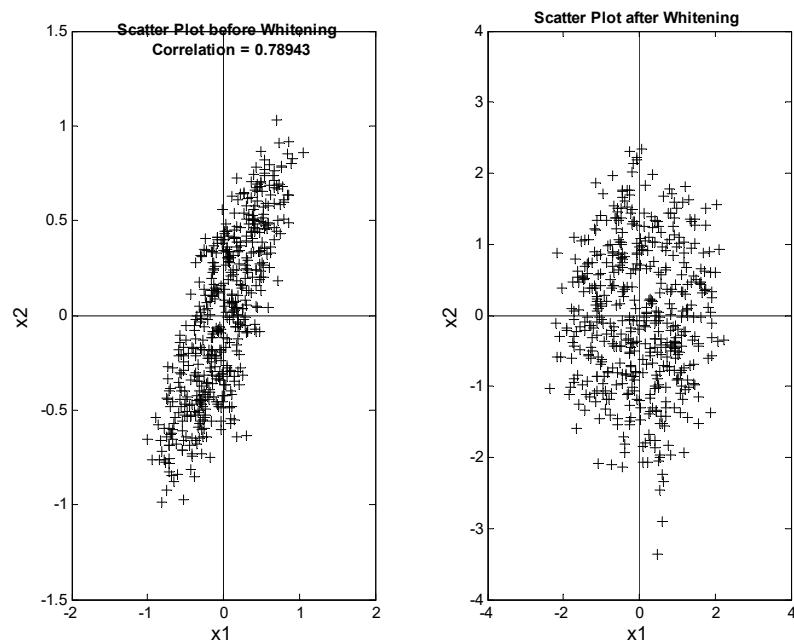


Mixtures of even two sinusoids have distributions that look Gaussian.

Data Whitening

Most ICA algorithms begin with
“data whitening:”

- o Centering the data (zero means).
- o Decorrelating the data (PCA)
- o Scaling the data so the variances equal to 1.0



ICA Algorithm (continued)

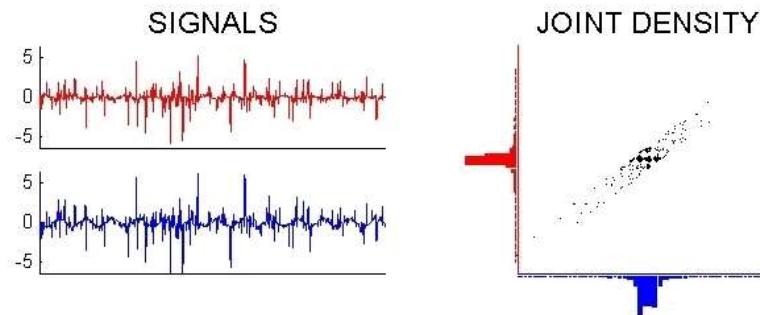
One approach to quantifying non-Gaussianity is to use kurtosis:

For data with zero mean: $E\{x^4\} = \sigma^4$ and if the data are whitened, $\sigma^2 = 1$, so the kurtosis equation becomes

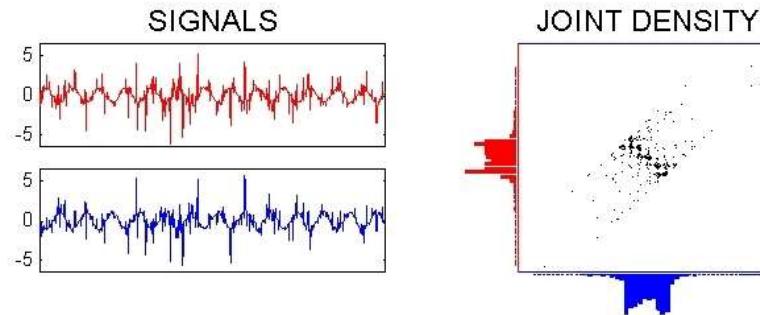
Using a metric that involves the 4th power greatly enhances the influence of outliers so a nonlinear function is often used to compress the data before taking the 4th power

Example Using NonGaussianity

Original

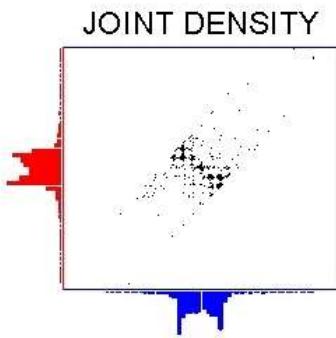
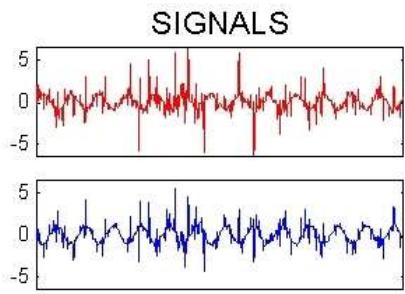


Whitened

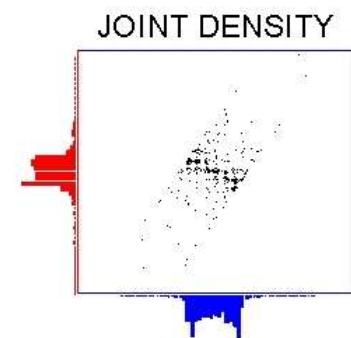
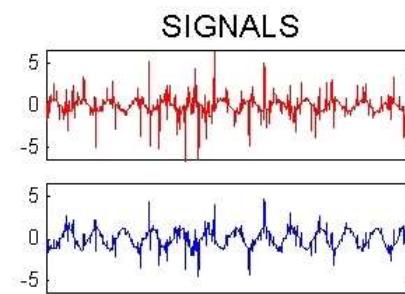


Whitened signals and density

ICA Process

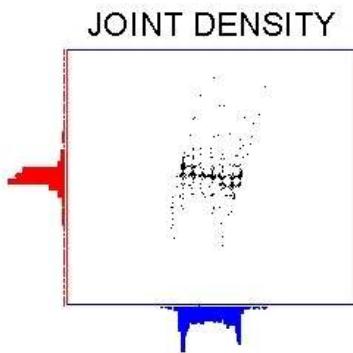
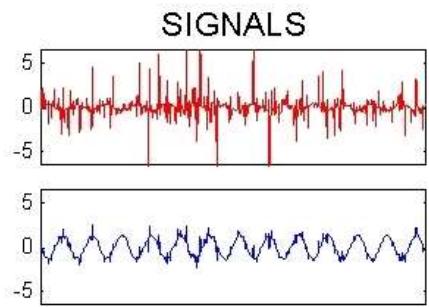


Separated signals after 1 step of FastICA

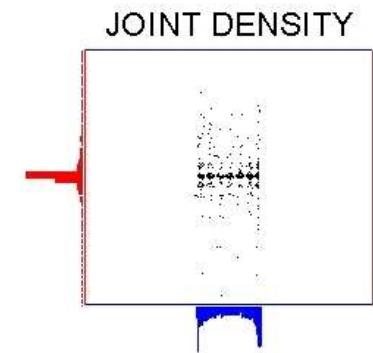
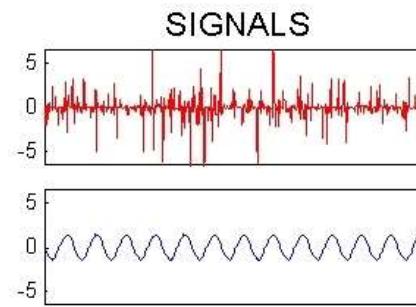


Separated signals after 2 steps of FastICA

ICA Process

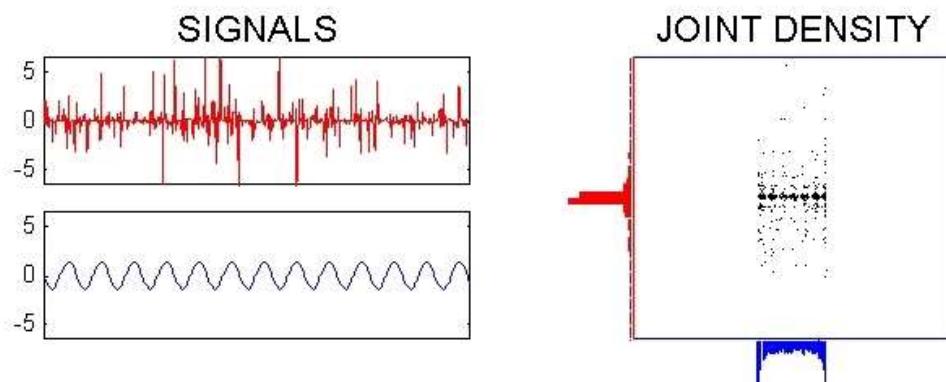


Separated signals after 3 steps of FastICA

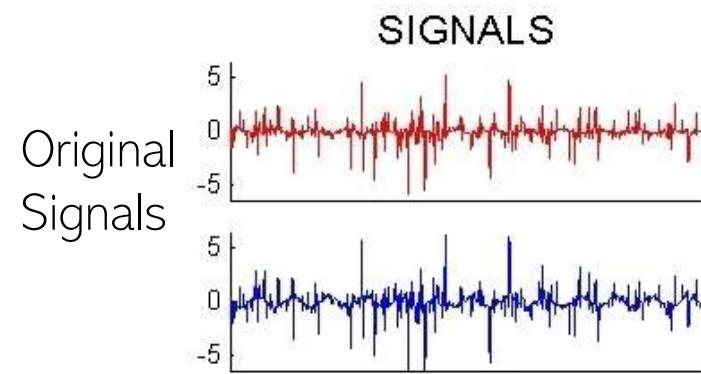


Separated signals after 4 steps of FastICA

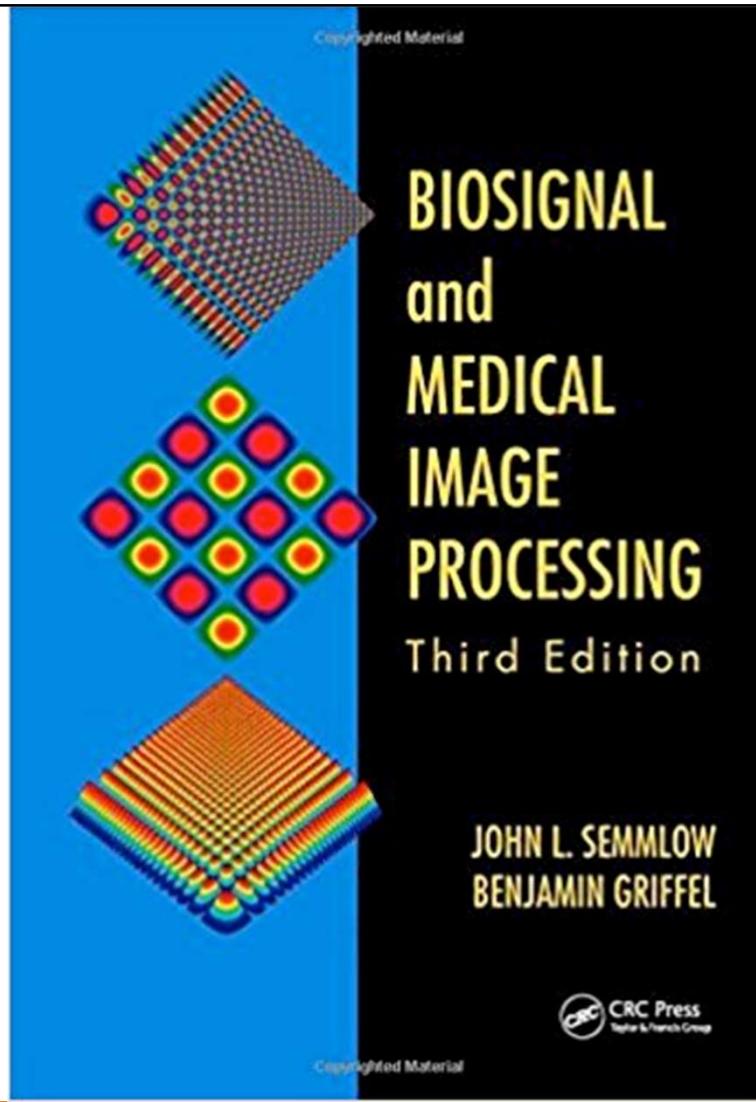
Final Rotation



Separated signals after 5 steps of FastICA



PCA/ICA Reference
material in this lecture:



Lecture 7

TAYLOR DEVET MASC.

PHD. CANDIDATE BIOLOGICAL AND BIOMEDICAL ENGINEERING
MCGILL UNIVERSITY

SHRINERS HOSPITAL FOR CHILDREN

Todays Aims...



Chronobiology



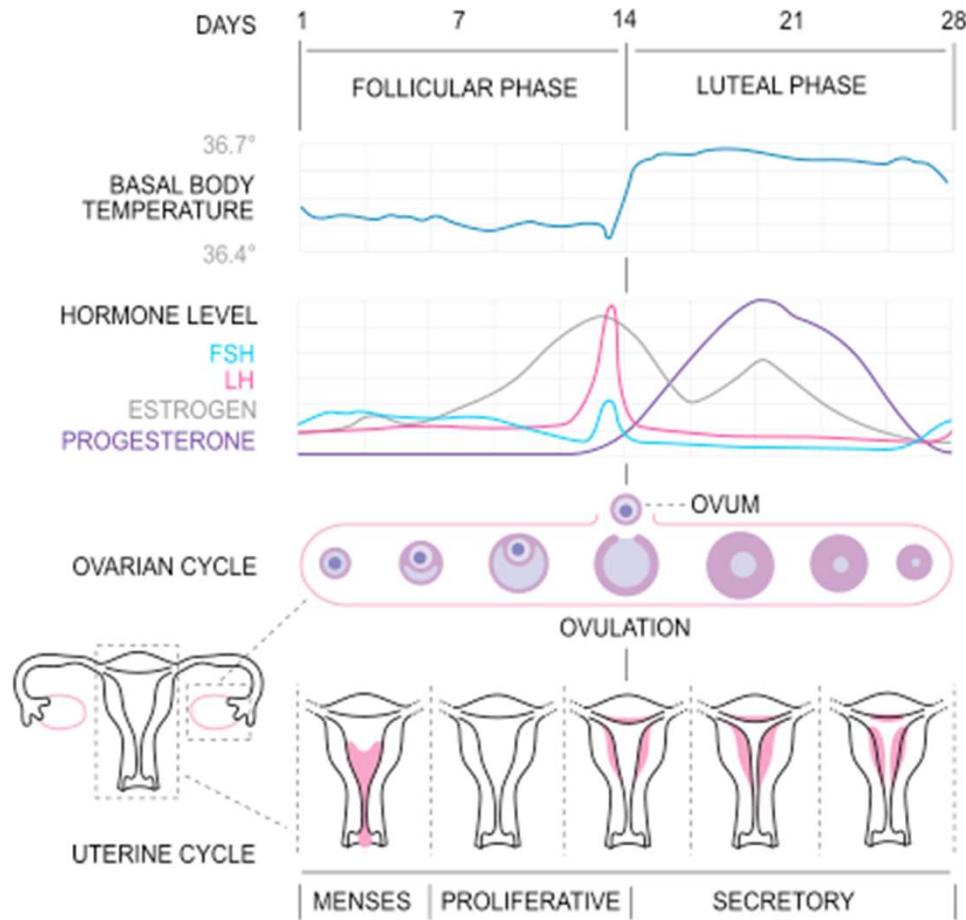
Linear Time Invariance



Correlation and
Coherence

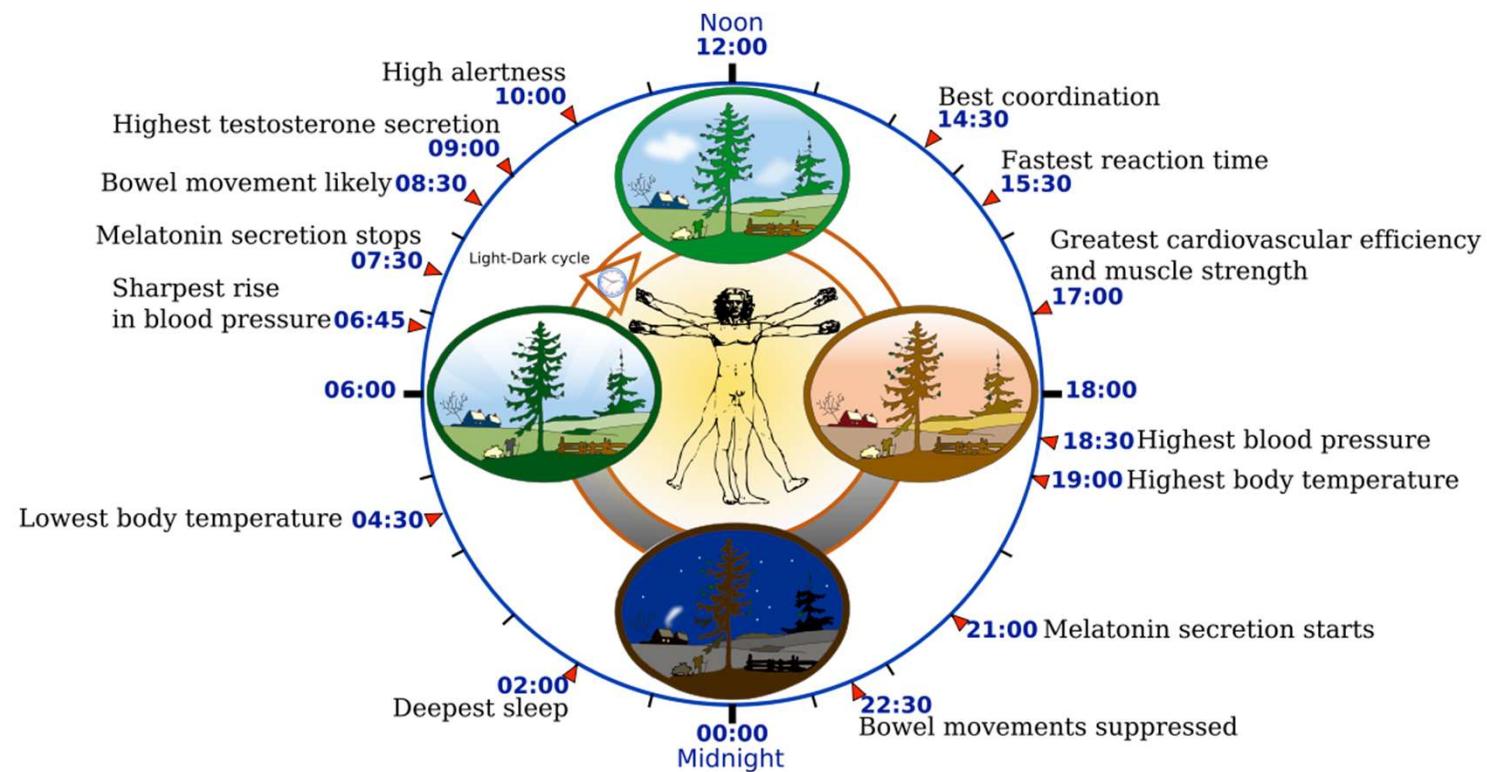
Measuring Physiological Signals in the Time Domain

- linear/multiple regression
- multivariate statistics (PCA/ICA)
- cosinor analysis
- autocorrelation analysis
- Fourier
- Wavelets
- Fractal and Chaos analysis
- etc.



Physiological Signals are Cyclical

Physiological Signals are Cyclical

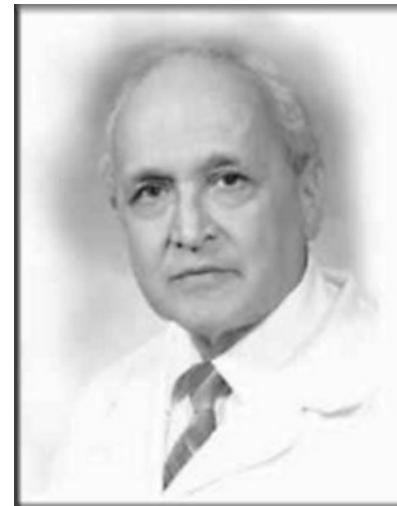


Chronobiology

Biology of periodic(cyclic) and other time dependent processes in living organisms
biological rhythms

Franz Halberg

In the 1950s, he introduced the word circadian
Father of Chronobiology



Circadian rhythms

Circa – around

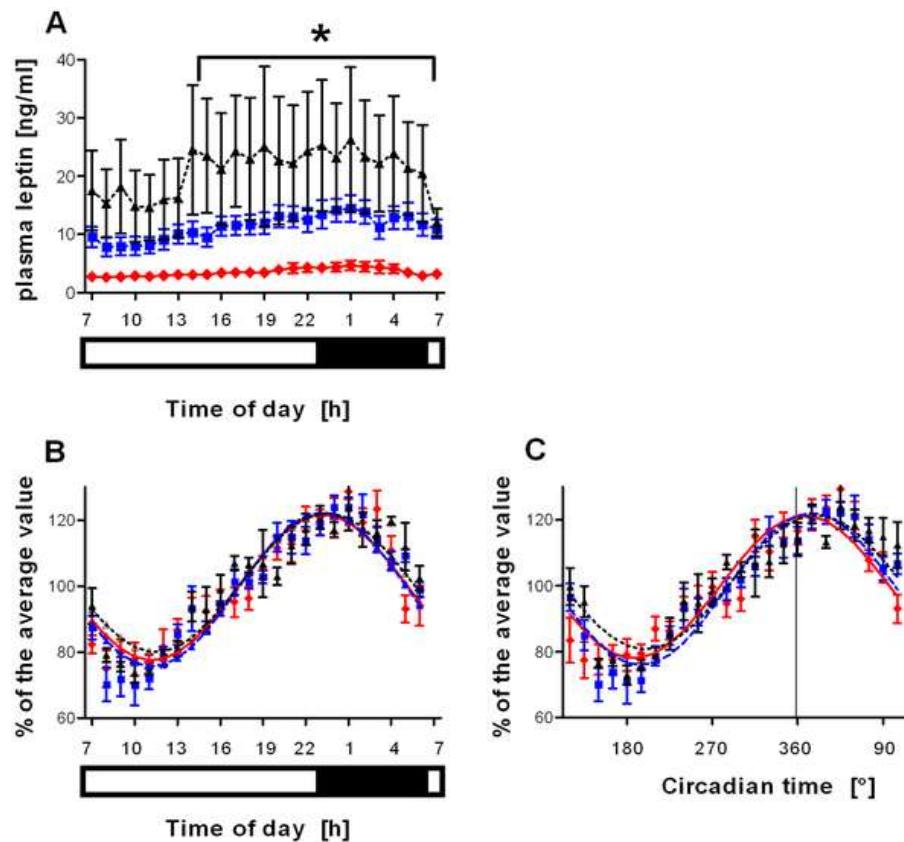
Diem - day

Biological rhythms with durations ~24 hours. Many behavioral and autonomic processes of organisms exhibit circadian rhythmicity

Diurnal rhythm

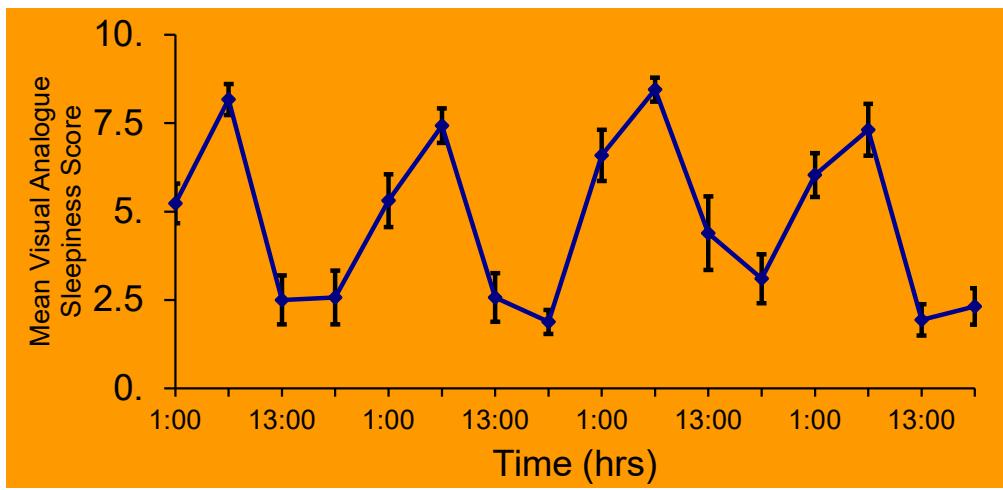
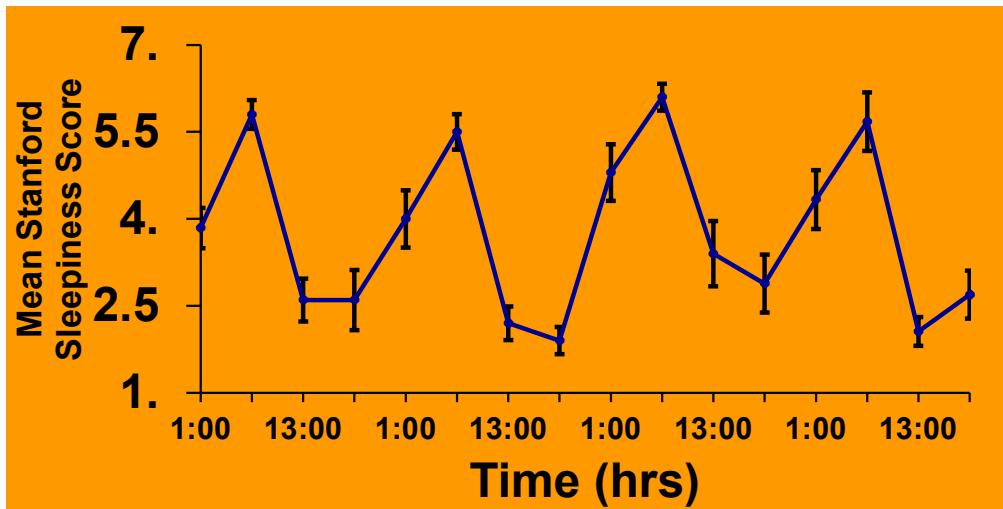
- a cycle that is synchronized/correlated with the day/night cycle
- Diurnal rhythms are circadian but not all circadian are diurnal

Diurnal rhythms of plasma leptin concentrations

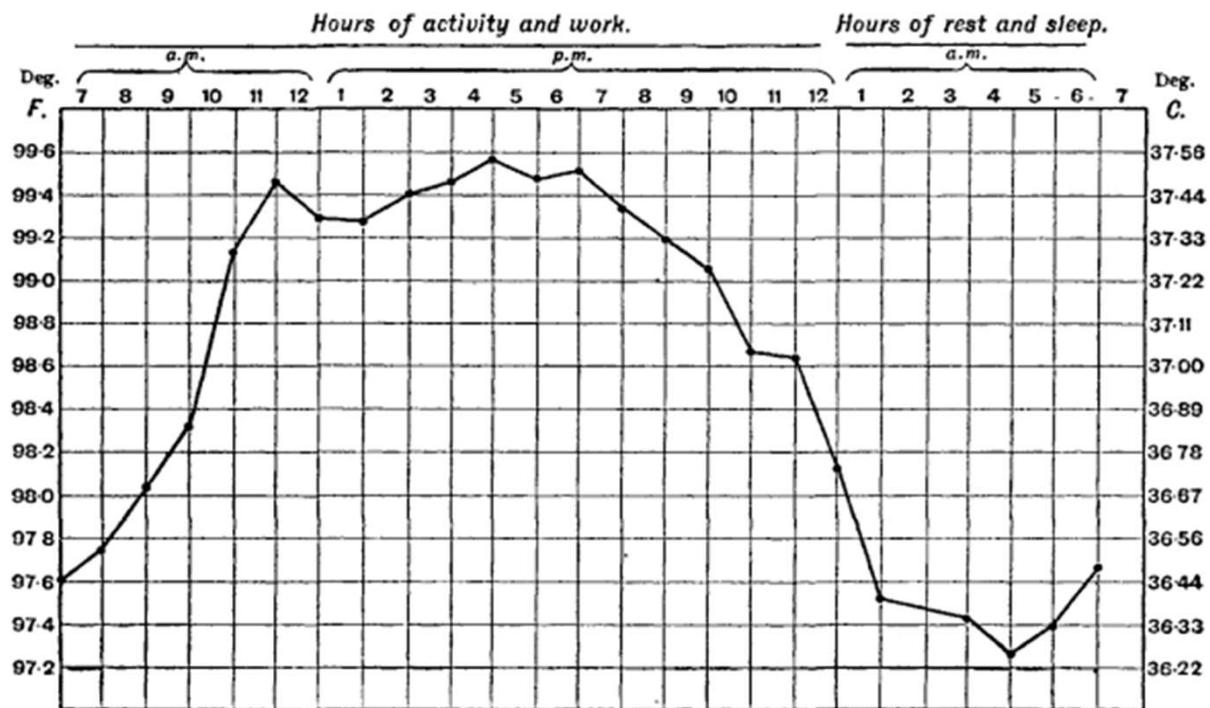


PLOS ONE

Mäntele S, Otway DT, Middleton B, Bretschneider S, et al. (2012) Daily Rhythms of Plasma Melatonin, but Not Plasma Leptin or Leptin mRNA, Vary between Lean, Obese and Type 2 Diabetic Men. PLoS ONE 7(5): e37123. doi:10.1371/journal.pone.0037123
<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0037123>



Basal body Temperature: Daily Rhythm



Ultradian rhythms

Biological rhythms with durations shorter than circadian rhythms (i.e., < ~19 hours)

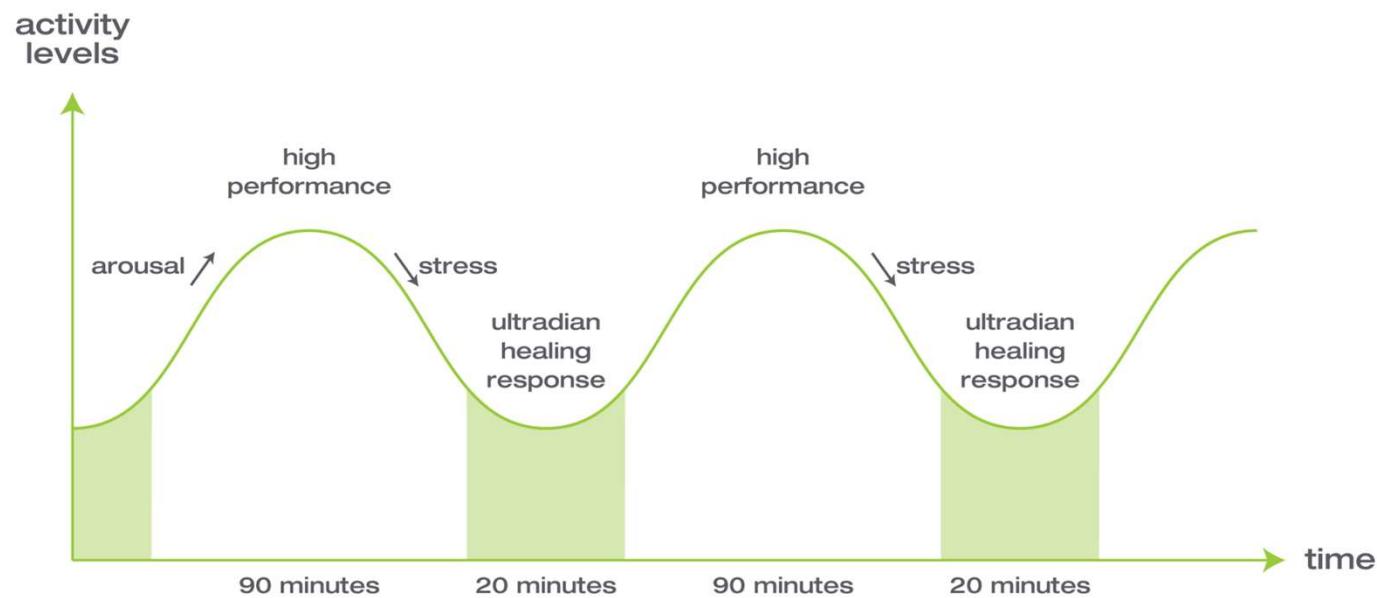
Rhythms longer than an hour

e.g. cardiac, respiratory, neuroendocrine, gastrointestinal, tidal, and other rhythms.

Many are endogenously generated by some sort of pacemaker

only tidal rhythms are regularly synchronized to environmental cycles.

General Human Performance



<https://www.kosmotime.com/ultradian-rhythm/>

Infradian rhythms

Biological rhythms with durations greater than circadian rhythms (i.e. $> \sim 28$ hours).

e.g. estrous, weekly, lunar, annual

Many infradian rhythms are endogenously generated by some sort of pacemaker
only lunar and annual rhythms can be fully synchronized to environmental cycles.

Infradian rhythms

THE EFFECTS OF THE FULL MOON ON HUMAN BEHAVIOR*

Edgecliff College

JODI TASSO AND ELIZABETH MILLER¹

SUMMARY

Data were gathered in a large metropolitan area over a period of one year as to nine categories of 34,318 criminal offenses committed during the phases of the full moon and non full moon. It was found that the eight categories of rape, robbery and assault, burglary, larceny and theft, auto theft, offenses against family and children, drunkenness, and disorderly conduct occurred significantly more frequently during the full moon phase than at other times of the year. Only the category of homicide did not occur more frequently during the full moon phase. The results support further exploration and research related to cosmic influences on man's behavior.

Infradian rhythms



The American Journal of Emergency Medicine

Volume 14, Issue 2, March 1996, Pages 161-164



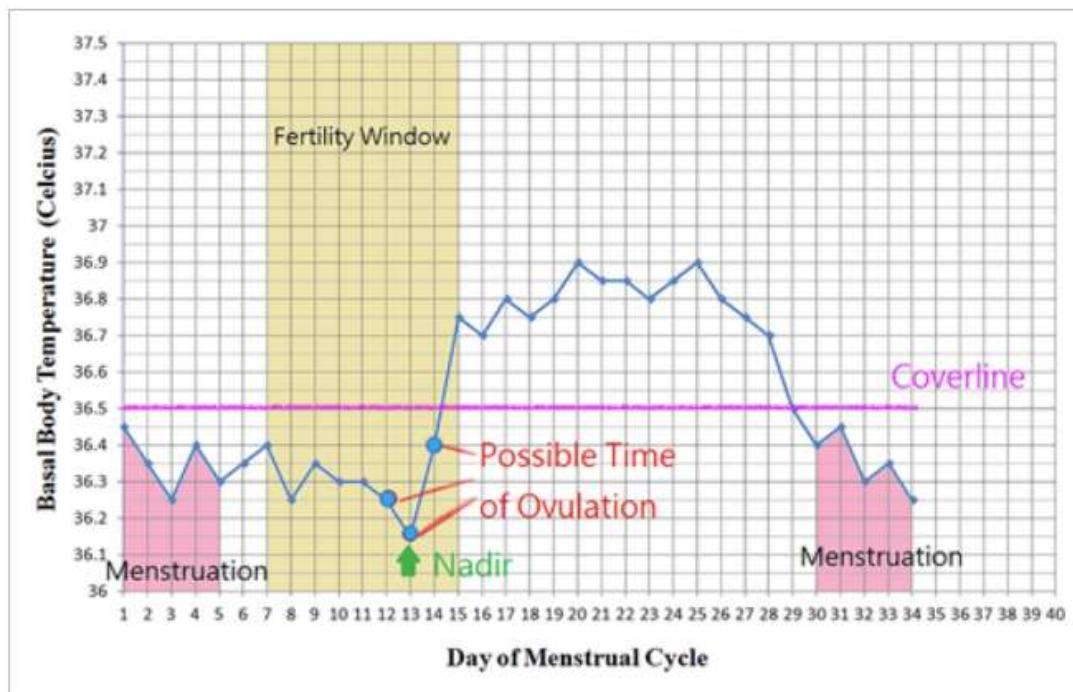
Original contribution

The full moon and ED patient volumes: Unearthing a myth

David A Thompson MD ^{8,*}, Stephen L Ad

ambulance. A total of 35,087 patients was admitted to the hospital and 11,278 patients were admitted to a monitored unit. No significant differences were found in total patient visits, ambulance runs, admissions to the hospital, or admissions to a monitored unit on days of the full moon. The occurrence of a full moon has no effect on ED patient volume, ambulance runs, admissions, or admissions to a monitored unit.

Infradian Rhythms



<https://aiche.onlinelibrary.wiley.com/doi/full/10.1002/btm2.10058>

Annual rhythms

A class of infradian rhythms that fluctuate on a yearly basis

e.g. body mass, cold-induced thermogenesis, food intake, heterothermy, melatonin secretion, molting, and reproductive capacity.

Many, but not all, are synchronized to annual environmental cycles (called circannual rhythms).

Migration of birds

Spring babies

Other Rhythms

- other cycles having periods of \sim 1 week, \sim 1 month, and \sim 1 year are also ubiquitous, as are some other newly discovered cycles with periods of about 5 and 16 months, and much longer periods.
- biological cycles are typically synchronized by environmental cycles (e.g. lighting and feeding schedules).
- More generally, environmental geophysical cycles such as the day-light cycle, tidal motions, moon phases, seasonal variation (circannual) are the dominant rhythms

Chronobiology Study Designs:

Longitudinal sampling

- obtaining data on the same individual (experimental unit) as a function of time
- (e.g. 24hr blood pressure monitoring of blood pressure every minute intervals for 7 days).

Transverse (cross-sectional) sampling

- obtaining only one value per individual (experimental unit),
- with different individuals providing data at the same or different sampling times.

Hybrid (linked cross-sectional) sampling

- taking serial measurements from several individuals (experimental units).
- e.g. circulating prolactin determined at 20-minute

intervals for 24 hours in women at low or high familial risk of developing breast cancer later in life.

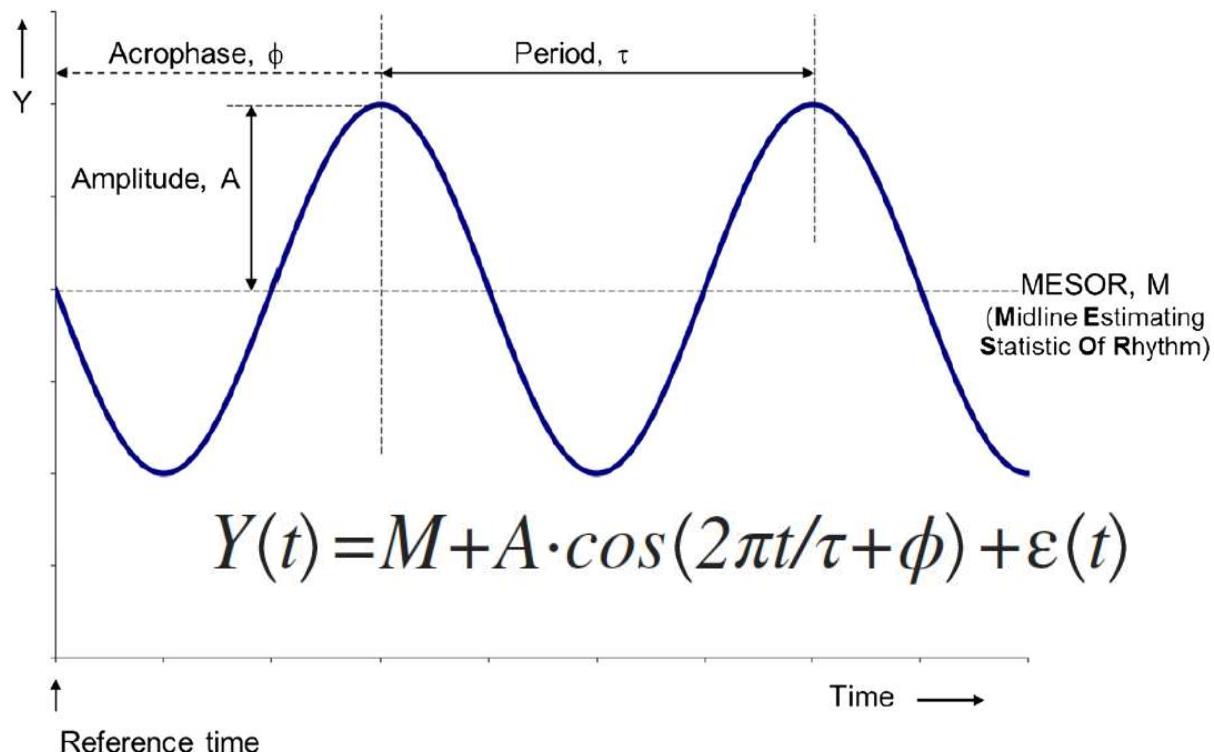
Chronobiology Study Design

- when sampling is performed on more than a single individual, it is important that they are synchronized
- Synchronizers are often environmental periodicities determining the temporal placement of biological rhythms.
- e.g. rest-activity or light–dark schedules (photoperiod) can be used to determine a reference time

Cosinor Analysis

- used in chronobiology
- a type of regression problem (i.e. least squares minimization)
- the biggest advantage over other time based analytical methods is this is applicable to non-equidistant data
- Assessment of non-random variations as a function of time at the cellular level, in tissue culture, as well as in multi-cellular organisms at different levels of physiologic organization
- biological time structure covers many different ranges from fractions of seconds in single neurons to seconds in the cardiac and respiratory cycles, and a few hours in certain endocrine functions.

Cosinor Analysis



Important Components

M = MESOR (Midline Statistic Of Rhythm), a rhythm-adjusted mean

A = amplitude (a measure of half the extent of predictable variation within a cycle)

ϕ = acrophase (a measure of the time of overall high values recurring in each cycle)

τ = cycle period

$\varepsilon(t)$ = error term

$$Y(t) = M + A \cdot \cos(2\pi t/\tau + \phi) + \varepsilon(t)$$

Cosinor Forms

$$Y(t) = M + A \cdot \cos(2\pi t/\tau + \phi) + \varepsilon(t)$$

- can be re-written as:

$$Y(t) = M + \beta x + \gamma z + \varepsilon(t)$$

- where:

$$\begin{aligned}\beta &= A \cdot \cos \phi & x &= \cos\left(\frac{2\pi t}{\tau}\right) & z &= \sin\left(\frac{2\pi t}{\tau}\right) \\ \gamma &= -A \sin \phi\end{aligned}$$

Least Squares Model

This is a least squares minimization problem.

So, minimize the residual sum of squares (RSS):

i.e. minimize the squared differences between measures (Y_i) taken at times t_i

- Estimates for M , β , and γ are obtained by solving the normal equations, obtained by expressing that RSS is minimal when its first-order derivatives with respect to each parameter are zero.

In Matrix Form

Therefore the normal equations are:

$$\begin{aligned}\sum Y_i &= MN + \beta \sum x_i + \gamma \sum z_i \\ \sum Y_i x_i &= M \sum x_i + \beta \sum x_i^2 + \gamma \sum x_i z_i \\ \sum Y_i z_i &= M \sum z_i + \beta \sum x_i z_i + \gamma \sum z_i^2\end{aligned}$$

These normal equations in matrix form:

$$\begin{pmatrix} \sum Y_i \\ \sum Y_i x_i \\ \sum Y_i z_i \end{pmatrix} = \begin{pmatrix} N & \sum x_i & \sum z_i \\ \sum x_i & \sum x_i^2 & \sum x_i z_i \\ \sum z_i & \sum x_i z_i & \sum z_i^2 \end{pmatrix} \begin{pmatrix} M \\ \beta \\ \gamma \end{pmatrix}$$

In Matrix Form

$$\begin{pmatrix} \sum Y_i \\ \sum Y_i X_i \\ \sum Y_i Z_i \end{pmatrix} = \begin{pmatrix} N & \sum X_i & \sum Z_i \\ \sum X_i & \sum X_i^2 & \sum X_i Z_i \\ \sum Z_i & \sum X_i Z_i & \sum Z_i^2 \end{pmatrix} \begin{pmatrix} M \\ \beta \\ \gamma \end{pmatrix}$$

i.e. $\hat{W} = \hat{X}\hat{b}$

$$\hat{b} = X^{-1} W$$

(K is # parameters to estimate)

Least Squares Model

- For rhythm detection, the total sum of squares (TSS) is partitioned into the sum of squares due to the regression model (MSS) and the residual sum of squares (RSS).

TSS = sum of squared differences between the data and the arithmetic mean.

MSS = sum of squared differences between the estimated values based on the fitted model and the arithmetic mean.

$$TSS = MSS + RSS \text{ or } \sum(Y_i - \bar{Y})^2 = \sum(\hat{Y}_i - \bar{Y})^2 + \sum(Y_i - \hat{Y}_i)^2$$

Calculate F-statistic to assess whether
the model is appropriate

k = # of parameters estimated (3)

n = number of samples

Least Squares Model

For parameter estimation, consider M and (β, γ) separately:

Calculate the confidence interval (based on $1-\alpha=95\%$)

$$\hat{M} \pm t_{\alpha/2, (N-k)} \hat{\sigma} \sqrt{X_{II}^{-1}}$$

where X_{jj}^{-1} are elements of X^{-1} and:

$$\hat{\sigma} = \sqrt{RSS/(N-k)}$$

Least Squares Model

The covariance matrix (β, γ) is given by:

$$\hat{\sigma} \begin{pmatrix} X_{22}^{-1} & X_{23}^{-1} \\ X_{32}^{-1} & X_{33}^{-1} \end{pmatrix}$$

$$\sum (x_i - \bar{x})^2 (\beta - \hat{\beta})^2 + 2 \sum (x_i - \bar{x})(z_i - \bar{z})(\beta - \hat{\beta})(\gamma - \hat{\gamma})$$

where

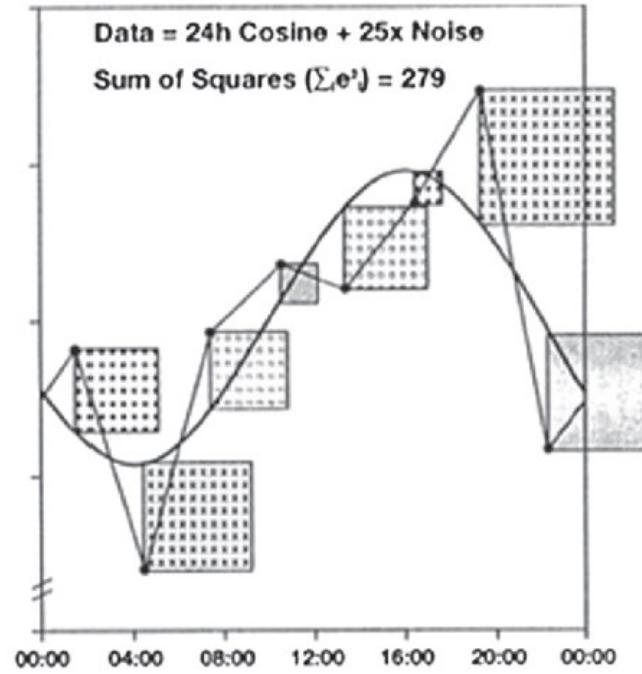
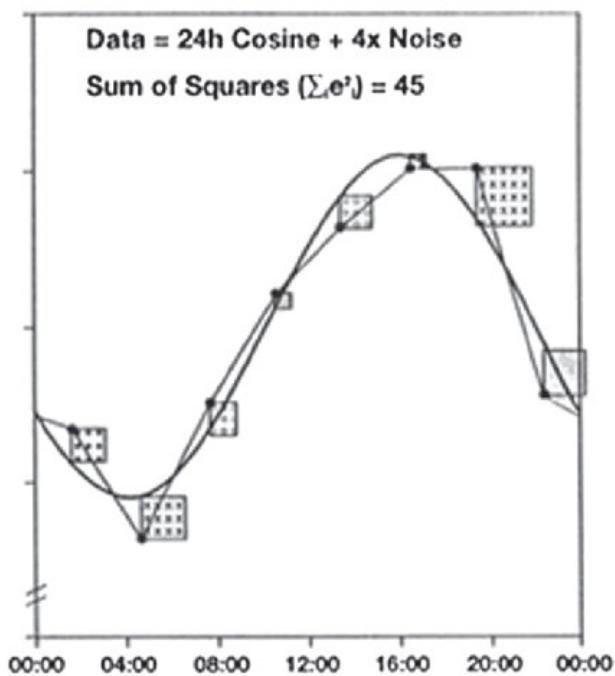
$$\bar{x} = \left(\sum_i x_i \right) / N$$

$$+ \sum (z_i - \bar{z})^2 (\gamma - \hat{\gamma})^2 \leq 2\hat{\sigma}^2 F_{1-\alpha, k-1, n-k}$$

and

$$\bar{z} = \left(\sum_i z_i \right) / N$$

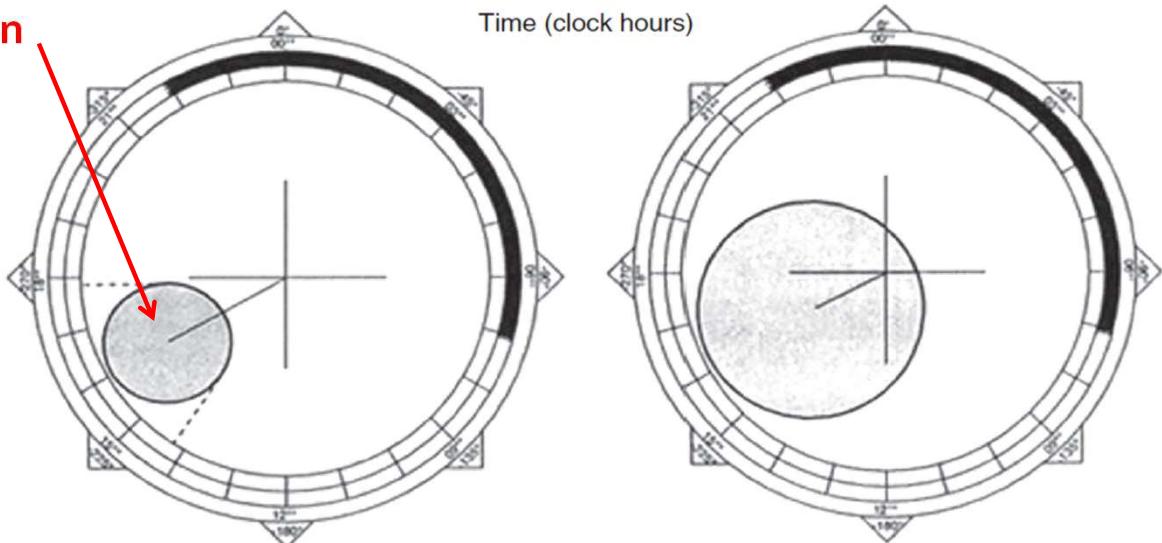
Least Squares Model



Cosinor Representation

confidence region

When the error ellipse covers the pole, the null hypothesis of no-rhythm (zero amplitude) is accepted (right).



Mesor = 120.00 ± 1.06
Amplitude = $10.98 (5.88 \text{ } 16.09)$
Acrophase = $-240^\circ (-213 \text{ } -268)$

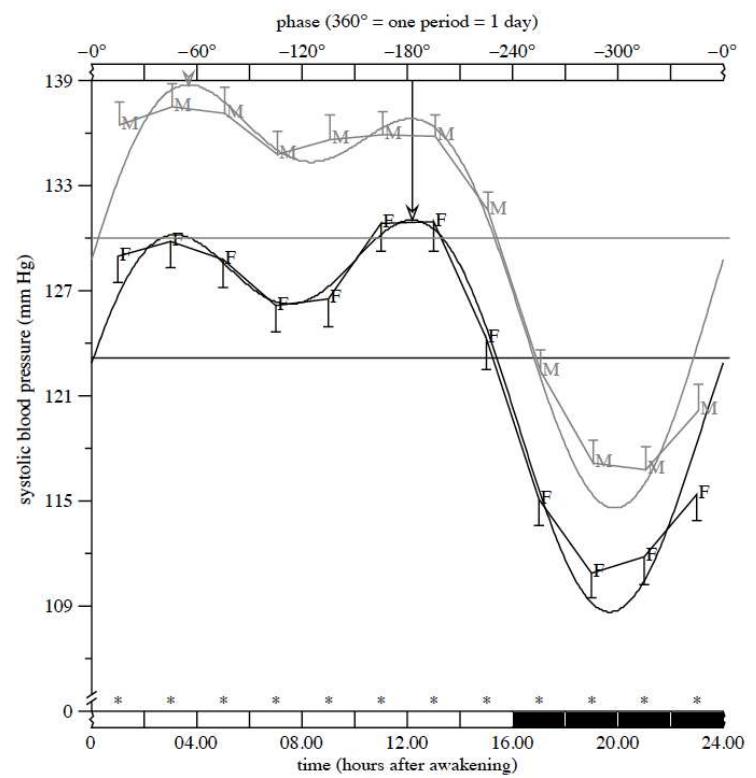
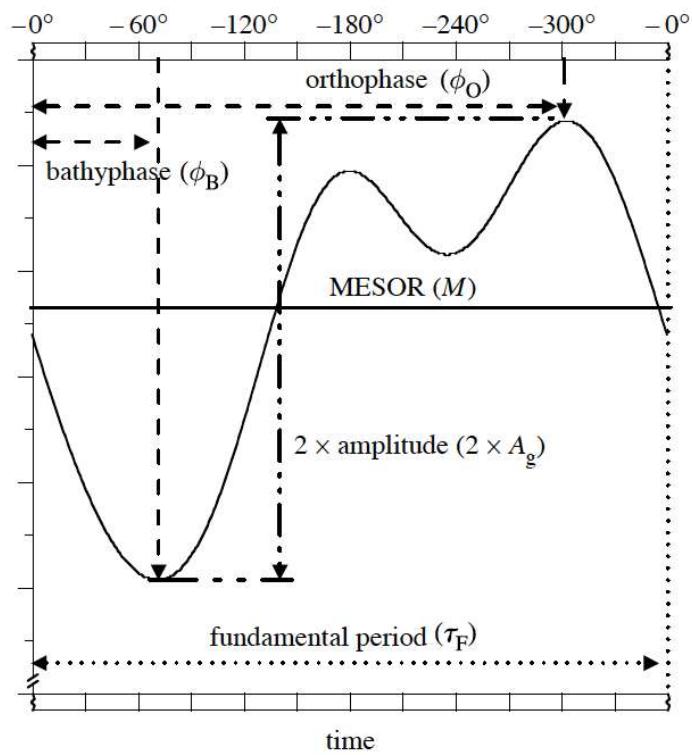
Mesor = 120.00 ± 2.64
Amplitude = $9.47 ($ $)$
Acrophase = $-241^\circ ($ $)$

Multiple Rhythms

$$y_n = M + \sum_{c=1}^C A_c \cos(\omega_c t_n + \phi_c) + e_n; \quad n = 1, \dots, N,$$

- where y_n is the observed value at time t_n (not necessarily equidistant) of the studied variable
- C is the number of sinusoidal components
- ω_c are the angular frequencies, i.e. $\omega_c = 2\pi/\tau_c$, where τ_c , with $c=1, \dots, C$, are the fitted periods
- N is the number of observed values (sample size)

Multiple Rhythms



Other Sinusoidally Varying Biological Data:

Heart Rate

- Heart Rate Variability (HRV)
- Heart Rate Variability (HRV) Lab Analyzes Data from Continuous Electronically-Stored ECGs

Holter device





Wearable Sensors: Over 900 Devices

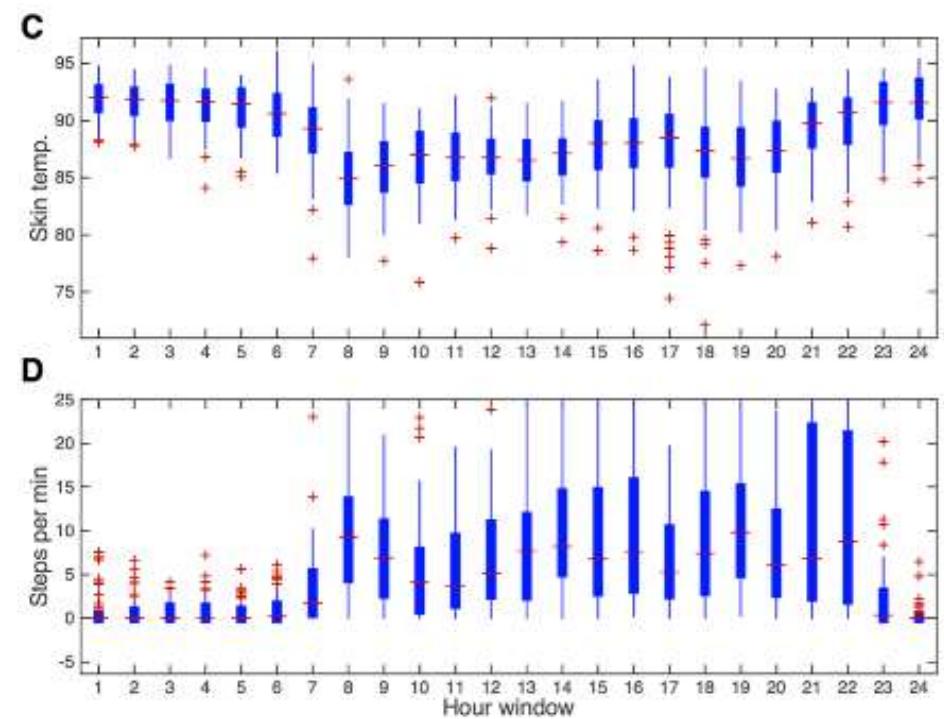
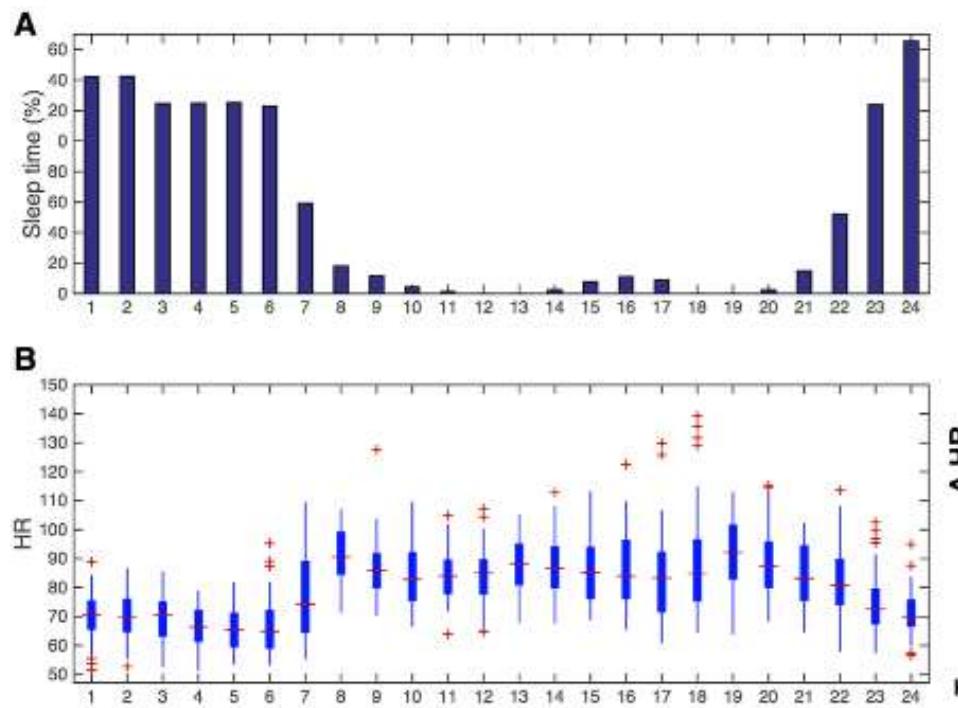


Li, Dunn et al.
PloS Biol 2017

- Worn by millions of people (20% of US)
- Make 100Ks of measurements each day
- Wearables can track many things: HR, HRV, Respiration Rate, SpO₂, Skin Temp, Blood Pressure

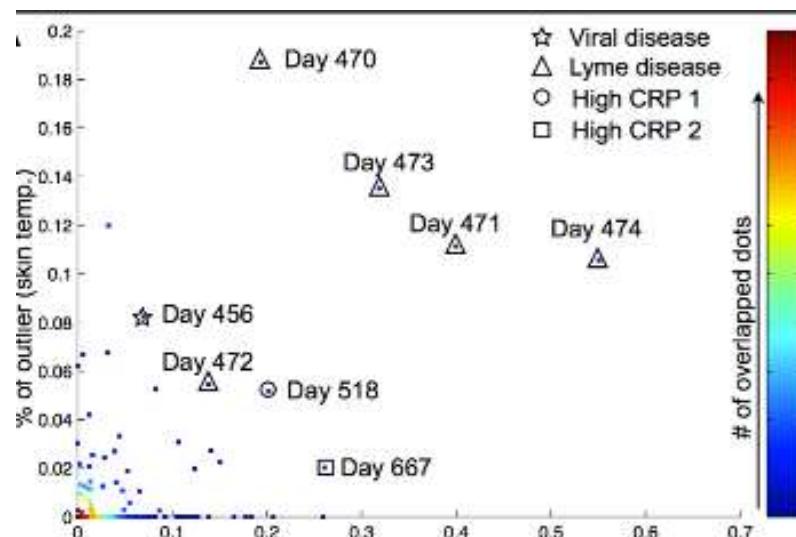
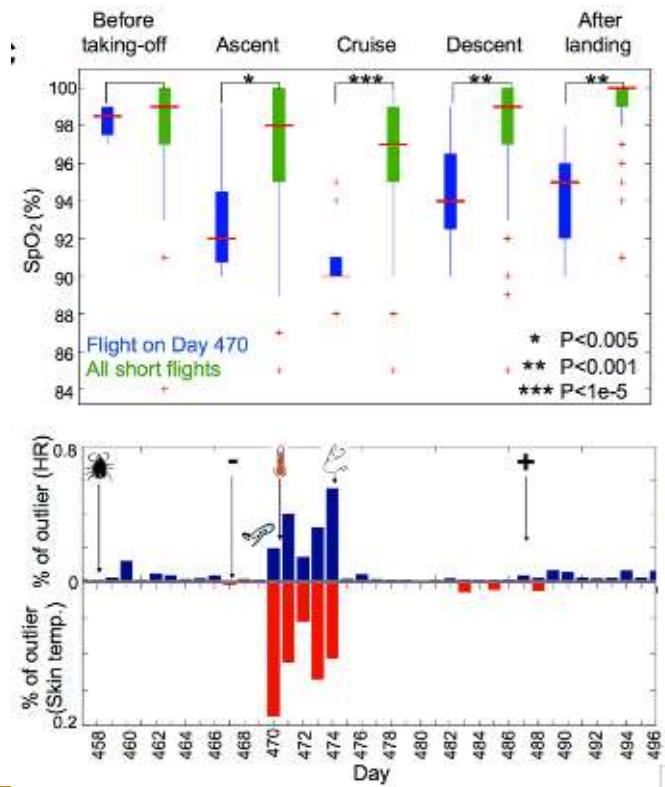
Slides: Dr Michael Snyder MD Stanford

Circadian and Diurnal Rhythms



Slides: Dr Michael Snyder MD Stanford

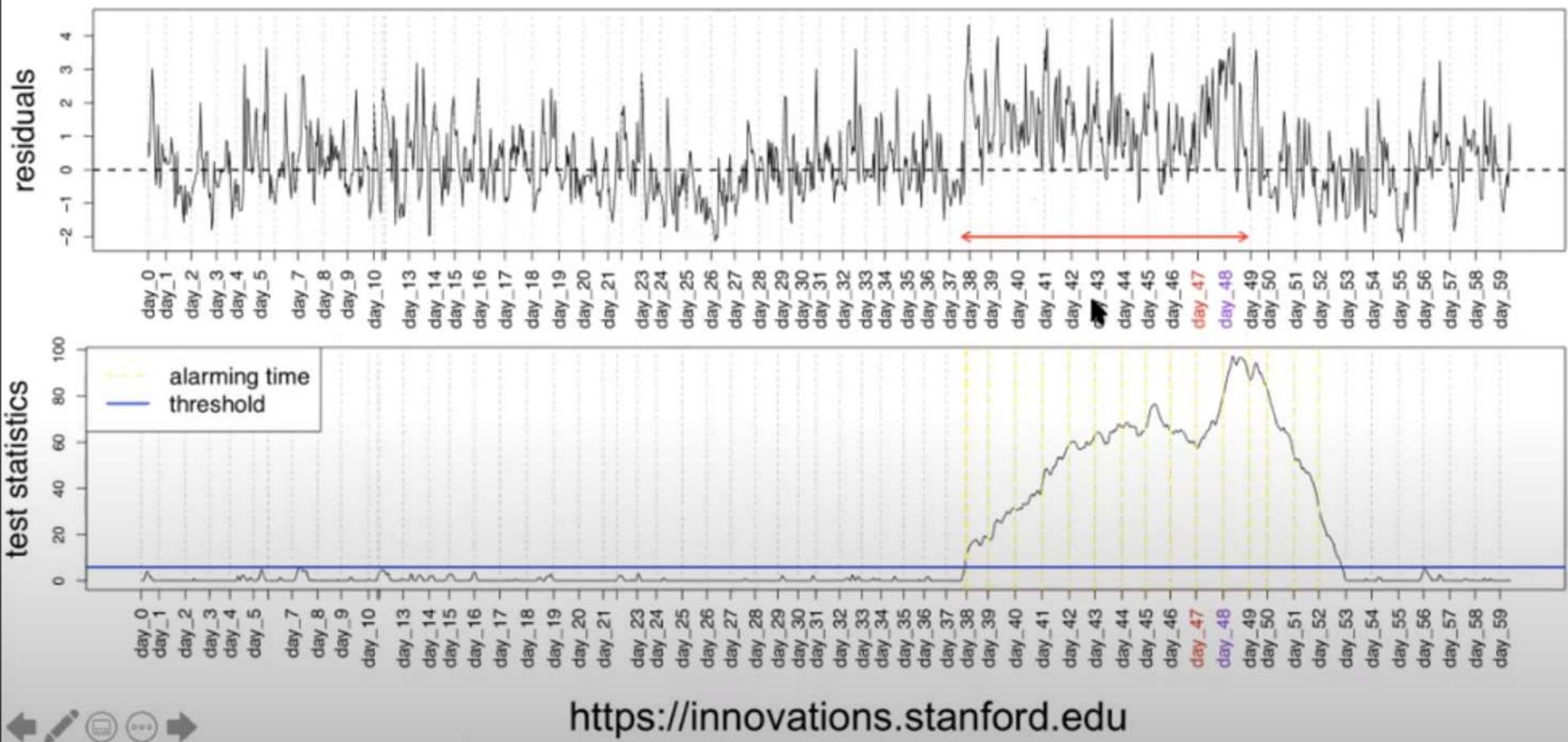
Detection of Lyme Disease



Slides: Dr Michael Snyder MD Stanford

<https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2001402>

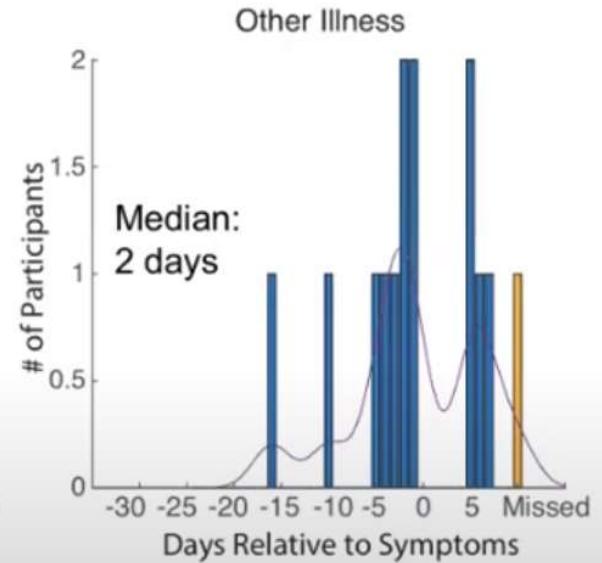
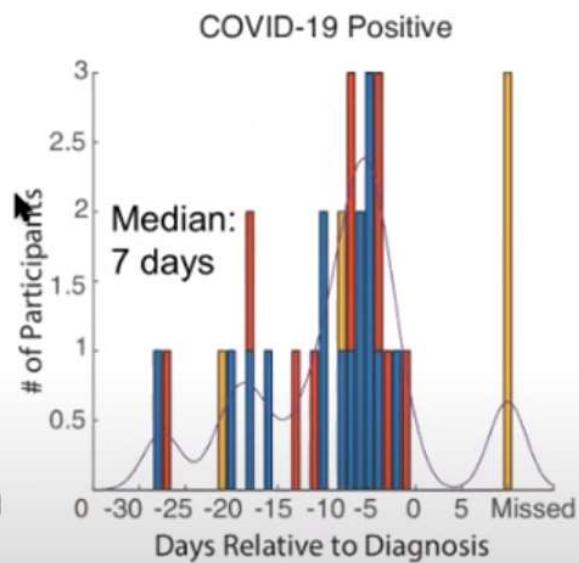
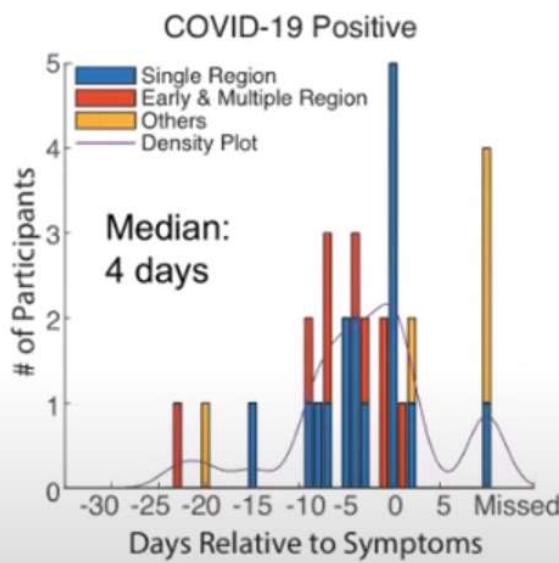
Identifying COVID-19 at early stage



<https://innovations.stanford.edu>

Slides: Dr Michael Snyder MD Stanford

Summary of Early Detection



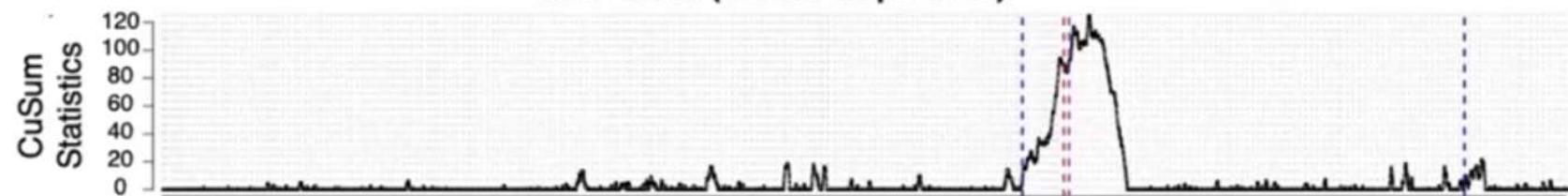
Elevated Heart Rate: 7 Beats/Min



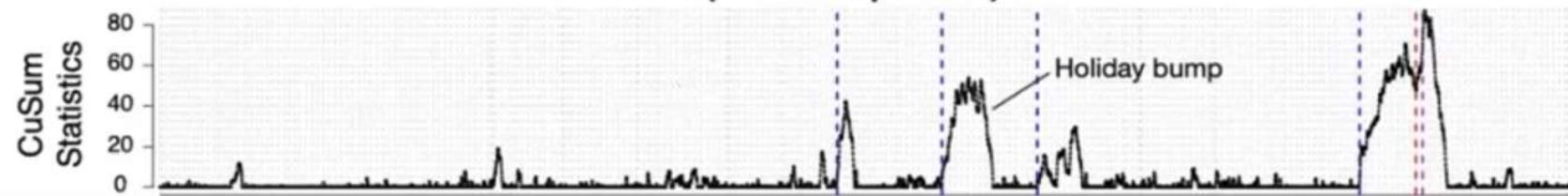
Slides: Dr Michael Snyder MD Stanford

Phase 2: Online Real-Time Detection - CuSum

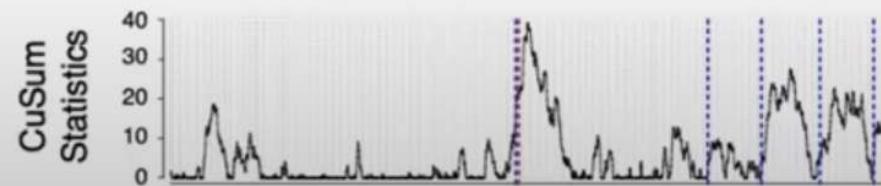
ASFODQR (COVID-19 positive)



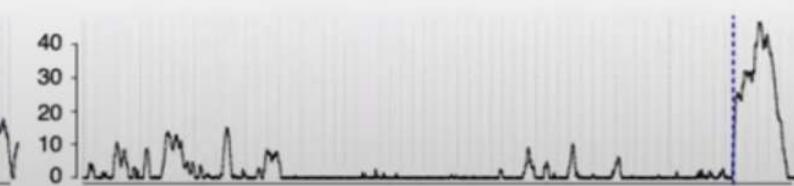
AJWW3IY (COVID-19 positive)



AR4FPCC (Other illness)



AFEFA29 (Healthy)



Slides: Dr Michael Snyder MD Stanford

Cohort and Data

- 1015 Fitbit participants
- 950 Apple Watch participants

Covid Positives

Apple Watch	33
Fitbit	34
Prospective	28
Retrospective	39

Presymptomatic/Asymptomatic detection rate via
NightSignal algorithm

(wrt [-14,+21] relative to symptom onset or test date)

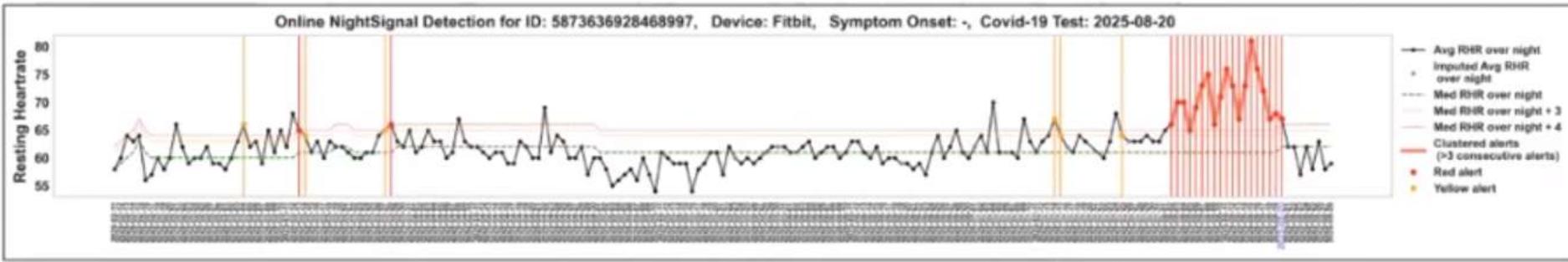
Total = 49/67 = 73.1%
Prospectives = 24/28

Slides: Dr Michael Snyder MD Stanford

Asymptomatic Detection Examples

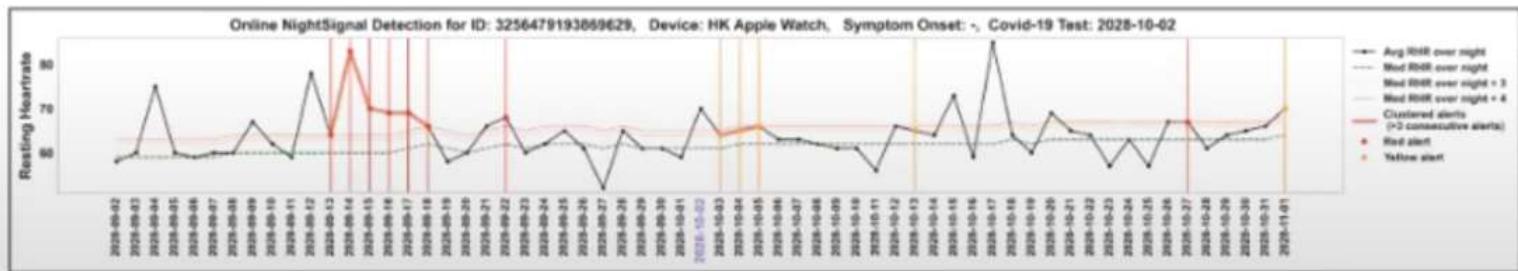
FitBit

NightSignal



Apple Watch

NightSignal

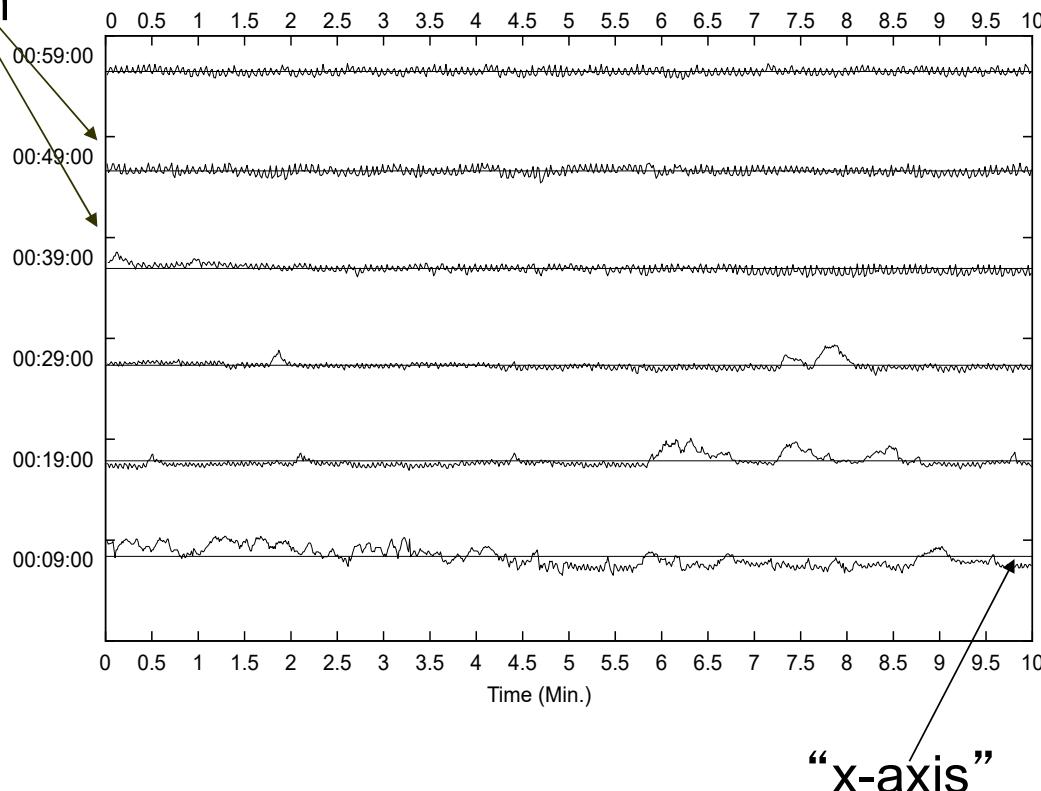


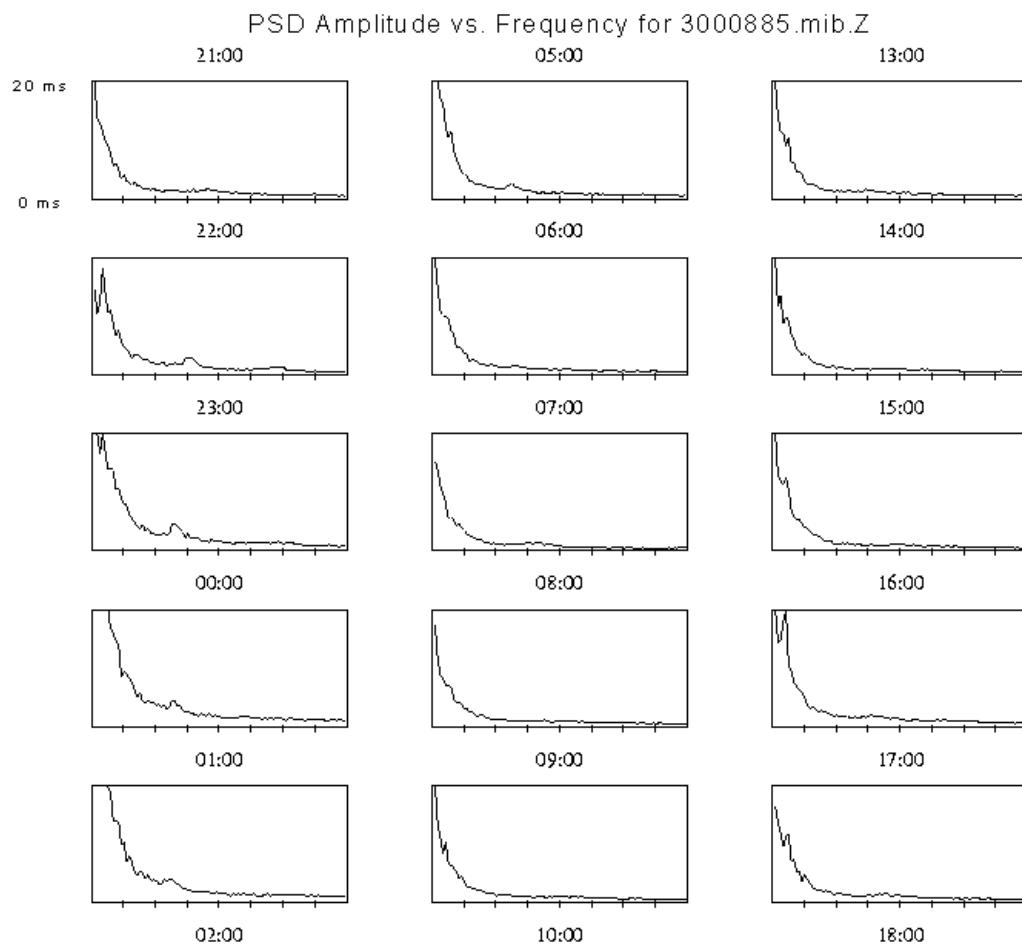
Slides: Dr Michael Snyder MD Stanford

Heart Rate Tachogram

0-100 bpm

- x-axis = time in minutes (0-10 minutes)
- y-axis for each 10-min plot is HR
- “x-axis” is mean HR for that 10-min segment





Hourly HRV Power Spectral Plots

By the way... What's wrong with these spectra?

Assessment of HRV

Approach 1

Physiologist's Paradigm

HR data collected over short period of time (~5-20 min), with or without interventions, under carefully controlled laboratory conditions.

<https://physionet.org/tutorials/hrv-toolkit/>

Assessment of HRV

Approach 2

Clinician's/Epidemiologists's Paradigm

Ambulatory Holter Recordings usually collected over 24-hours or less, usually on outpatients.

Approaches 1 and 2 can be combined

HRV Perspectives

Longer-term HRV

- quantifies changes in HR over periods of >5min.

Intermediate-term HRV

- quantifies changes in HR over periods of <5 min.

Short-term HRV

- quantifies changes in HR from one beat to the next

Ratio HRV

- quantifies relationship between two HRV indices.

Sources of Heart Rate Variability

Intrinsic Periodic Rhythms

- Respiratory sinus arrhythmia
- Baroreceptor reflex regulation
- Thermoregulation
- Neuroendocrine secretion
- Circadian rhythms
- Other, unknown rhythms

Extrinsic

- Activity - Sleep Apnea
- Mental Stress - Smoking
- Physical Stress

Ways to Quantify HRV

Approach 1: How much variability is there?

Time Domain and Geometric Analyses

Approach 2: What are the underlying rhythms? What physiologic process do they represent? How much power does each underlying rhythm have?

Frequency Domain Analysis

Approach 3: How much complexity or self-similarity is there?

Non-Linear Analyses

Commonly used time-domain measures

AVNN Average of all NN intervals (i.e. normal to normal)

SDNN Standard deviation of all NN intervals

SDANN Standard deviation of the averages of NN intervals in all 5-minute segments of a 24-hour recording

SDNNIDX Mean of the standard deviations of NN intervals in all 5-minute segments of a 24-hour recording

rMSSD Square root of the mean of the squares of differences between adjacent NN intervals

pNN50 Percentage of differences between adjacent NN intervals that are greater than 50 ms

Time Domain HRV

Longer-term HRV

- SDNN

Standard deviation of NN intervals (in msec) (Total HRV)

- SDANN
- Standard deviation of mean values of QRS-QRSs for each 5 minute interval in msec
- (Reflects circadian, neuroendocrine and other rhythms + sustained activity)

Time Domain HRV

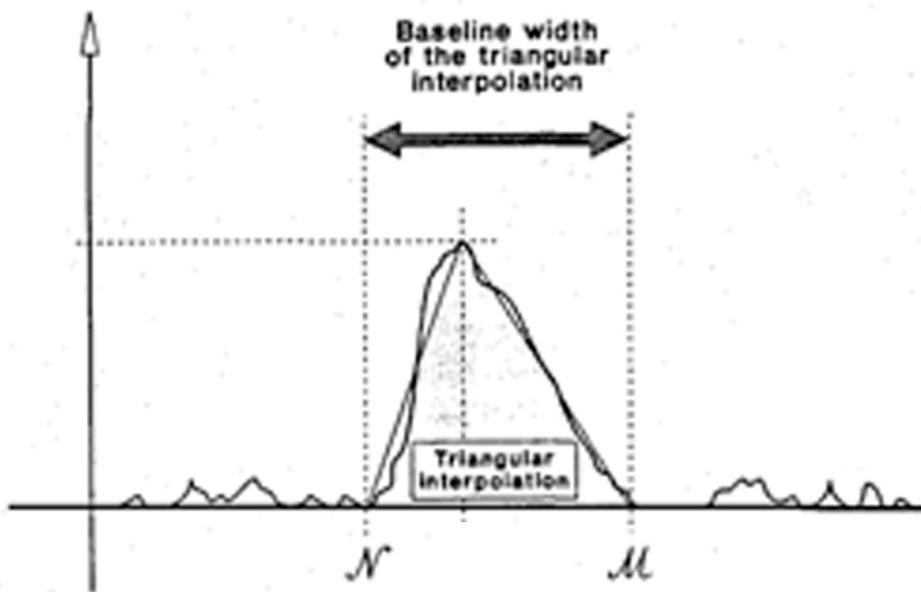
Intermediate-term HRV

- SDNNIDX
- Average of standard deviations of QRS to QRS for each 5 min interval in ms (Combined SNS and PNS HRV)

Coefficient of variance (CV)

- = SDNNIDX / AVNN.
- Heart rate normalized SDNNIDX.

Geometric HRV



HRV Index-Measure of longer-term HRV

From Farrell et al, J Am Coll Cardiol 1991;18:687-97

Examples of Normal and Abnormal Geometric HRV

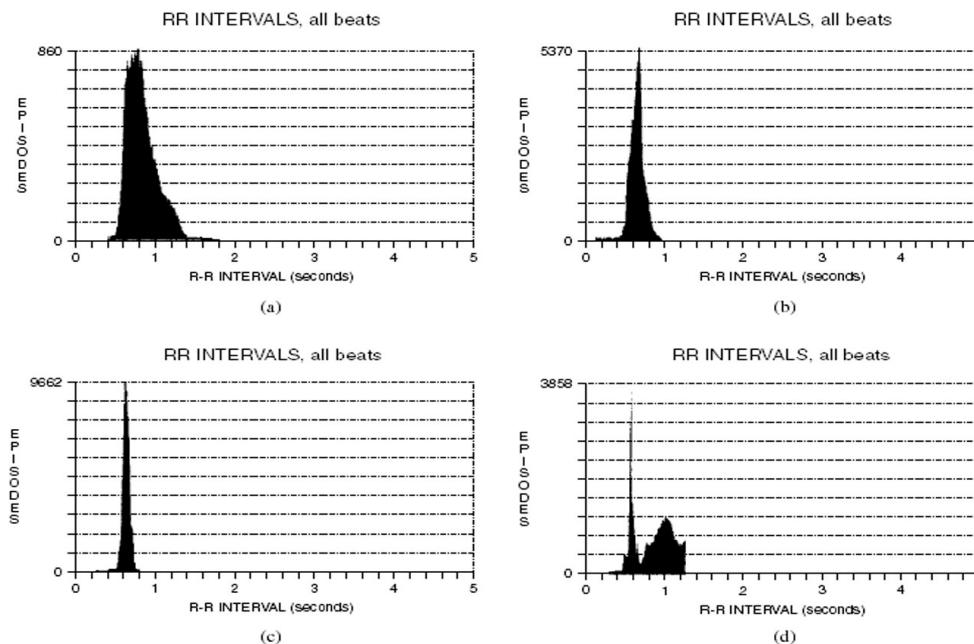


Fig. 1. Examples of R-R histograms from (a) a normal subject, (b) a cardiac patient with decreased HRV, (c) a cardiac patient with very low HRV and (d) a patient with an extremely abnormal R-R interval distribution and a large number of ventricular ectopic beats.

Frequency Domain HRV

Based on autoregressive techniques or fast Fourier transform (FFT).

Partitions the total variance in heart rate into underlying rhythms that occur at different frequencies.

These frequencies can be associated with different intrinsic, autonomically-modulated periodic rhythms.

Commonly used frequency-domain measures

TOTPWR

- Total spectral power of all NN intervals up to 0.04Hz

ULF

- Total spectral power of all NN intervals up to 0.003Hz

VLF

- Total spectral power of all NN intervals between 0.003 and 0.04Hz

LF

- Total spectral power of all NN intervals between 0.04 and 0.15Hz

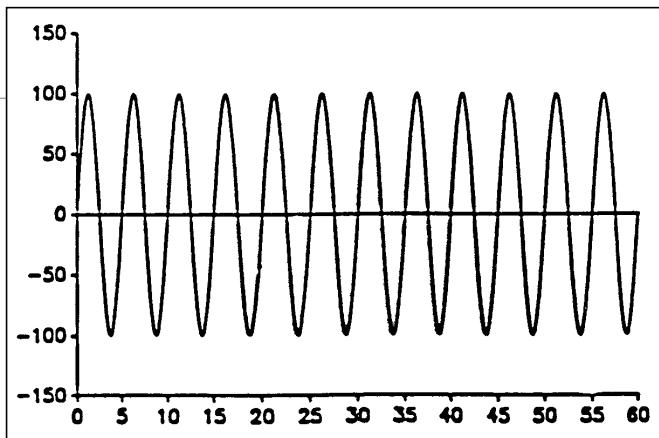
HF

- Total spectral power of all NN intervals between 0.15 and 0.4Hz

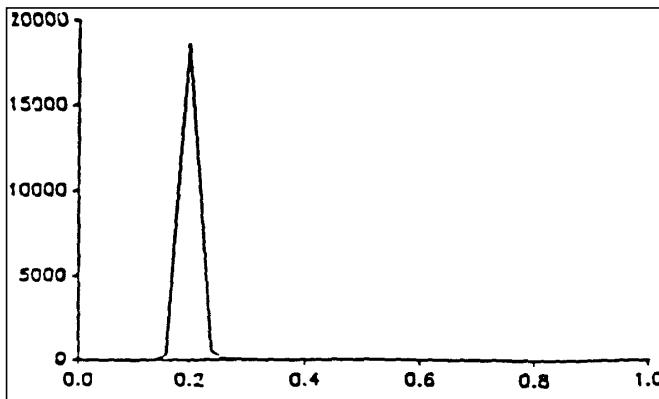
LF/HF

- Ratio of low to high frequency power

What are the Underlying Rhythms?



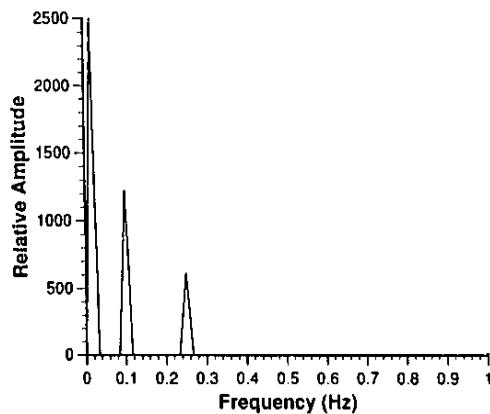
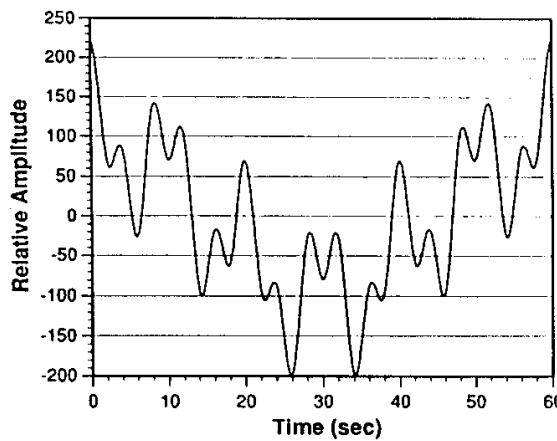
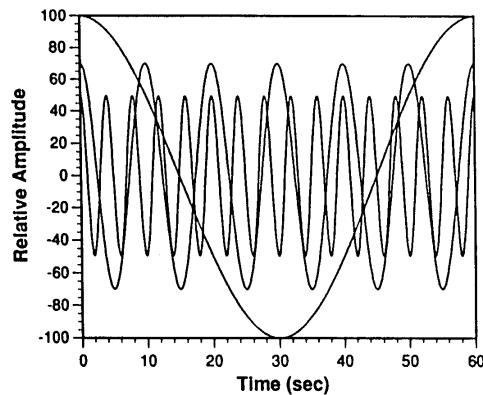
One rhythm
5 seconds/cycle or
12 times/min



5 seconds/cycle=
1/5 cycle/second

1/5 cycle/second=
0.2 Hz

What are the Underlying Rhythms?



Three Different Rhythms

High Frequency = 0.25 Hz (15 cycles/min)
Low Frequency = 0.1 Hz (6 cycles/min)
Very Low Frequency = 0.016 Hz
(1 cycle/min)

Ground Rules for Measuring Frequency Domain HRV

Only normal-to-normal (NN) intervals included

At least one normal beat before and one normal beat after each ectopic beat is excluded

Cannot reliably compute HRV with >20% ectopic beats

With the exception of ULF, HRV in a 24-hour recording is calculated on shorter segments (5 min) and averaged.

Frequency Domain HRV

Longer-Term HRV

Total Power (TP)

Sum of all frequency domain components.

Ultra low frequency power (ULF)

At >every 5 min to once in 24 hours. Reflects circadian, neuroendocrine, sustained activity of subject, and other unknown rhythms.

Frequency Domain HRV

Very low frequency power (VLF)

At ~20 sec-5 min frequency

Reflects activity of renin-angiotensin system, vagal activity, activity of subject.

Exaggerated by sleep apnea. Abolished by atropine

Low frequency power (LF)

At 3-9 cycles/min Baroreceptor influences on HR, mediated by SNS and vagal influences.
Abolished by atropine.

Frequency Domain HRV

Short-term HRV

High frequency power (HF)

At respiratory frequencies (9-24 cycles/minute, respiratory sinus arrhythmia but may also include non-respiratory sinus arrhythmia).

Normally abolished by atropine.

Vagal influences on HR with normal patterns.

Frequency Domain HRV

Ratio HRV

LF/HF ratio may reflect SNS:PNS balance under some conditions.

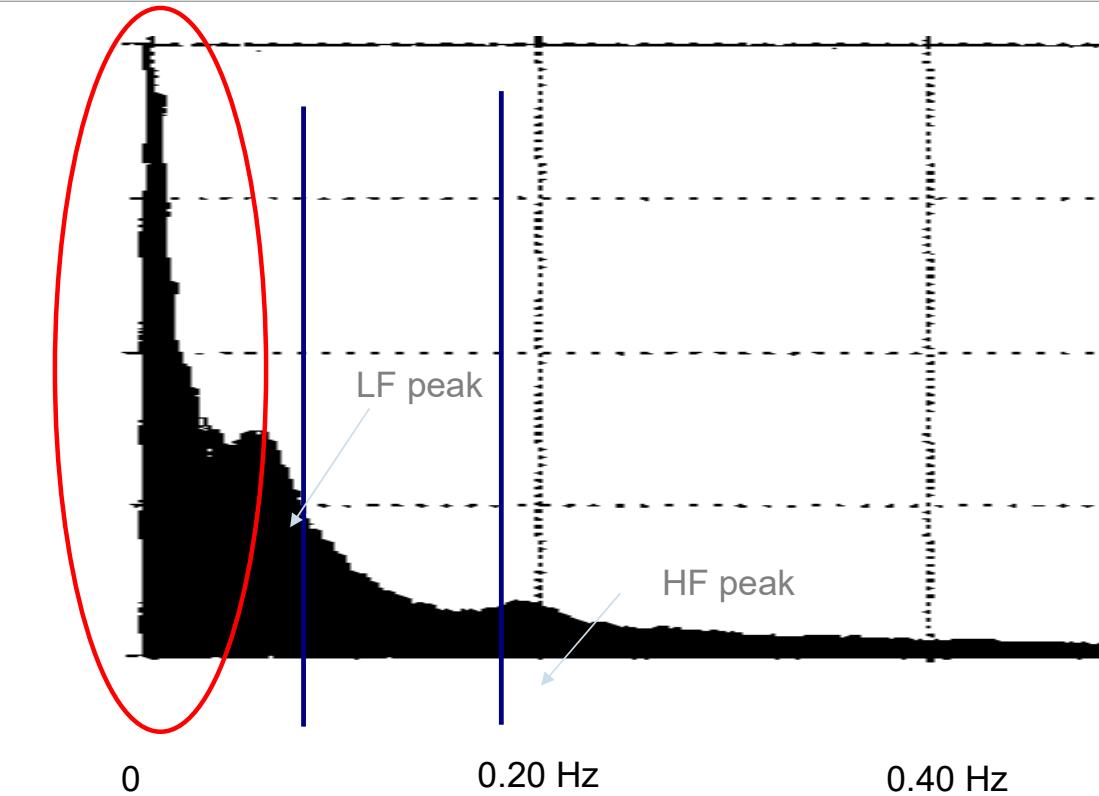
Normalized LF power= $\text{LF}/(\text{TP-VLF})$ correlates with SNS activity under some conditions.

Normalized HF power= $\text{HF}/(\text{TP-VLF})$ proposed as a measure of relative vagal control of HR. Increased for abnormal HRV.

Frequency Domain HRV

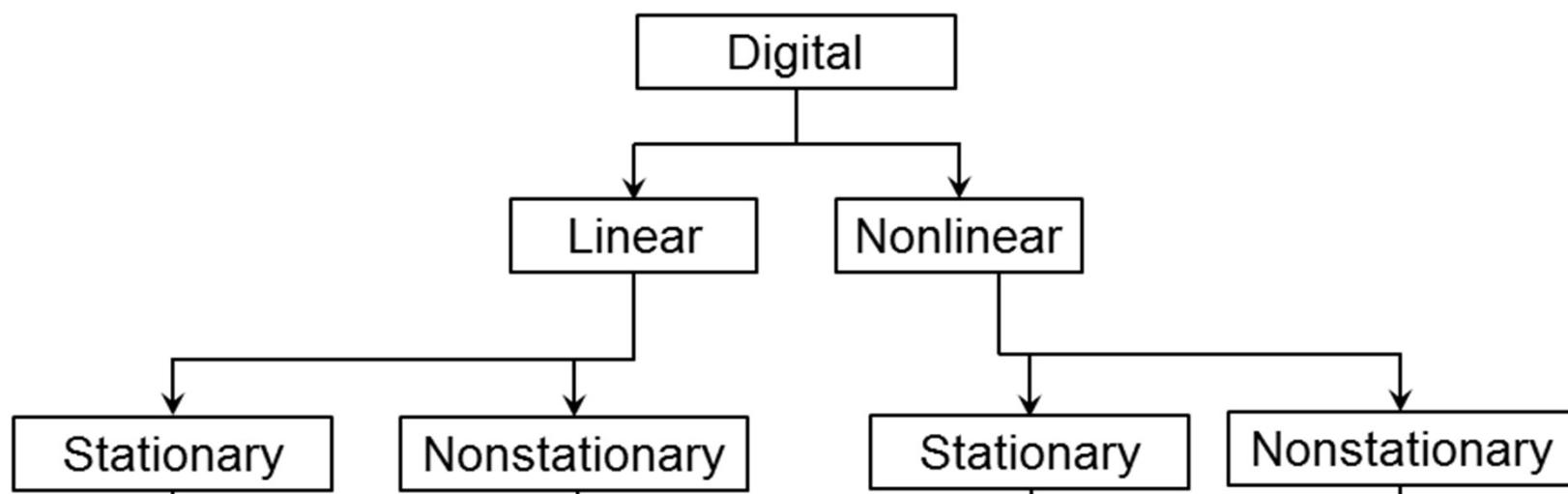
24-hour average of 2-min power spectral plots in a healthy adult

FYI- what's wrong about this spectrum?



How else can we assess variability?

Digital Signal Types



Signal Encoding

All signals involve some type of encoding scheme.

Most encoding strategies can be divided into two broad categories or domains:

- continuous
- discrete.

Continuous signals usually encode information in terms of signal amplitude (the intensity of the signal, voltage, or current values) as a function of time.

Example

The temperature in a room can be encoded so that 0 volts represents 0.0 °C, 5 volts represents 10 °C, 10 volts represents 20 °C, and so on. If *linear*, the encoding equation would be:

Temp	Voltage
0	0
5	10
10	20
15	30
20	40

Linear Signals

input (temperature), output (voltage) following the classic linear relationship:

$$y = mx + b$$

where m is the slope of the input-output relationship and b is the offset which in this case is 0.0.

The temperature can be found from the voltage output of the transducer as:

$$\text{temperature} = 2 * \text{voltage } ^\circ\text{C}$$

When the information is encoded in terms of signal amplitude, it is known as an analog signal.

Linearity

If you double the input into a linear system, you will double the output.
basic concept is proportionality.

- if the independent variables of linear function are multiplied by a constant, k , the output of the function is simply multiplied by k .

If $y = f(x)$ where f is a linear function:

Properties of Linear Signals

If f is a linear function:

$$f(x_1(t)) + f(x_2(t)) = f(x_1(t) + x_2(t))$$

- If $z = \frac{df(x)}{dx}$ then $\frac{df(kx)}{dx} = k \left(\frac{df(x)}{dx} \right) = kz$
- If $z = \int f dx$ then $\int f(kx) dx = k \int f(x) dx = kz$
- Derivation and integration are linear operations.
- Systems that contain derivative and integral operators and other linear operators produce linear signals.

Time Invariance

If a system's response characteristics do not change over time, it is said to be *time-invariant*.

Time invariance is a stricter version of stationarity since a time-invariant system would also be stationary.

Mathematically: if f is a linear function, then for time invariance:

$$y(t - T) = f(x(t - T))$$

LTI Systems

A system that is both linear and time-invariant is referred to as a *linear time-invariant (LTI)* system.

The LTI assumptions allow us to apply a powerful array of mathematical tools known collectively as linear systems analysis or linear signal analysis.

Most living systems change over time, they are adaptive, and they are often nonlinear, but the power of linear systems analysis is simplifies everything so much that simplifying assumptions or approximations are made so that these tools can be used.

Causality

A system that responds only to current and past inputs is termed *causal*.

Systems that exist in the real-world (e.g., analog electronic filters) must be causal.

Computer programs can operate using values that appear to be in the future with respect to a given operation.

Such systems are *noncausal*.

Superposition

Linearity is required for the application of an important concept known as *superposition*.

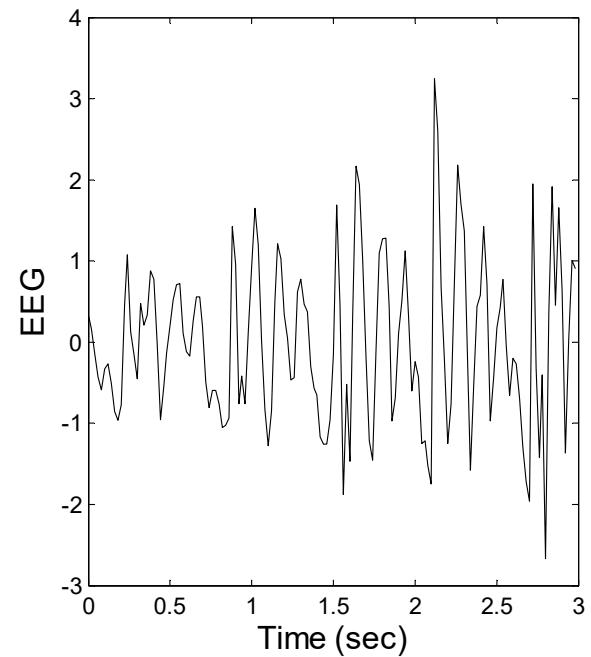
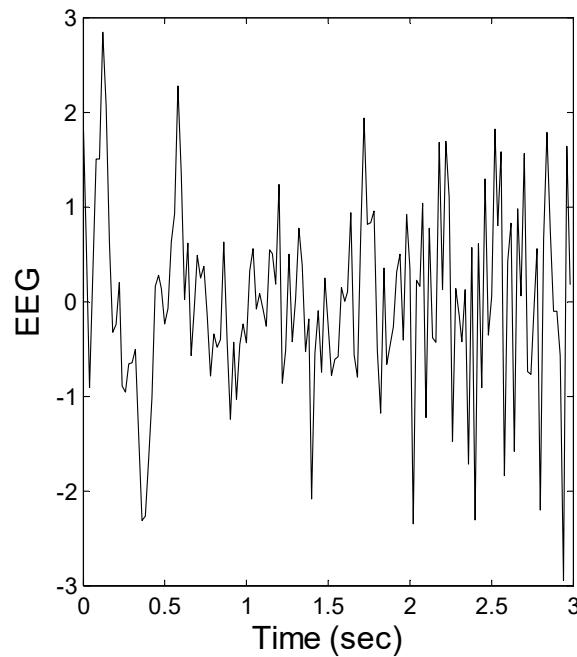
Superposition states that if there are two (or more) inputs acting on a linear system, the system responds to each as if it were the only input

This allows a “divide and conquer” approach

Data Functions and Transforms

Basic measurements do not definitively describe signals.

For example, these two EEG segments have the same mean, RMS, and variance, but are clearly different.



Describing Signals

We would like some method to capture the differences between these two (and other) signals, and preferably to be able to quantify these differences.

Other functions (or waveforms) can be used to describe signals and their differences.

In signal processing, functions fall into two categories:

- 1) Data, including waveforms and images;
- 2) Functions that operate on data.

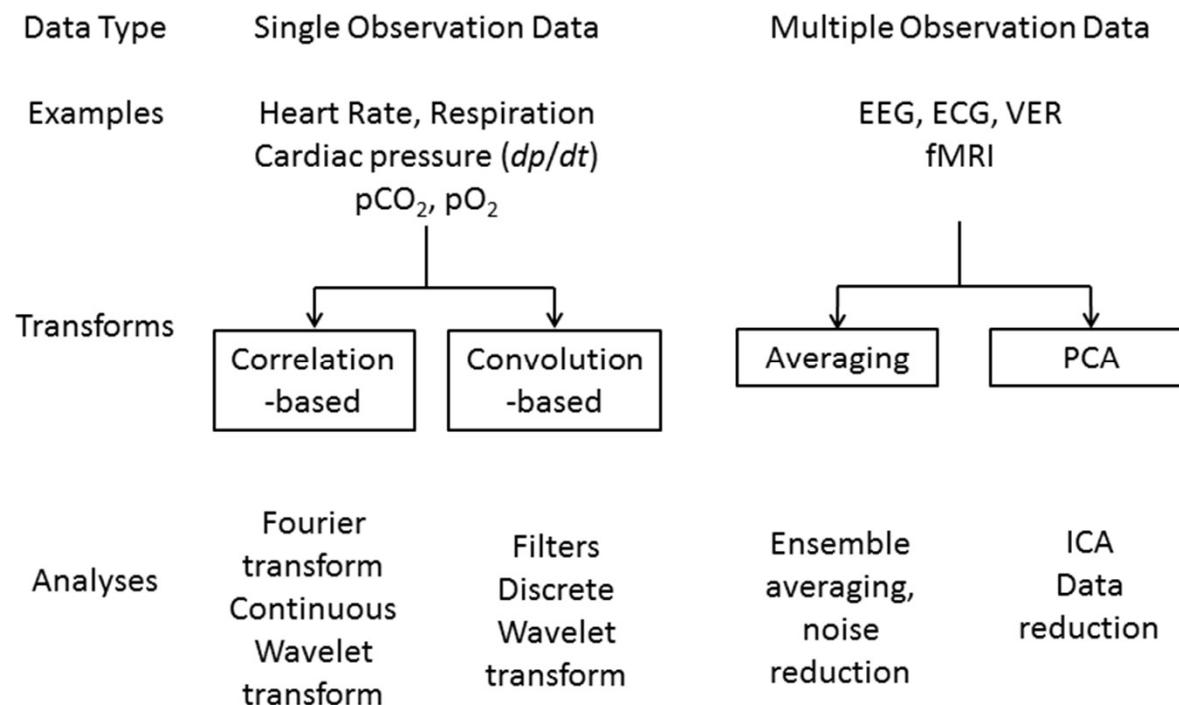
Transformations: Functions that Operate on Signals

Transformations are operators that modify data.

Transformations are used to:

- Improve data quality
- make the data easier to interpret
- Reduce the size of the data by removing unnecessary components

Transformations: depend on data type



Comparing Waveforms: Correlation

Correlation seeks to quantify how much one function (i.e., signal) is like another.

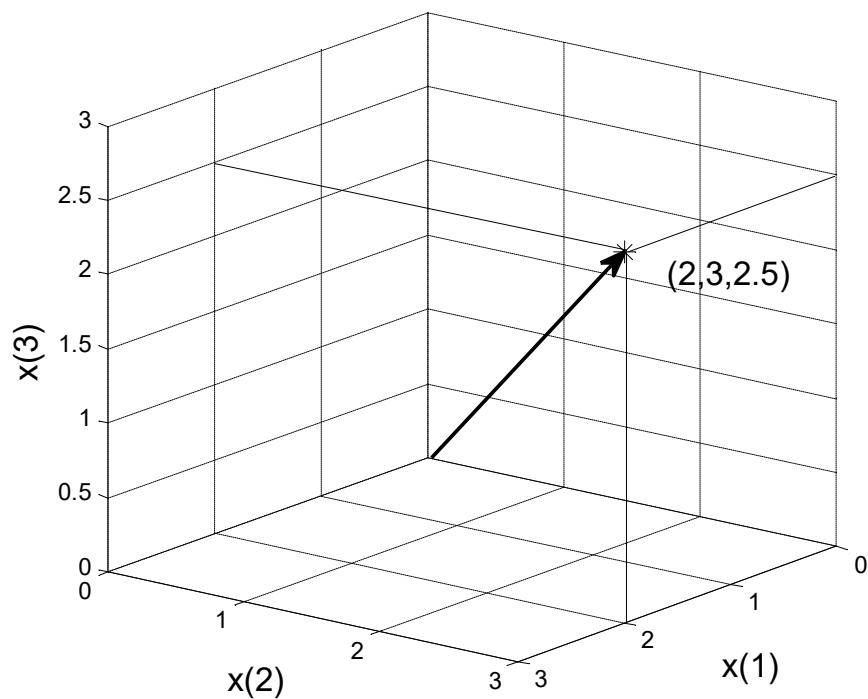
All correlation-based approaches involve taking the sum of the sample-by-sample product of the two functions:

$$\begin{aligned} r_{xy} &= x[1]y[1] + x[2]y[2] + x[3]y[3] + \dots + x[N]y[N] \\ &= \sum_{n=1}^N x_n y_n \end{aligned}$$

where r_{xy} is used to indicate correlation and the subscripts x and y indicate what is being correlated.

- Different normalizations ($\frac{1}{N} \sum_{n=1}^N x_n y_n$) can be used.

Vector Representation



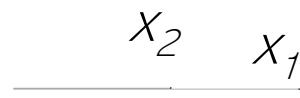
A string of numbers can be thought of as a vector in N -dimensional space:
 $x[n] = [x_1, x_2, x_3, \dots, x_n]$

Data sequence:
 $x[n] = 2, 3, 2.5$
represented as a vector in 3-dimensional space.

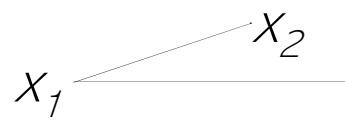
This curious way of thinking about a data string does have its uses.

Signal Comparison using Vector Representation

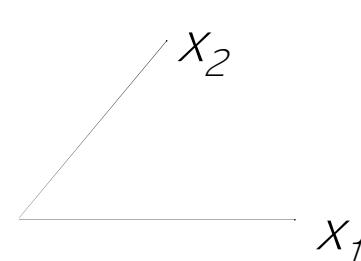
If two strings are mathematically similar, their vector representations will project on one another:



Completely similar
 $\vartheta = 0$ deg.



Highly similar
 $\vartheta = \text{small}$



Moderately similar
 $\vartheta = \text{larger}$



Completely different
 $\vartheta = 90$ deg

Correlation and the Scalar Product

The projection of one vector on another is found by taking the scalar product of the two vectors.

This shows the relationship between vector projection and correlation.

The scalar product is defined as:

$$\begin{aligned} \text{Scalar product of } x \& y &\equiv \langle x, y \rangle &= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \\ &= x_1 y_1 + x_2 y_2 + \dots + x_N y_N \\ &= \sum_{n=1}^N x_n y_n \end{aligned}$$

Scalar Product – Correlation (cont)

Note that the scalar product results in a single number (i.e., a scalar), not a vector.

The scalar product can also be defined in terms of the magnitude of the two vectors and the angle between them:

where θ is the angle between the two vectors.

Projection (correlation) is an excellent way to compare two signals or to compare a signal with a 'probing' or 'test' waveform.

In MATLAB, the scalar product is just: `sum(x.*y)`

Orthogonality

In common usage, “orthogonal” means perpendicular: if two lines are orthogonal they are perpendicular.

In vector representation, orthogonal signals would have orthogonal vectors.

The formal definition for orthogonal signals is that their correlation (or scalar product) is zero:

Orthogonality (cont)

An important characteristic of signals that are orthogonal (i.e., uncorrelated) is that when they are combined or added together they do not interact with one another.

Orthogonality simplifies many calculations and some analyses could not be done, at least not practically, without orthogonal signals.

Orthogonality is not limited to two signals. Whole families of signals can be orthogonal (or orthonormal*) and are called orthogonal or orthonormal sets.

Coherence

- The coherence C between two signals x and y is defined as the cross-spectrum S_{xy} normalized by the power spectra S_{xx} and S_{yy}
- However, to make the coherence a dimensionless number between 0 and 1, S_{xy} is squared:

NOTE: S_{xy} will usually be a complex function, whereas S_{xx} and S_{yy} are both real functions

Coherence (cont)

If we calculate the normalized cross-spectrum as a complex number for a single frequency and a single trial, the outcome always has magnitude 1 and phase angle φ .

E.g. define:

$$X(\omega) = a + bj \quad Y(\omega) = c + dj$$

Can write expression for each of the components of the coherence equation as:

$$S_{xx}(\omega) = X(\omega)X^*(\omega) = (a + bj)(a - bj) = a^2 + b^2,$$

$$S_{yy}(\omega) = Y(\omega)Y^*(\omega) = (c + dj)(c - dj) = c^2 + d^2, \quad \text{and}$$

$$|S_{xy}(\omega)|^2 = |X(\omega)Y^*(\omega)|^2 = |(a + bj)(c - dj)|^2 = |(ac + bd) - j(ad - bc)|^2$$

Coherence (cont)

Because S_{xy} is a complex number, the magnitude squared is the sum of the squares of the real and imaginary parts:

$$|S_{xy}(\omega)|^2 = (ac + bd)^2 + (ad - bc)^2 = a^2c^2 + b^2d^2 + a^2d^2 + b^2c^2$$

Looking at the original coherence equation it is easy to see why it equals 1:

$$C(\omega) = \frac{a^2c^2 + b^2d^2 + a^2d^2 + b^2c^2}{(a^2 + b^2)(c^2 + d^2)} = 1$$

Coherence (cont)

- coherence is typically estimated by averaging over several epochs or frequency bands
- Thus S_{xy} is determined by averaging over n epochs:

$$C(\omega) = \frac{|\langle S_{xy}(\omega) \rangle_n|^2}{\langle S_{xx}(\omega) \rangle_n \langle S_{yy}(\omega) \rangle_n}$$

VERY Important Note: The averaging of cross-spectrum S_{xy} occurs before the absolute value is taken. A common beginner's mistake is to average the absolute which will always come out to one!

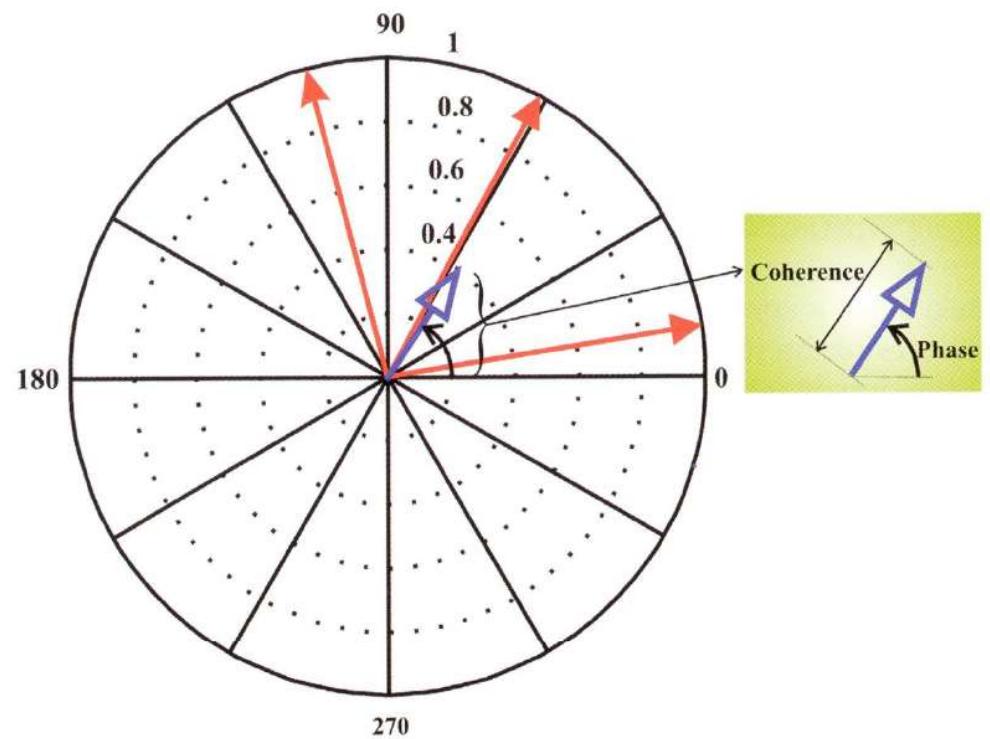
Coherence(cont)

- When determining $C(\omega)$ for a single frequency, ω over different samples (from an ensemble) one obtains several vectors on the unit circle, typically with different phase angles.
- The magnitude of sum of individual vectors indicates the degree of coherence, and the resulting phase angle is the phase coherence.

Coherence (cont)

Complex numbers (red vectors) in complex plane are values for normalized cross-spectrum obtained from different samples out of an ensemble.

The blue arrow is average of the 3 (i.e. amplitude coherence, or just coherence), and the phase is the phase coherence.



Coherence (cont)

The magnitude of a single coherence estimate is always 1. The use of the coherence metric therefore only makes sense if the value is determined repeatedly and subsequently averaged.

Usually the coherence values are:

- (1) averaged over different frequencies in a frequency band
- (2) averaged for a given frequency band for different epochs, or
- (3) averaged over both frequencies and epochs of the signal

Example

An example of coherence calculations associated with subdural electrode arrays implanted over the frontal cortex (red and blue, 1cm spacing) and temporal cortex (green-yellow, 5-mm spacing) of a patient with medically intractable epilepsy.

- colored pipes indicate pairs of electrodes with high coherence between them.

White pipes are not associated with a phase shift.

Green pipes indicate a phase delay at the blue end of the pipe.



Lecture 8

TAYLOR DEVET MASC.

PHD. CANDIDATE BIOLOGICAL AND BIOMEDICAL ENGINEERING
MCGILL UNIVERSITY

SHRINERS HOSPITAL FOR CHILDREN

Todays Aims...



Fourier Transforms



STFT



Wavelets

Fourier Transform

Process that breaks down functions depending on space or time into functions depending on spatial or temporal frequency

Time domain → Frequency Domain

Break signal down into sum of sines and cosines

Can be continuous or discrete



https://en.wikipedia.org/wiki/Fourier_transform

Jean-Baptiste Joseph Fourier

Jean-Baptiste Joseph Fourier

(21 March 1768 – 16 May 1830)

French mathematician and physicist

Fourier's law of conduction

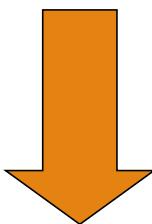
Discovery of the greenhouse effect



Timeline

1807, J.B. Fourier:

- All periodic functions can be expressed as a weighted sum of trigonometric function
- Denied publication by Lagrange, Legendre and Laplace
- 1822: Fourier's work is finally published
- ...
- ...
- ...
- ...
- ...
- ...
- ...
- 1965, Cooley & Tukey: Fast Fourier Transform



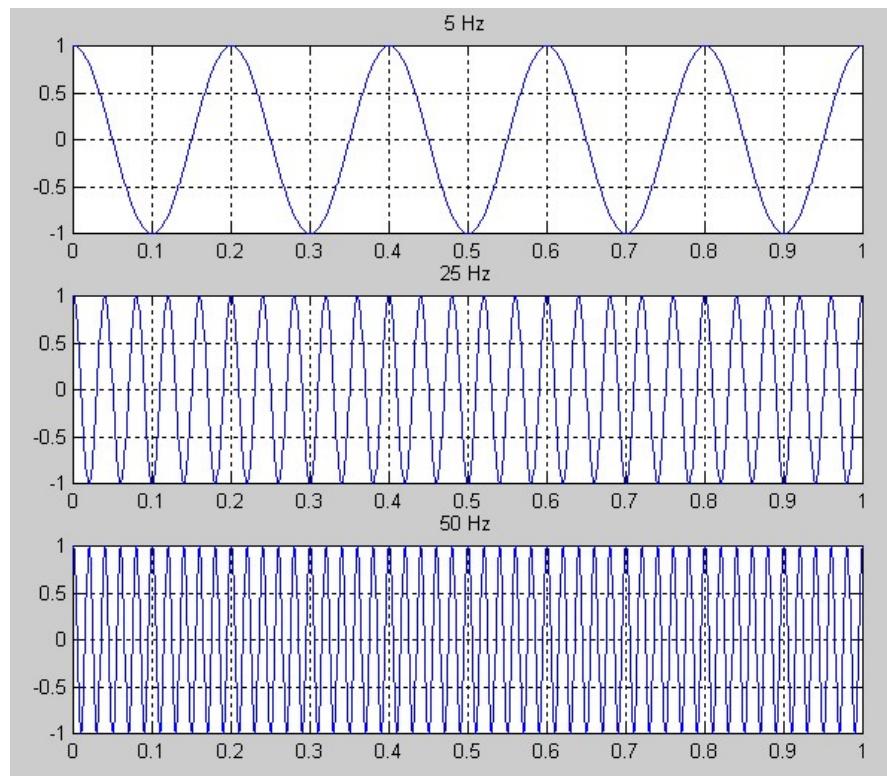
143 years

Fourier Transform (FT)

$$X_1(t) = \cos(2\pi \cdot 5 \cdot t)$$

$$X_2(t) = \cos(2\pi \cdot 25 \cdot t)$$

$$X_3(t) = \cos(2\pi \cdot 50 \cdot t)$$

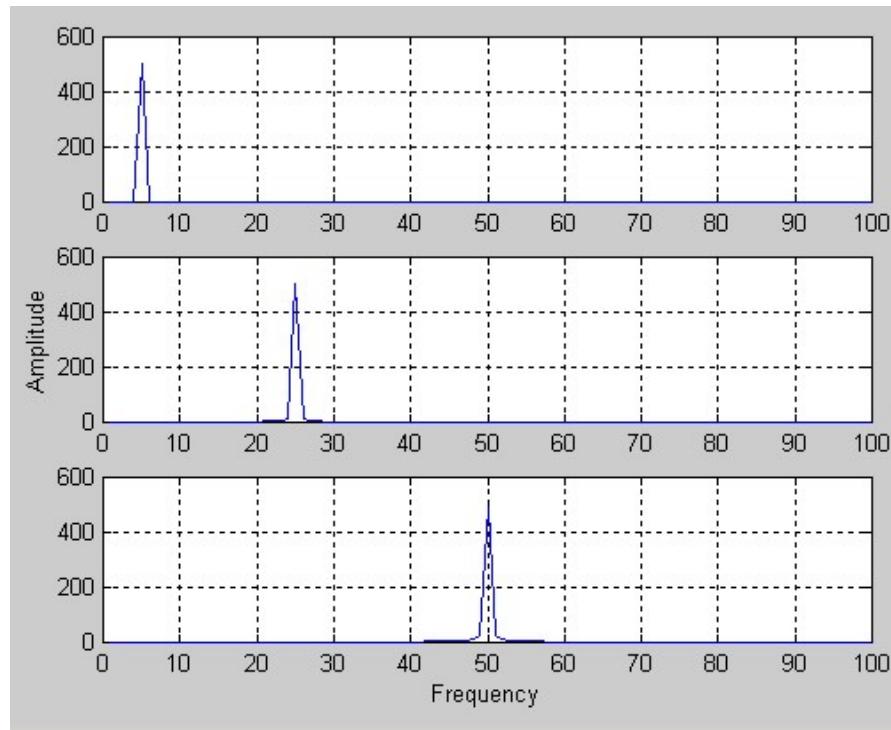


The Fourier Transform at work:

$$x_1(t) \xleftrightarrow{\mathcal{FT}} X_1(\omega)$$

$$x_2(t) \xleftrightarrow{\mathcal{FT}} X_2(\omega)$$

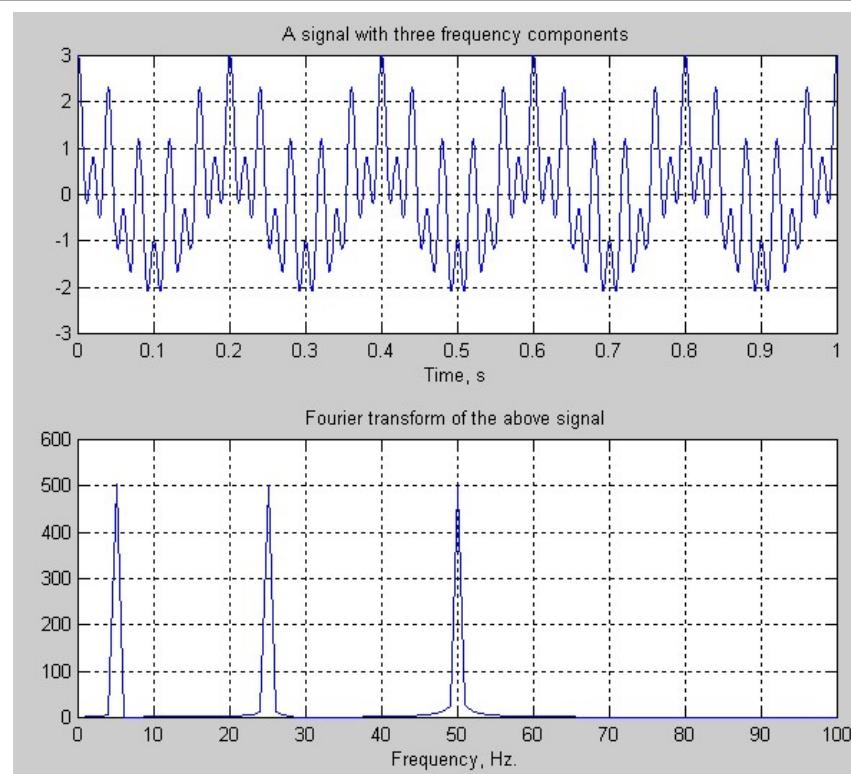
$$x_3(t) \xleftrightarrow{\mathcal{FT}} X_3(\omega)$$



FT At Work

$$x_4(t) = \cos(2\pi \cdot 5 \cdot t) + \cos(2\pi \cdot 25 \cdot t) + \cos(2\pi \cdot 50 \cdot t)$$

$$x_4(t) \xleftrightarrow{\mathcal{FT}} X_4(\omega)$$



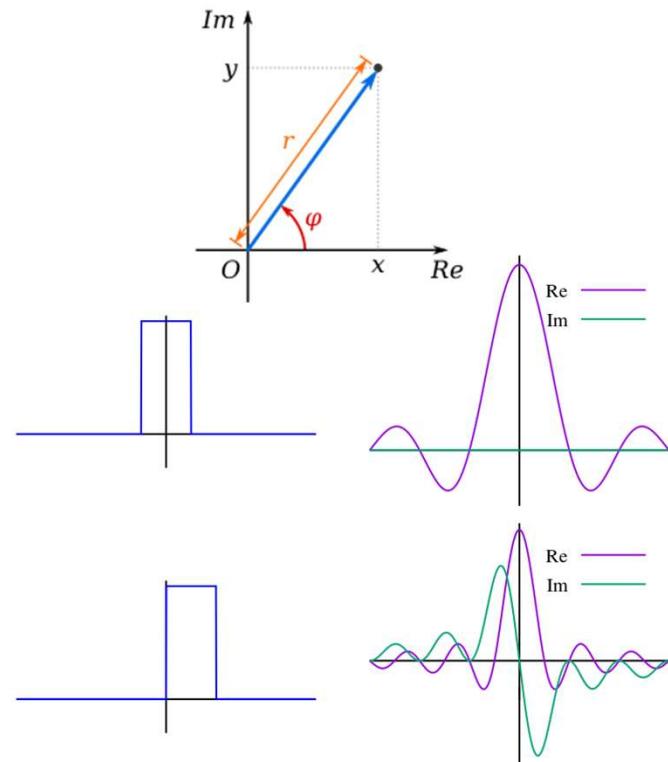
Fourier Transform

The amplitude is $|f^\wedge(\omega)|$

Phase is in $\arg f^\wedge(\omega)$ (Angle between positive real axis and point)

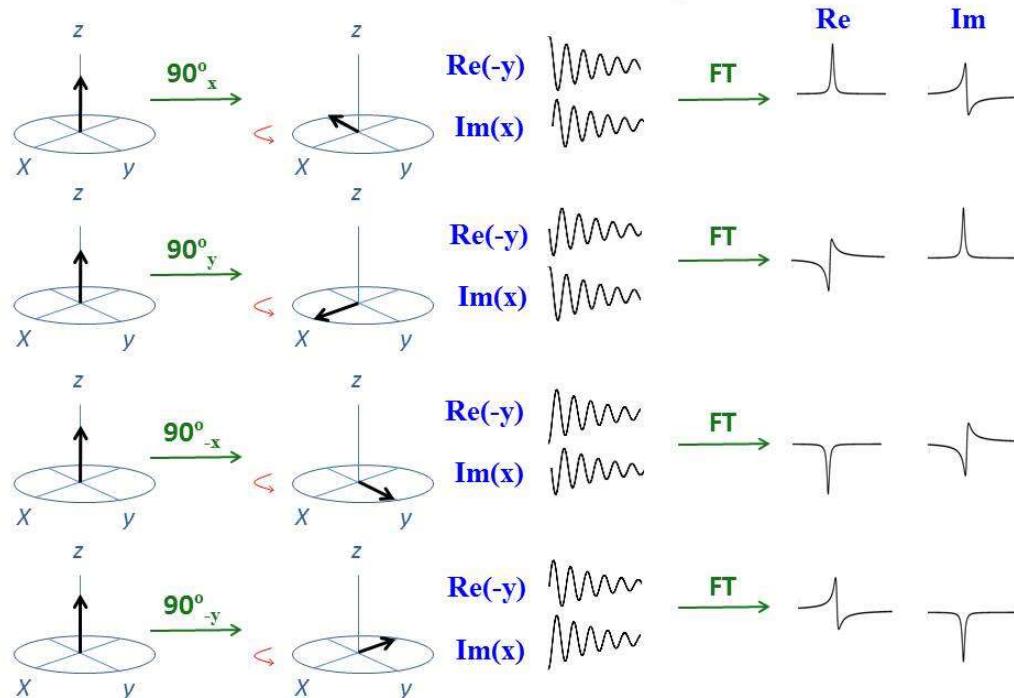
$$f(\omega) = |f^\wedge(\omega)| e^{i \arg f^\wedge(\omega)}$$

Shift in time is multiplication by $e^{-ia\omega}$ in the frequency domain

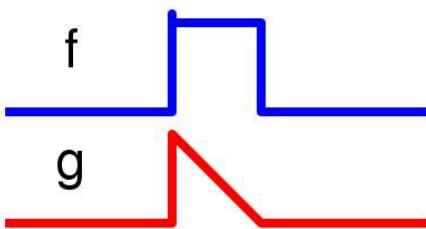


A great example showing effects of phase on the FT

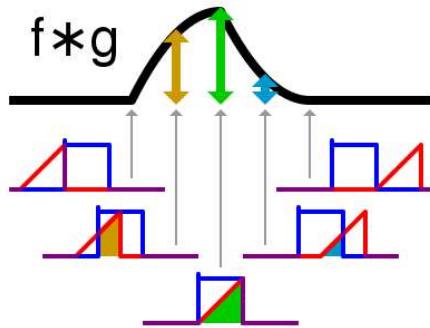
The Phase of an NMR Spectrum



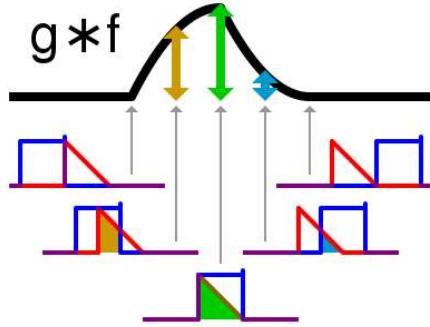
Convolution



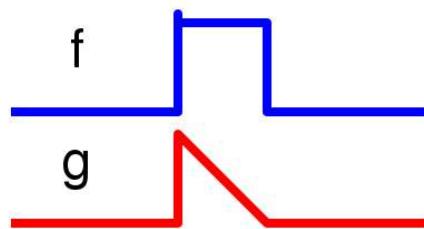
$f * g$



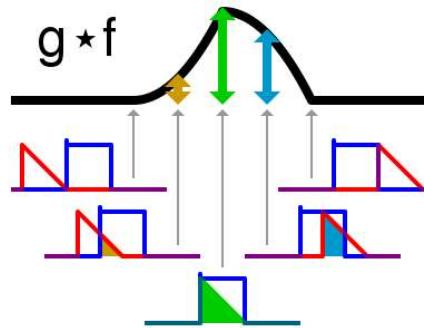
$g * f$



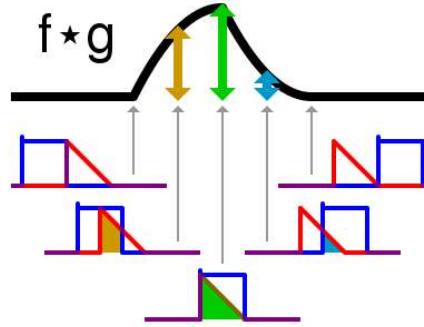
Cross-correlation



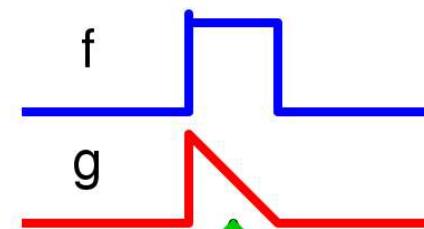
$g \star f$



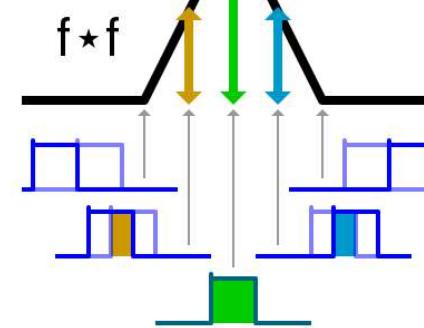
$f \star g$



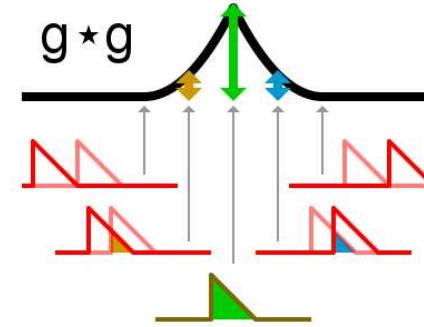
Autocorrelation



$f \star f$



$g \star g$



Data Truncation (Windowing)

A digitized waveform must necessarily be truncated to the length of the memory storage array, a process described as “[windowing](#).”

The windowing process can be thought of as multiplying the data by some shaped function.

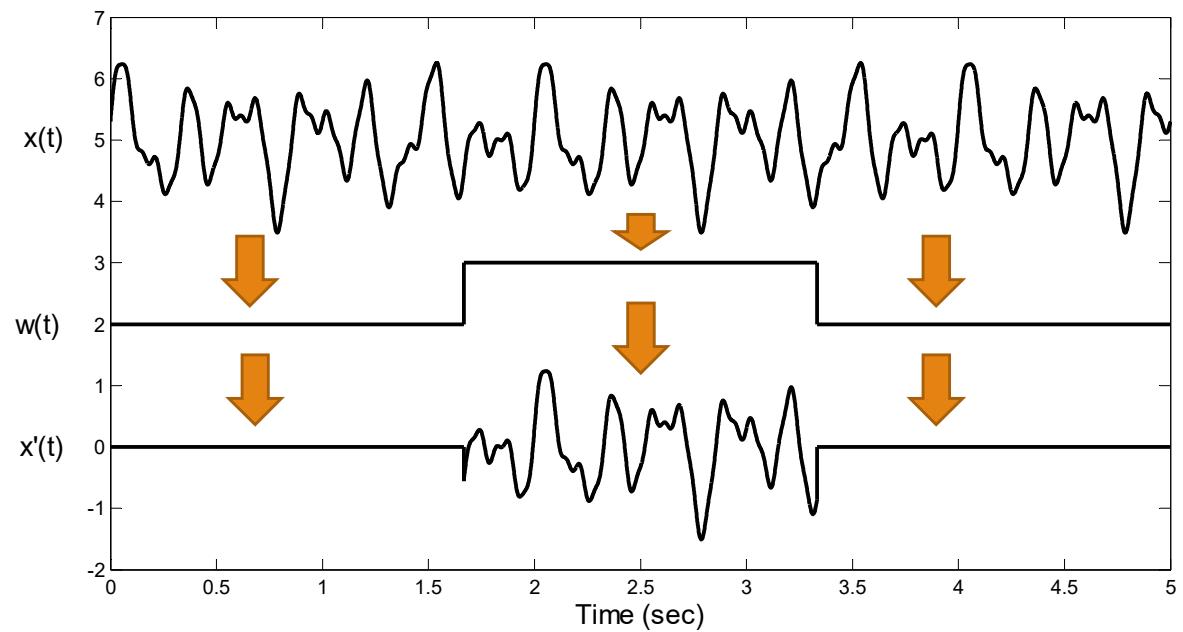
If the waveform is simply truncated (as often the case), then the window shape is rectangular.

[Multiplying in the time domain](#) is equivalent to [convolution in the frequency domain](#) (and vice versa)

Windowing

Truncation is the same as multiplying the data by a function that is 1. for the length of the data and 0.0 everywhere else.

This function is called 'rectangular window.'

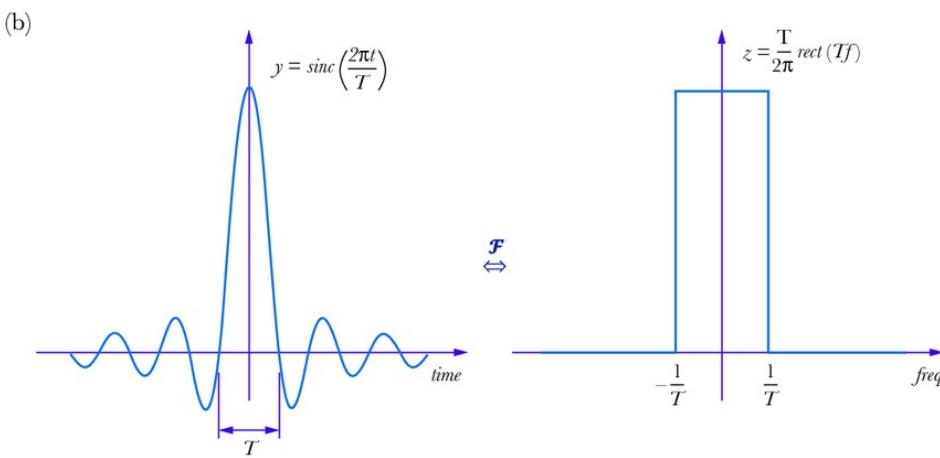
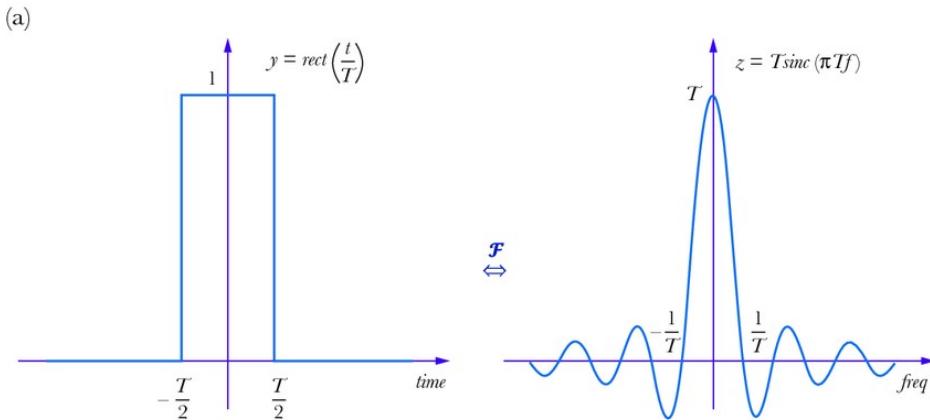


Windowing in Time Domain

Windowing is multiplication in the time domain: the signal is multiplied by the window.

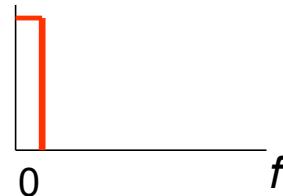
Multiplication in the time domain is like convolution in the frequency domain (and vice versa).

Every point in the frequency domain is modified by the frequency characteristics of the window function.



Windowing (cont)

The ideal window would have a spectrum that was 1.0 at $f = 0$ Hz and 0 everywhere else.



The width of a window's spectrum at $f = 0$ Hz is termed the main lobe.

The nonzero height of a window's spectrum away from $f = 0$ Hz is termed the sidelobe.

The wider the mainlobe the less accurate the frequency resolution: nearby spectral points become averaged together.

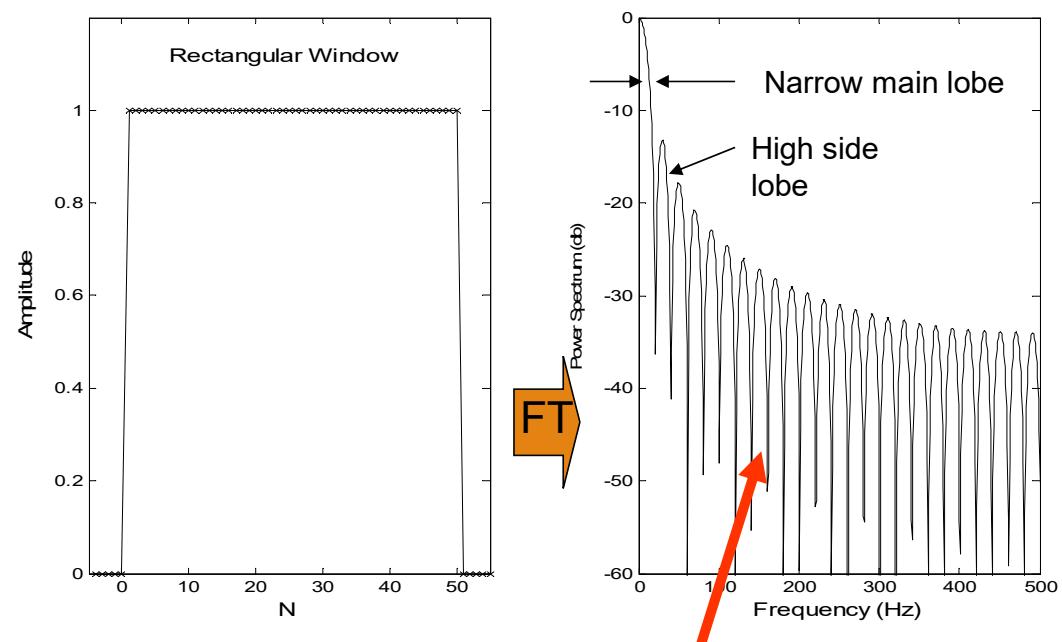
The higher the sidelobes the more adjacent frequencies influence the spectrum and are merged into the spectrum.

Windowing

Each point in a spectrum obtained from windowed signal is influenced by adjacent frequencies.

The spectrum of a rectangular window has high 'sidelobes' (15 dB) which mixes nearby frequencies into each calculated spectral point.

- also has a narrow 'mainlobe' so that its frequency resolution will be 'good'

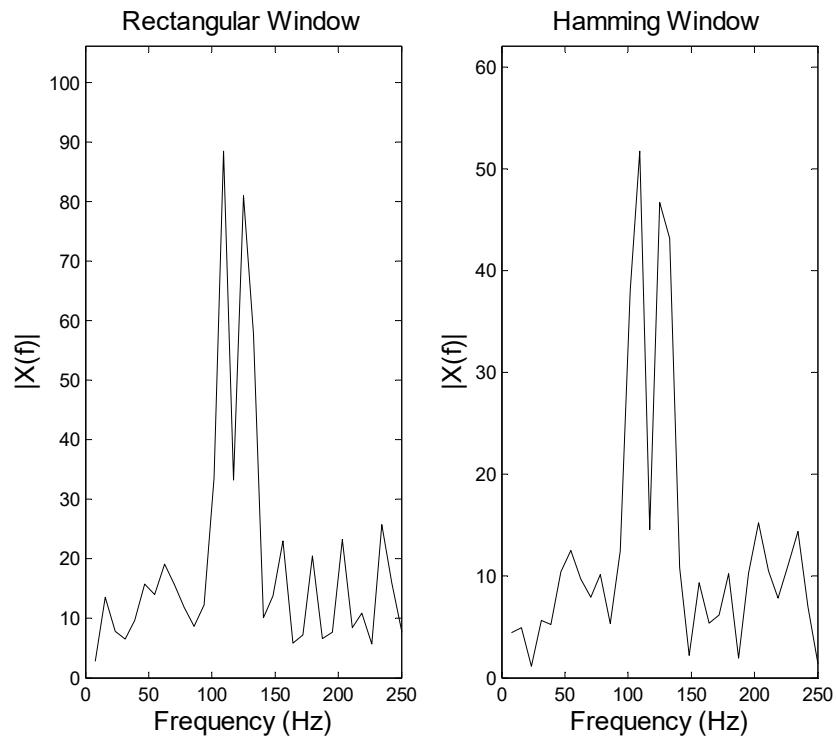


Every spectral point in the FFT is convolved with this curve

Windowing

If the data set is fairly long (perhaps 256 points or more), the benefits of a non-rectangular window are slight.

Figure shows spectra obtained with a rectangular and Hamming window.
They are nearly the same except for a scale difference produced by the Hamming window.



Power Spectrum

In the direct approach, the power spectrum is calculated as the **magnitude squared of the Fourier transform** of the waveform of interest:

$$PS(f) = |X(f)|^2$$

The **power spectrum does not contain phase information** so the power spectrum is not a bilateral transformation -- it is not possible to reconstruct the signal from the power spectrum.

Spectral Averaging: Periodogram

Just as time signals can be averaged, Power Spectra can be **averaged**.

Even if only one signal is available, isolated **segments** of the data can be used.

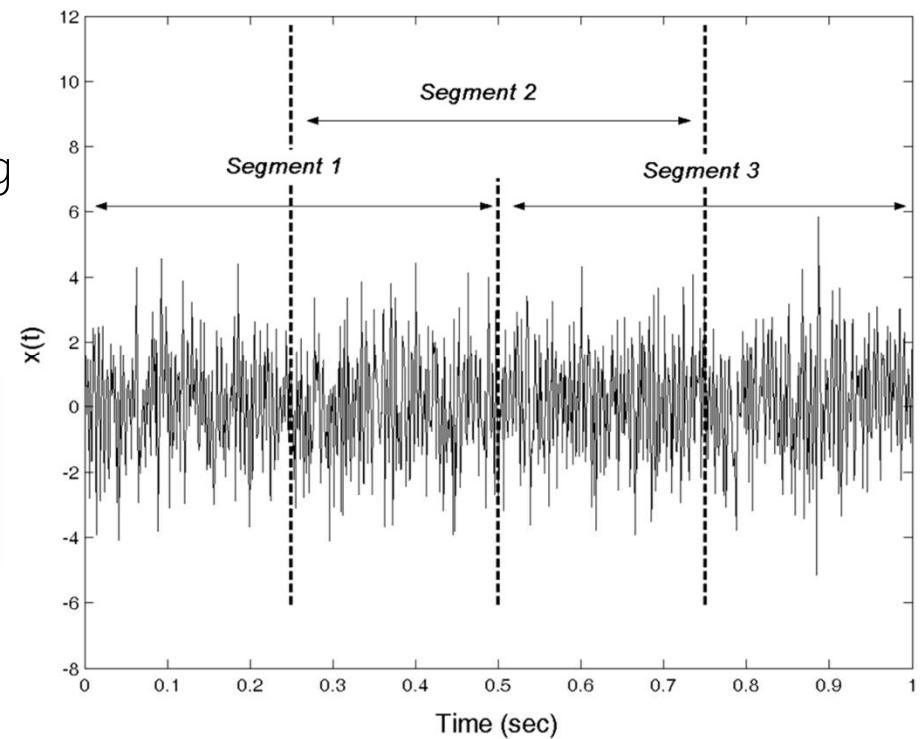
The Power Spectra determined from each segment is averaged to produce a spectrum that better represents the broad, or **global**, features of the spectrum.

When the Power Spectrum is based on a direct application of the Fourier Transform followed by averaging, it is commonly referred to as an averaged **periodogram**.

Spectral Averaging: Welch Method

One of the most common methods to evaluate the average periodogram is attributed to Welch which uses overlapping segments.
(here with 50% overlap)

A shaping window is often applied to each segment because the segments tend to be short.



Spectral Averaging (cont)

Averaging spectra can only be applied to the magnitude spectrum or power spectrum because magnitude spectra are insensitive to time translation.

Averaged periodograms traditionally average spectra from half-overlapping segments, i.e., segments that overlap by 50%

Frequency resolution \propto f_s/N

f_s = sampling frequency

N is # of samples in a segment,

Stationary and Non-stationary Signals

FT identifies all spectral components present in the signal, however it does not provide any information regarding the temporal (time) localization of these components.

Why?

Stationary signals consist of spectral components that do not change in time

- all spectral components exist at all times
- no need to know any time information
- FT works well for stationary signals

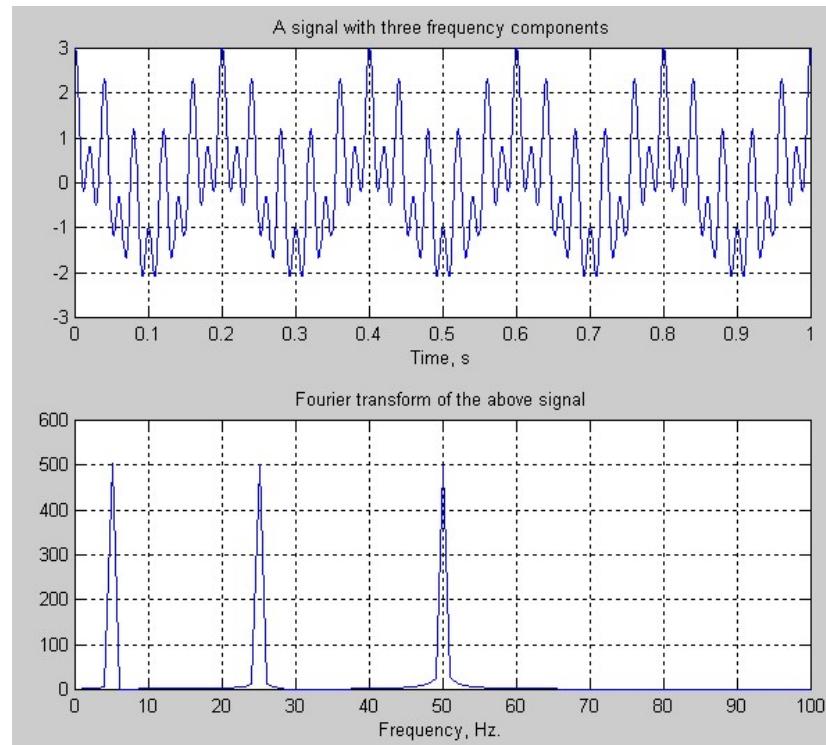
However, non-stationary signals consists of time varying spectral components

- How do we find out which spectral component appears when?
- FT only provides *what spectral components exist*, not where in time they are located.
- Need some other ways to determine *time localization of spectral components*

Recall: FT At Work

$$x_4(t) = \cos(2\pi \cdot 5 \cdot t) + \cos(2\pi \cdot 25 \cdot t) + \cos(2\pi \cdot 50 \cdot t)$$

$$x_4(t) \xleftrightarrow{\mathcal{FT}} X_4(\omega)$$



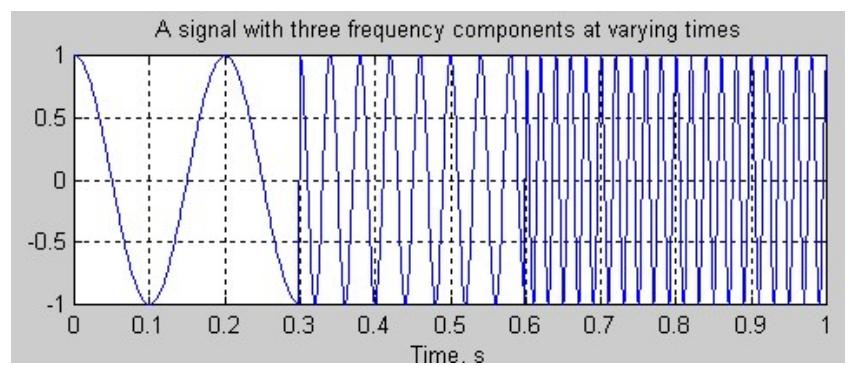
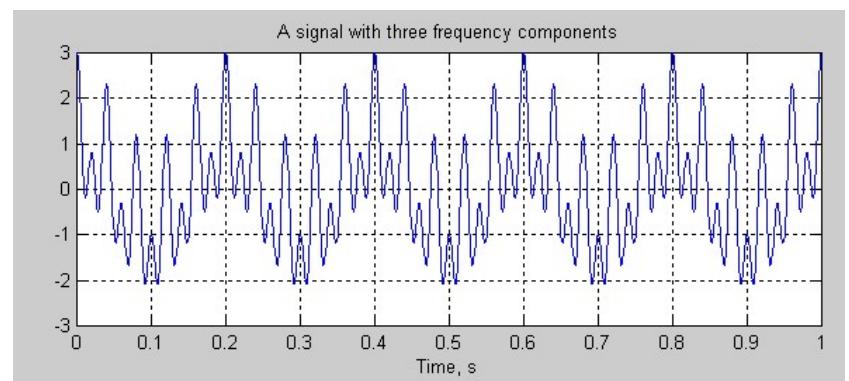
Stationary and Non-stationary Signals

Stationary signals' spectral characteristics do not change with time

$$x_4(t) = \cos(2\pi \cdot 5 \cdot t) + \cos(2\pi \cdot 25 \cdot t) + \cos(2\pi \cdot 50 \cdot t)$$

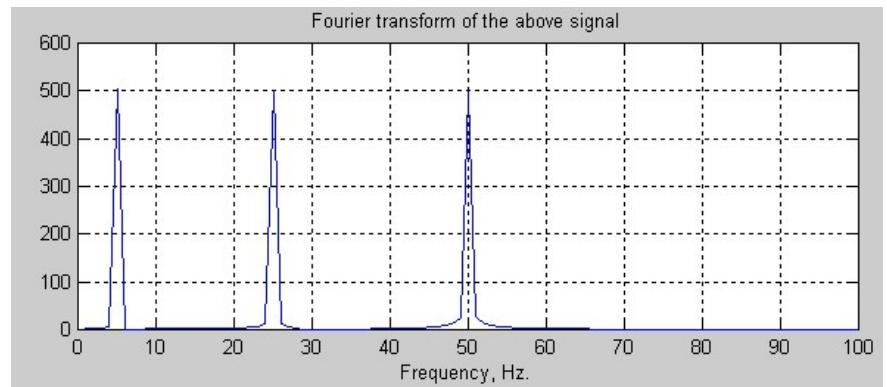
Non-stationary signals have time varying spectra

$$x_5(t) = [x_1 \oplus x_2 \oplus x_3] \oplus \text{Concatenation}$$



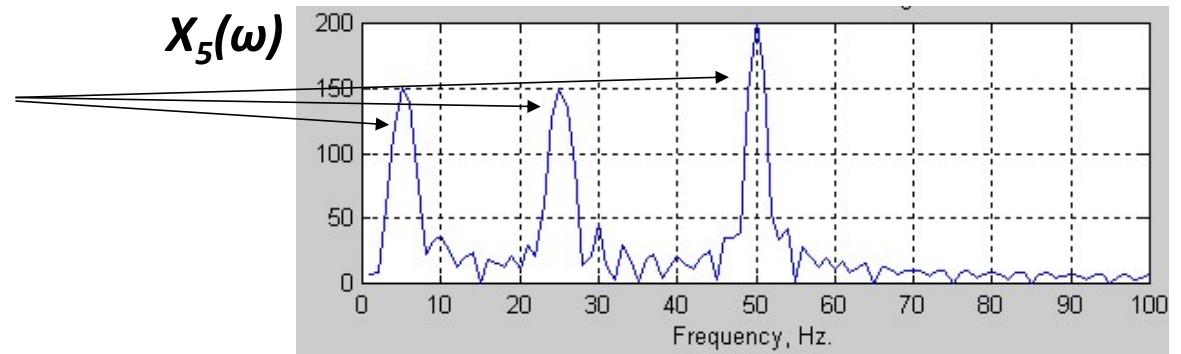
Stationary vs. Non-Stationary

$X_4(\omega)$

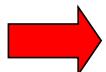


Perfect knowledge of what frequencies exist, but no information about where these frequencies are located in time

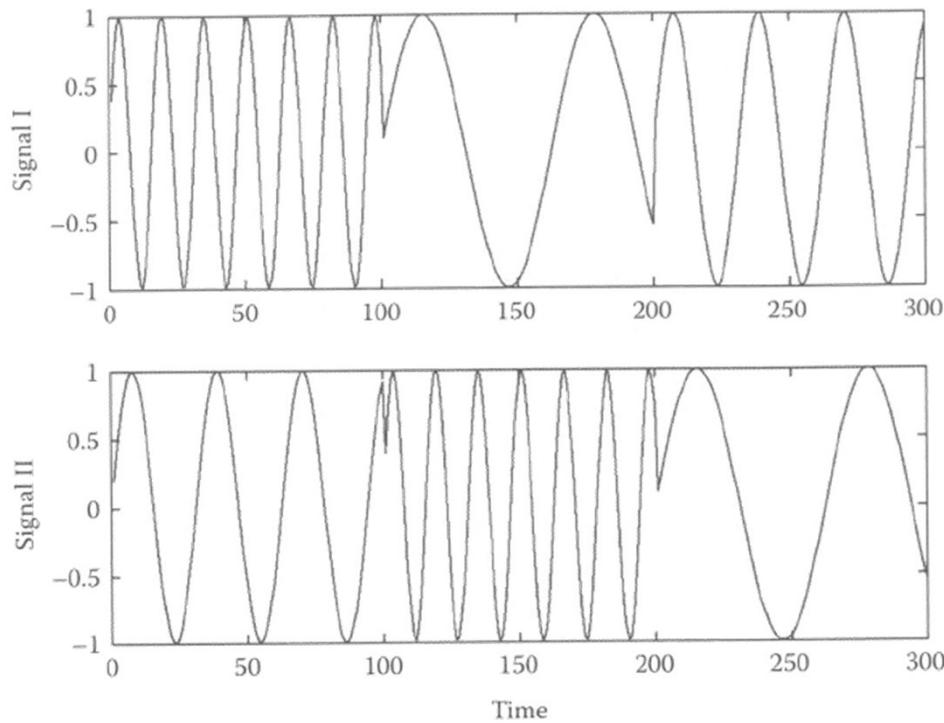
$X_5(\omega)$



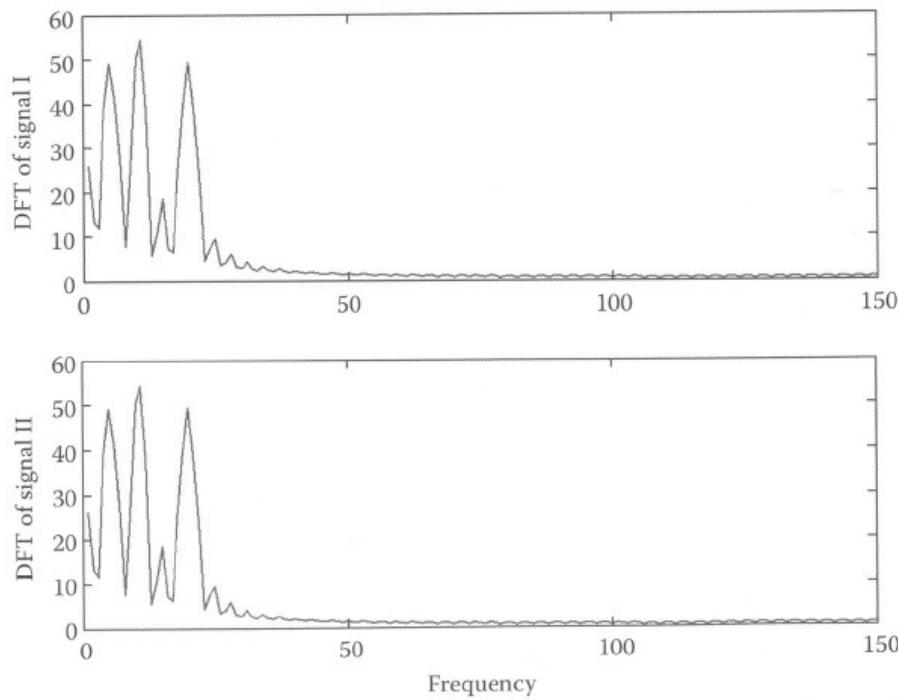
Shortcomings of the FT

- Sinusoids and exponentials
 - Stretch into infinity in time,  no time localization
 - Instantaneous in frequency,  perfect spectral localization
 - *Global* analysis does not allow analysis of non-stationary signals
- Need a *local* analysis scheme for a **time-frequency representation** (TFR) of nonstationary signals
 - Windowed F.T. (AKA Short Time FT, STFT) : Segmenting the signal into narrow time intervals, narrow enough to be considered stationary, and then take the FT of each segment, (Gabor 1946).
 - Followed by other TFRs, which differed from each other by the selection of the windowing function

Example: What will FTs look like?



Example: FTs are identical



Short Time Fourier Transform

define a new version of the FT in which the time localization is preserved. This attempt leads us to the definition of a particular form of the FT called the short-time Fourier transform or STFT.

For a signal $x(t)$, the STFT is defined as

$$X_{\text{STFT}}(a, f) = \int_{-\infty}^{+\infty} x(t)g^*(t - a)e^{-j2\pi ft}dt$$

STFT

$$X_{\text{STFT}}(a, f) = \int_{-\infty}^{+\infty} x(t)g^*(t-a)e^{-j2\pi ft}dt$$

Diagram illustrating the STFT formula:

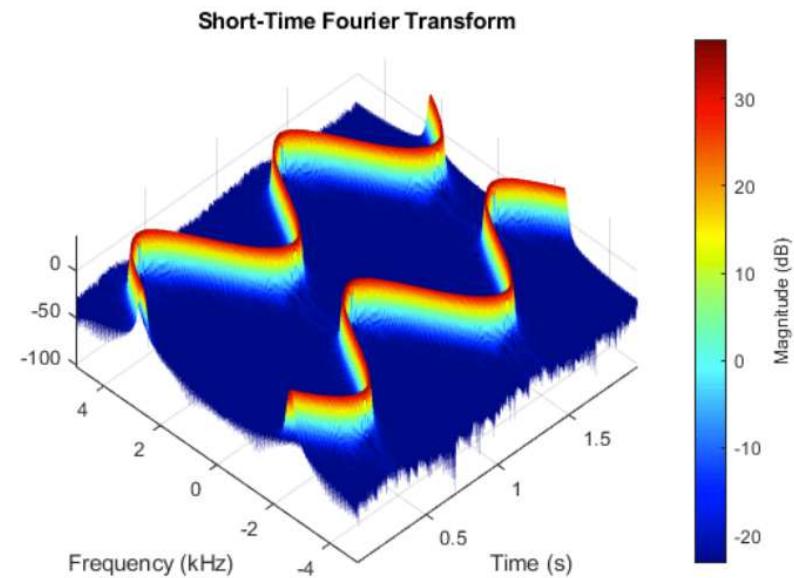
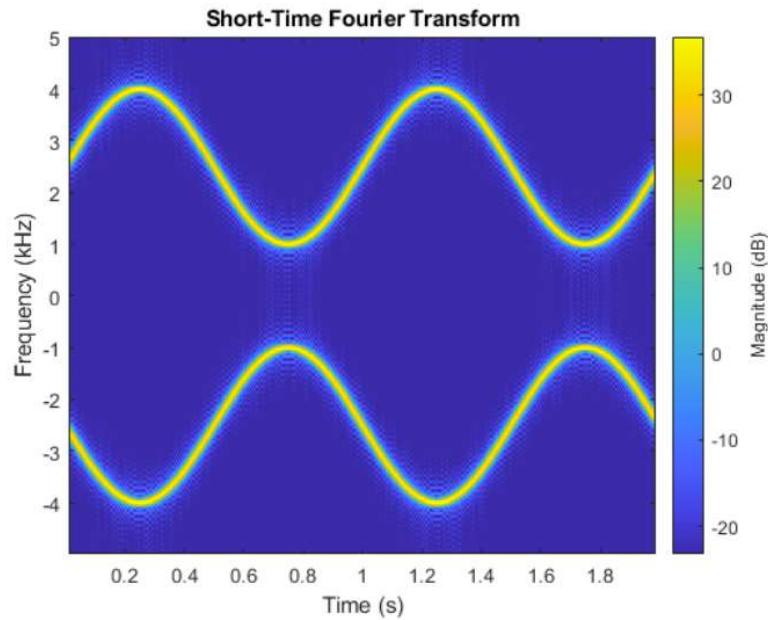
- Time parameter: a (Time axis)
- Frequency parameter: f (Frequency axis)
- Signal to be analyzed: $x(t)$
- FT Kernel (basis function): $g^*(t-a)e^{-j2\pi ft}$

Annotations below the formula:

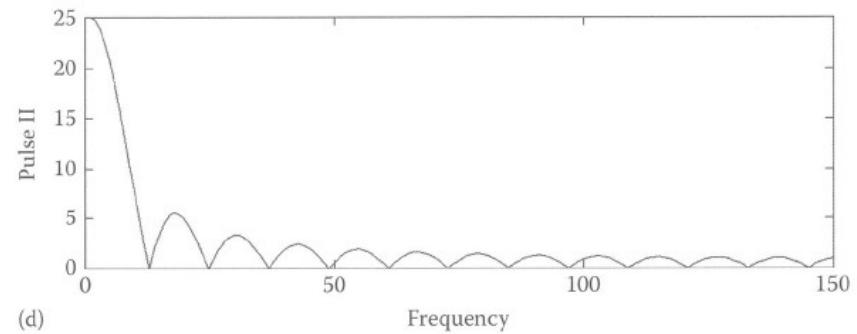
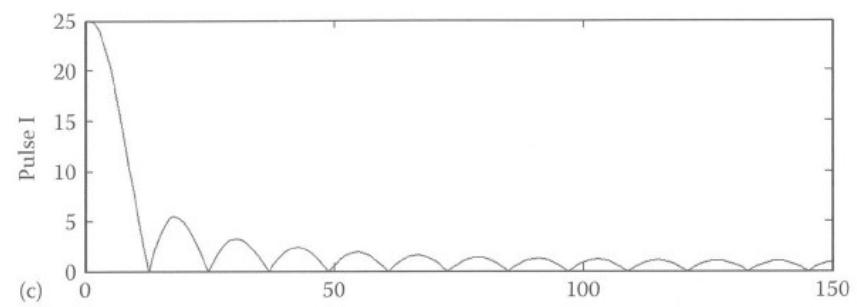
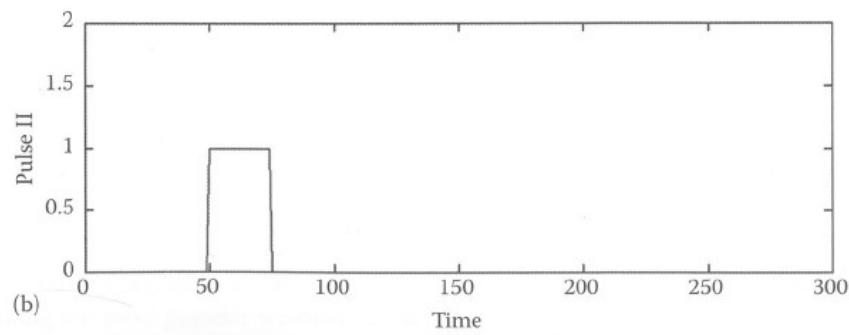
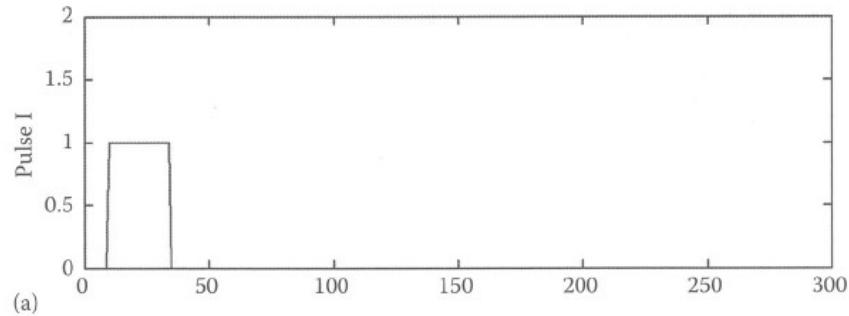
- STFT of signal $x(t)$: Computed for each window centered at $t=a$
- Windowing function
- Windowing function centered at $t=a$

MATLAB

```
stft(x,fs,'Window',kaiser(256,5),'OverlapLength',220,'FFTLength',512);
```



Windowing



Windowing

$$\chi_{\text{STFT}}(a, f) = \int_{-\infty}^{+\infty} x(t)g^*(t-a)e^{-j2\pi ft}dt$$

By definition, $g(t - a)$ is a shifted version of a time window (gate) $g(t)$ that extracts a portion of the signal $x(t)$

i.e. the “gate” $g(t - a)$, having a limited time span, selects and extracts only a portion of the signal $x(t)$ to be analyzed by the FT.

This time window is often a real-time function, and, therefore,

$$g^*(t - a) = g(t - a)$$

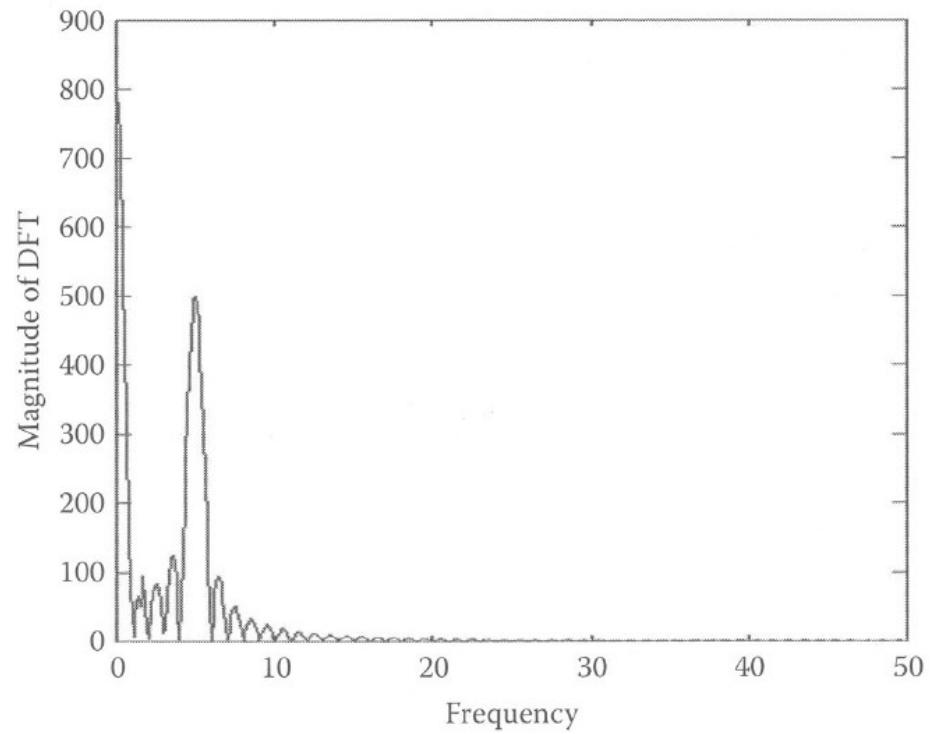
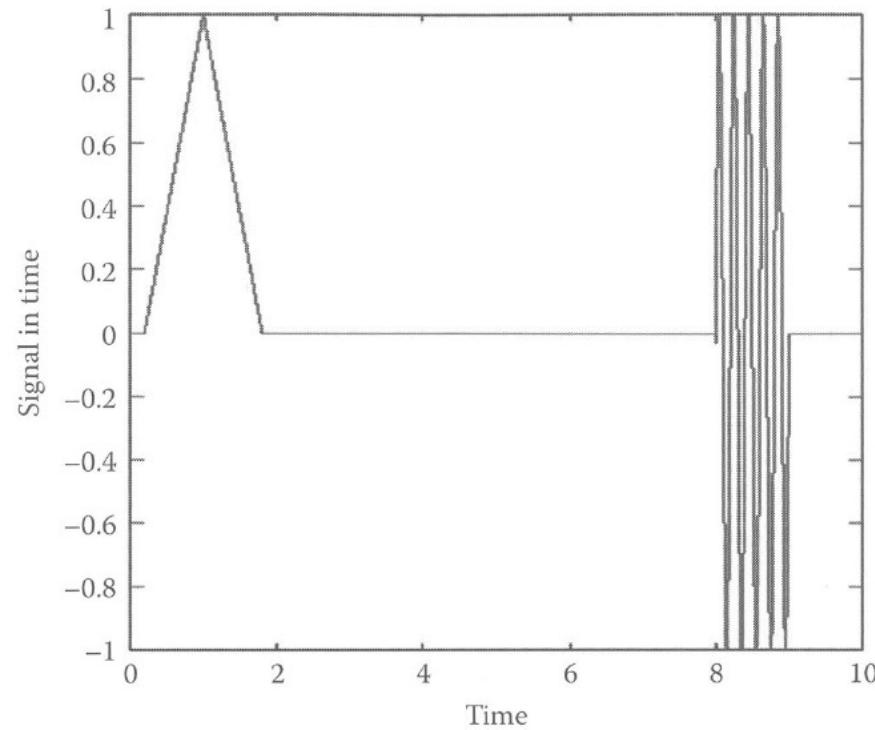
Windowing

So for the STFT, a time window selects a portion of $x(t)$ and then the regular old FT is calculated for this selected part of the signal.

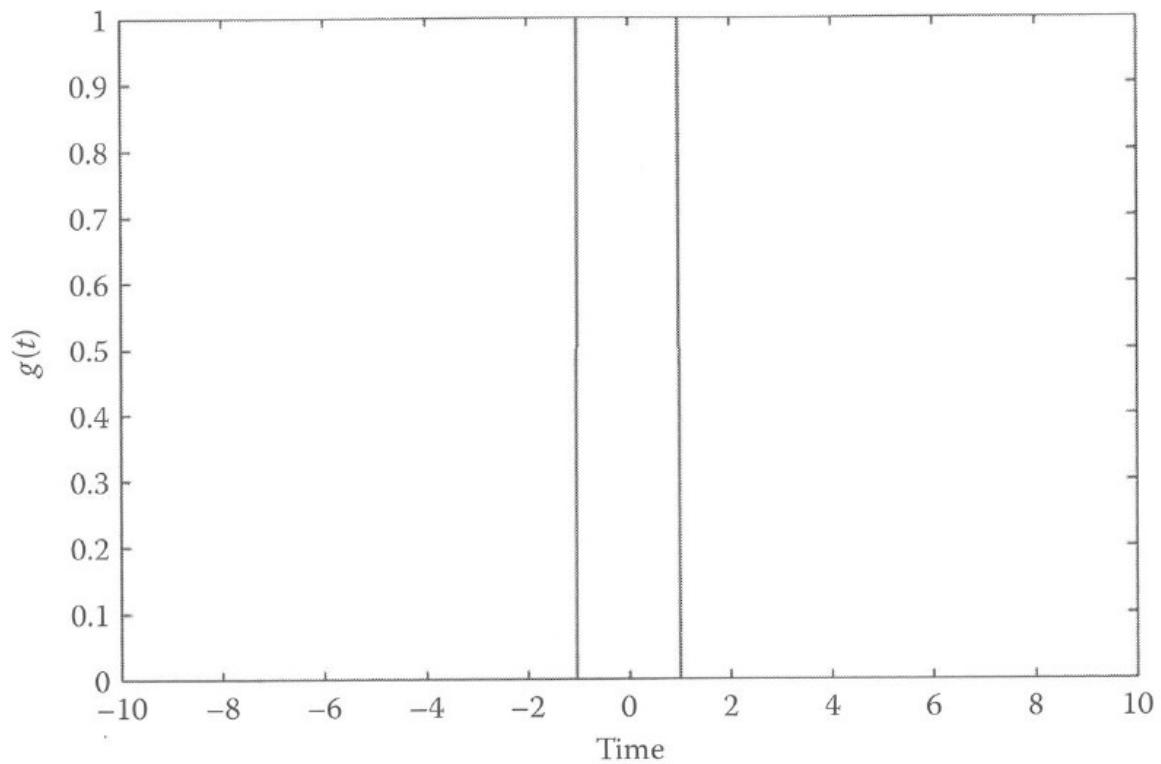
- By changing the amount of shift parameter a , one obtains not only the FT of every part of the signal, but also the time localization of each part as these portions are extracted at known time intervals identified by the shift factor a

STFT has two parameters, f and a . So, there is more computation (compared to FT)

STFT: Example



Window function to be used by the STFT



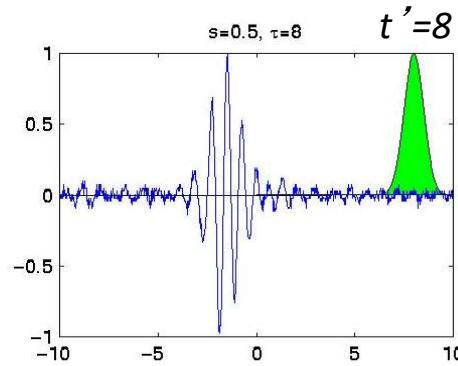
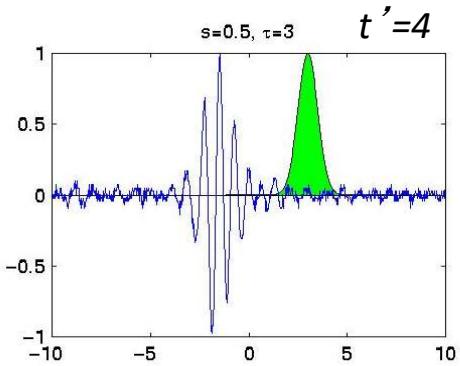
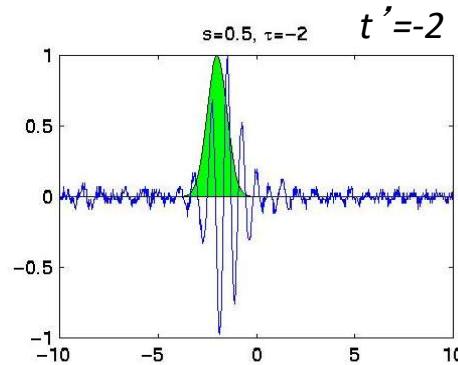
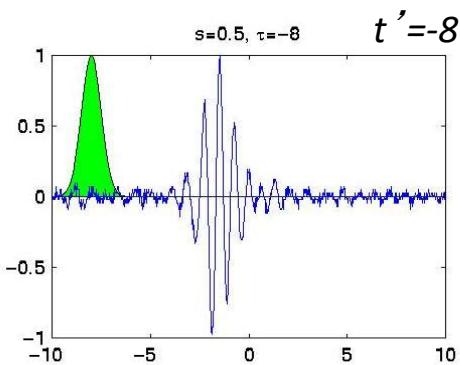
Short Time Fourier Transform (STFT)

- 
1. Choose a window function of finite length
 2. Place the window on top of the signal at $t=0$
 3. Truncate the signal using this window
 4. Compute the FT of the truncated signal, save.
 5. Incrementally slide the window to the right
 6. Go to step 3, until window reaches the end of the signal

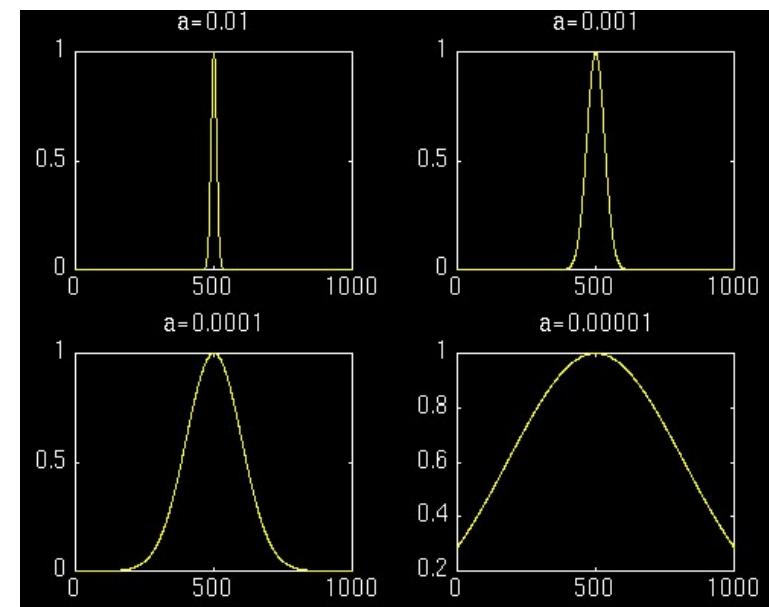
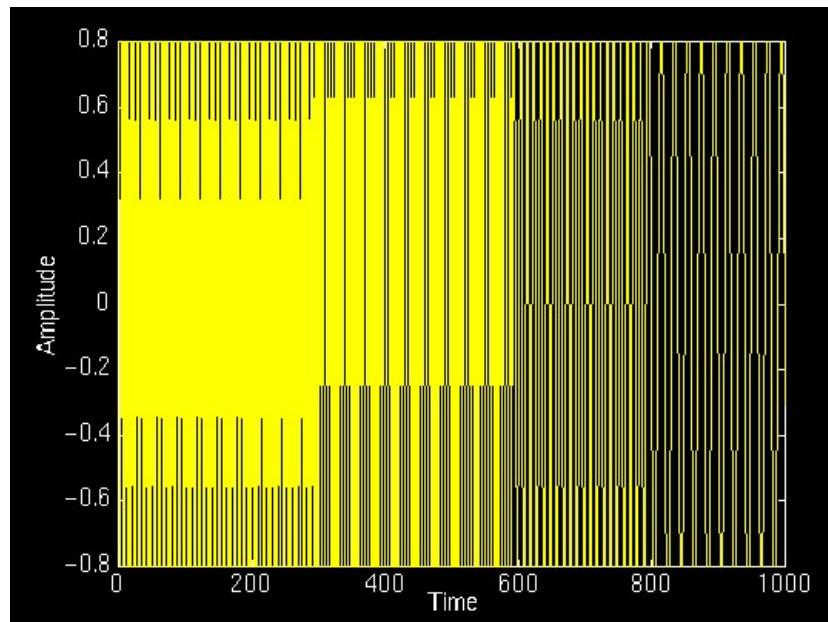
For each time location where the window is centered, we obtain a different FT

- Hence, each FT provides the spectral information of a separate time-slice of the signal, providing simultaneous time and frequency information

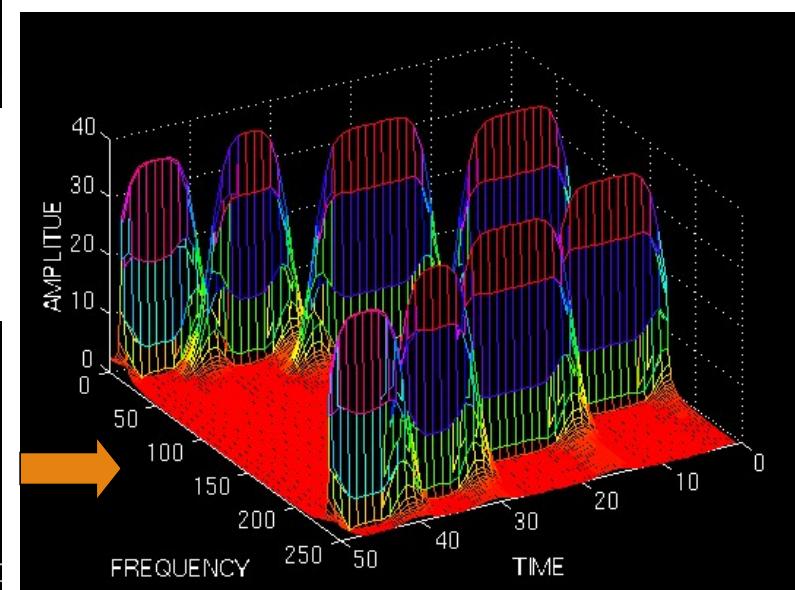
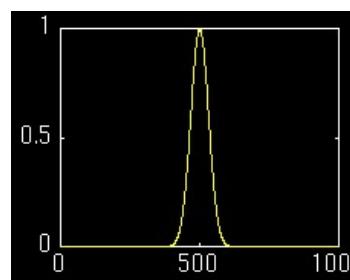
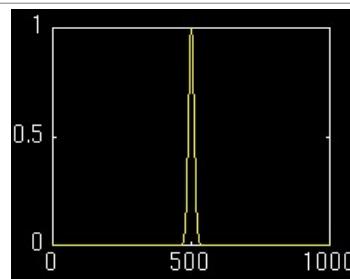
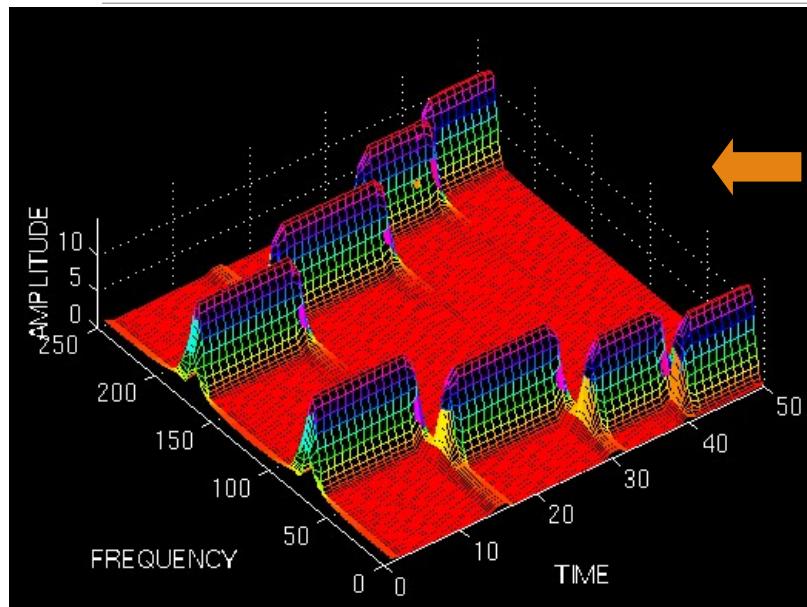
STFT



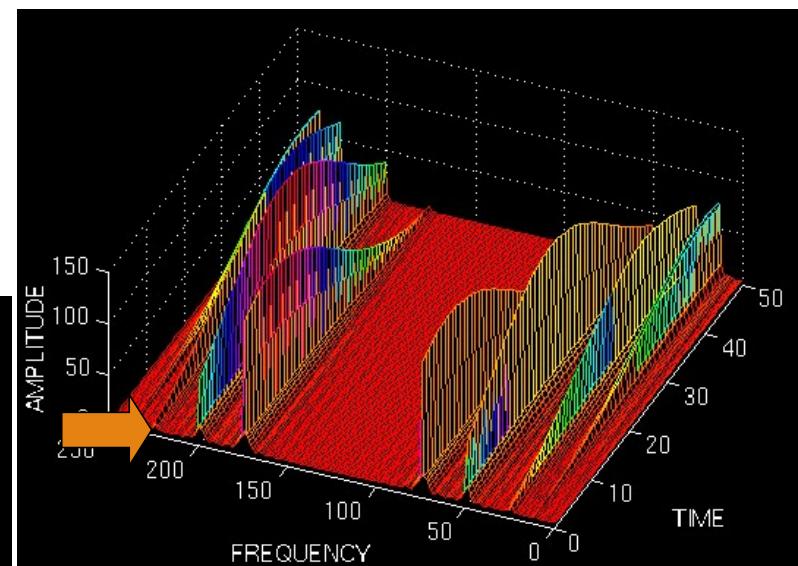
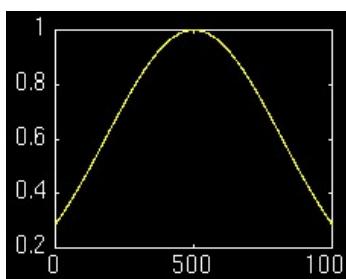
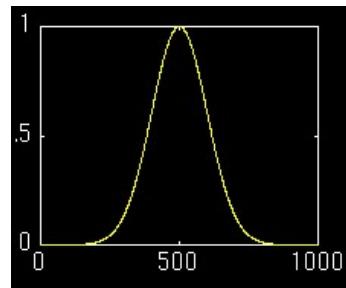
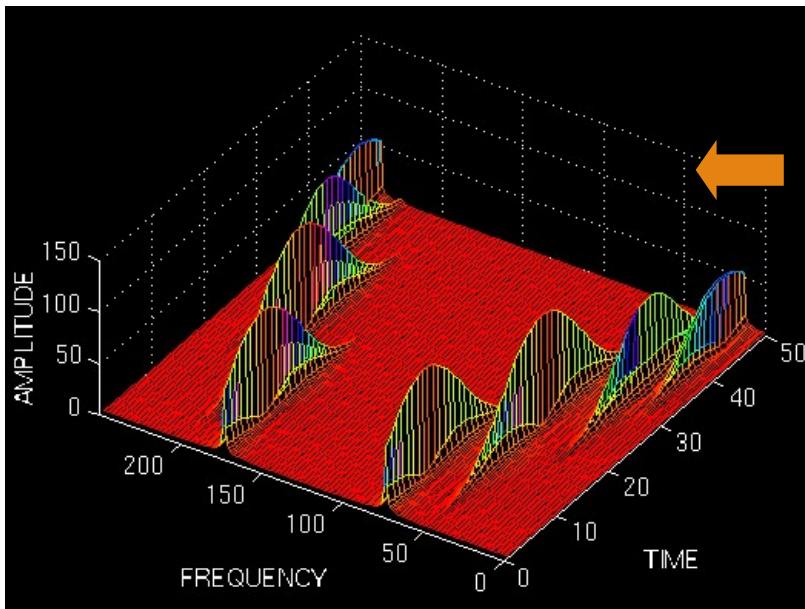
Different Sized Windows



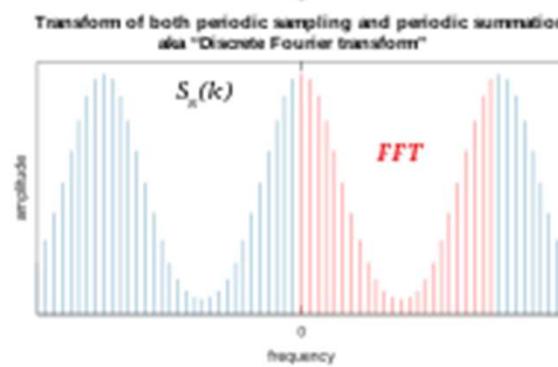
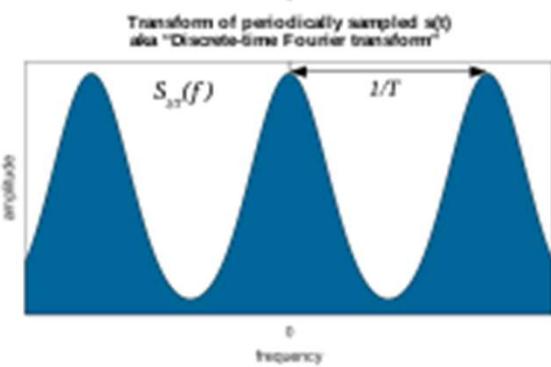
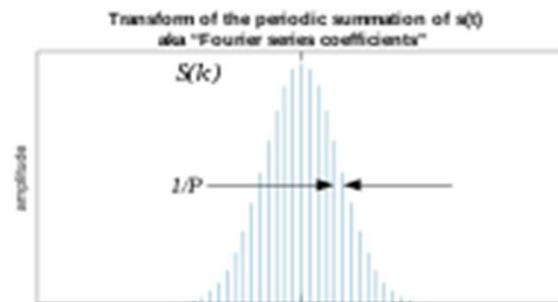
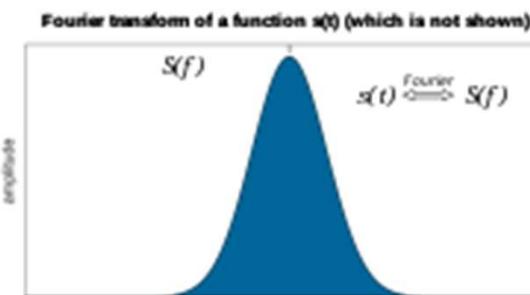
STFT At Work



STFT At Work

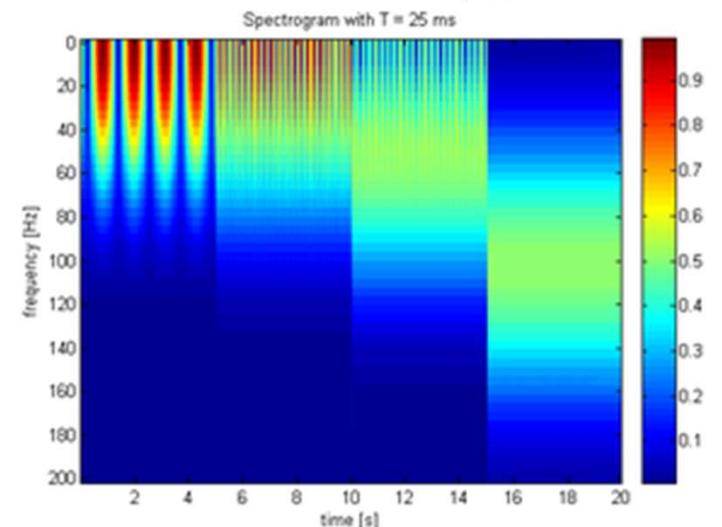


Continuous vs Discrete



$$\text{STFT}\{x[n]\}(m, \omega) \equiv X(m, \omega) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-j\omega m}$$

$$\text{STFT}\{x(t)\}(\tau, \omega) \equiv X(\tau, \omega) = \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-j\omega t} dt$$



STFT

STFT provides the time information by computing a different FTs for consecutive time intervals, and then putting them together

- Time-Frequency Representation (TFR)
- Maps 1-D time domain signals to 2-D time-frequency signals

Consecutive time intervals of the signal are obtained by truncating the signal using a sliding windowing function

How to choose the windowing function?

- What shape? Rectangular, Gaussian, Elliptic...?
- How wide?
 - Wider window require less time steps → low time resolution
 - Also, window should be narrow enough to make sure that the portion of the signal falling within the window is stationary
 - Can we choose an arbitrarily narrow window...?

Selection of STFT Window

Two extreme cases:

$W(t)$ infinitely long: $W(t) = 1$

→ STFT turns into FT, providing excellent frequency information (good frequency resolution), but no time info.

Wide analysis window → poor time resolution, good frequency resolution

$$X_{\text{STFT}}(a, f) = \int_{-\infty}^{+\infty} x(t)g^*(t-a)e^{-j2\pi ft} dt \quad X_{\text{STFT}}(a, f) = \int_t [x(t) \cdot \delta(t-a)] \cdot e^{-j\omega t} dt = x(a) \cdot e^{-j\omega a}$$

Selection of STFT Window

$W(t)$ infinitely short: $W(t) = \delta(t)$

→ STFT then gives the time signal back, with a phase factor. Excellent time information (good time resolution), but no frequency information

Narrow analysis window → good time resolution, poor frequency resolution

Once the window is chosen, the resolution is set for both time and frequency.

Heisenberg Principle

TIME RESOLUTION:

How well two spikes in time can be separated from each other in the transform domain

Both time and frequency resolutions cannot be arbitrarily high!!!

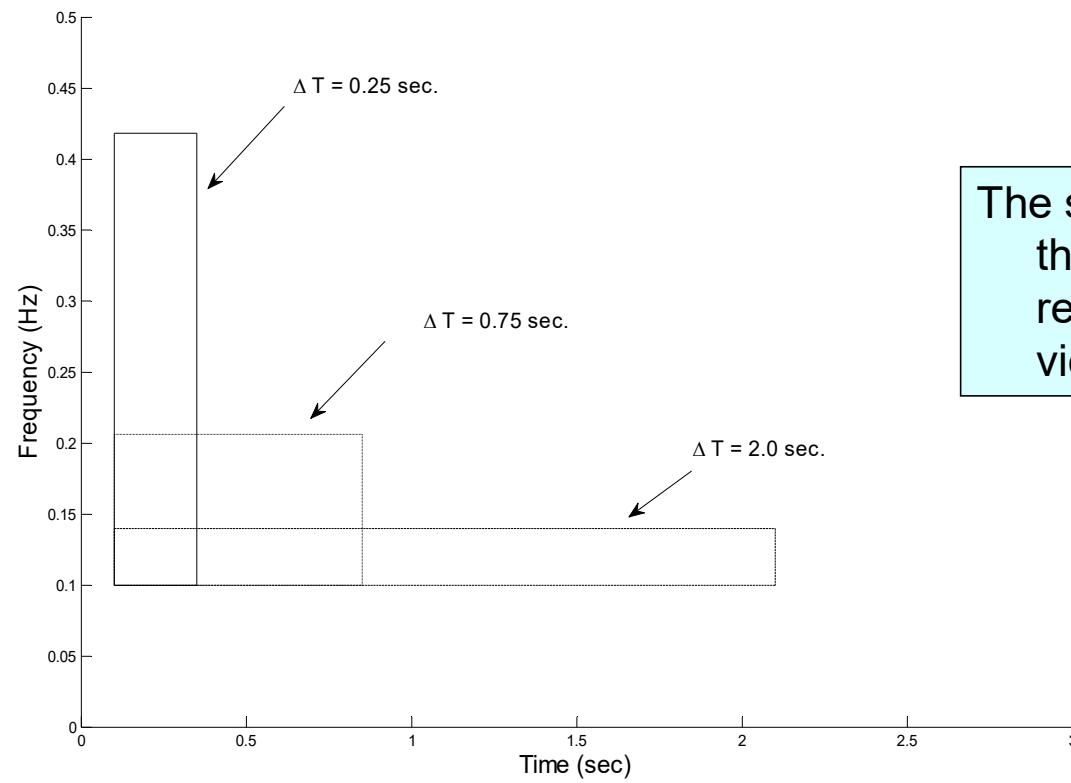
We cannot precisely know at what time instance a frequency component is located.
We can only know what *interval of frequencies* are present in which *time intervals*

$$\Delta t \cdot \Delta f \geq \frac{1}{4\pi}$$

FREQUENCY RESOLUTION:

How well two spectral components can be separated from each other in the transform domain

Time Frequency Tradeoff



The smallest time resolution gives the poorest frequency resolution (black rectangle) and vice versa (dashed rectangle).

STFT Conclusion

Despite the favorable properties of the STFT, this transform is not the best solution to address the time and frequency localization of signal events.

shortcomings of the STFT,

- choice of the window length:
 - A too short window may not capture the entire duration of an event
 - too long window may capture two or more events in the same shift.
- nature of the basis functions used in the FT, i.e., complex exponentials.
 - The term $e^{-i2\pi ft}$ describes sinusoidal variations in real and complex spaces. Such functions are not limited in time span or duration

WHAT TO DO NOW?

The Wavelet Transform

Overcomes the preset resolution problem of the STFT by using a variable length window

Analysis windows of different lengths are used for different frequencies:

- Analysis of high frequencies → Use narrower windows for better time resolution
- Analysis of low frequencies → Use wider windows for better frequency resolution

This works well, if the signal to be analyzed mainly consists of slowly varying characteristics with occasional short high frequency bursts.

Heisenberg principle still holds!!!

The function used to window the signal is called *the wavelet*

Wavelets

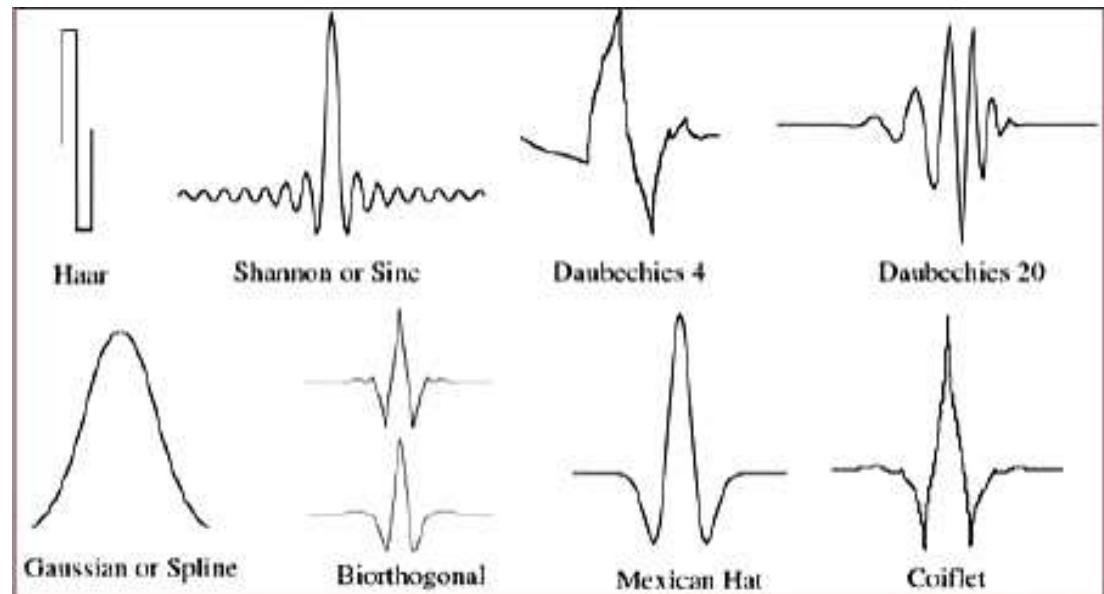
Wave-like function

Amplitude starts and ends at 0

Crafted with specific functions
in mind

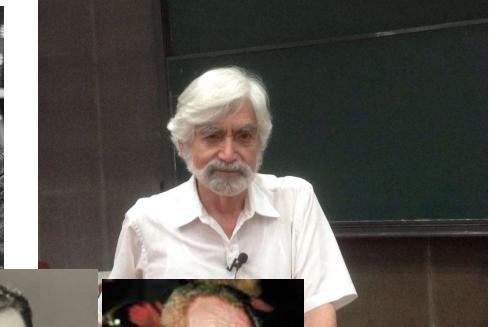
Can be used for 1d or 2d
signals

- i.e. audio signals, eeg signals etc
- Images

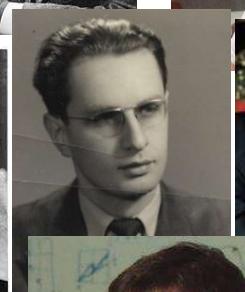


Mother and Father's of Wavelets

Haar – started the train of thought

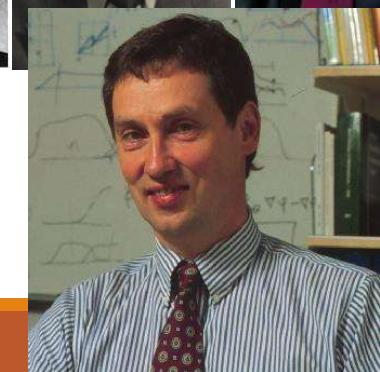


Gabor – Gabor atoms



Zweig – CWT

Gopillard, Grossman, Morlet – Refined
Zweigs work



Jan-Olov Strömberg – DWT

Ingrid Daubechies – orthogonal wavelet

JPEG 2000 – image compression

Wavelet “Frequency”

Before introducing the Wavelet Transform (WT), let's take a closer look at the basic definition of "frequency":

- the fundamental concept of the Fourier Transform (FT)
 - uses periodic time-unlimited functions
 - based on what we have seen we need to use time-limited basis functions for our new (i.e. WT) transform
- not using periodic sinusoidal basis functions for the new transform, we need to think of a concept that replaces frequency
- Time-limited basis functions are obviously not periodic, and therefore, we need to invent a new concept that can represent a concept similar to frequency

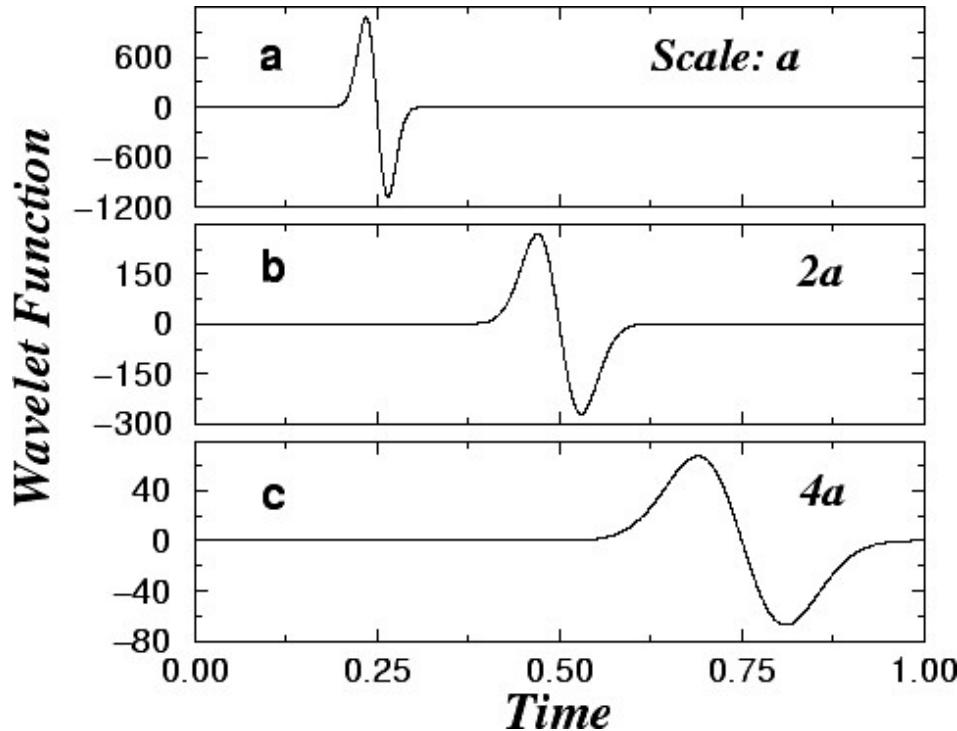
Wavelet “Frequency”

- Consider a sinusoidal basis function with frequency = 0.1 Hz.
- another basis function in the Fourier decomposition of this signal would be the 2nd harmonic (i.e., a sinusoidal basis function with frequency = 0.2 Hz).
- harmonic relations among the basis signals is the fundamental concept of signal transformation and decomposition

Therefore, the relation among harmonics is something that can somehow represent a new concept that will replace frequency.

Wavelet “Frequency”

- make the following important observation about harmonics:
 - by warping the time axis "t" one can obtain the harmonics from the original signal
 - E.g. By replacing the time axis "t" in the original signal with " $2t$ " time axis results in the second harmonic.
- This is essentially "scaling" the signal in time to generate other basis functions.
- main characteristic of harmonic frequencies can be drawn from a more general concept: "scale"



Wavelet Scale

- unlike frequency (defined only for periodic signals) scale is equally applicable to nonperiodic signals.
- Using scale as a variable, the new transform, which will be based on time-limited basis function, can be meaningfully applied to both time-unlimited and time-limited signals.

Continuous Wavelet Transform

$$W_{\Psi,X}(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} x(t) \Psi^* \left(\frac{t-b}{a} \right) dt, \quad a \neq 0$$

$\Psi(t)$ is a function with limited duration in time

b is a shifting parameter. Translates the function across $x(t)$ (like the t in the STFT) a is a time scaling parameter (replaces frequency parameter f)

So, the basis functions of the CWT are the shifted and scaled version of the probing function $\Psi(t)$ (mother wavelet)

The * indicates complex conjugation and dividing by $\sqrt{|a|}$ normalizes energy.

Continuous Wavelet Transform

Wavelets are a class of basis functions that incorporate two parameters:

1. translation in time
2. scaling in time

main point is to accommodate temporal information

(e.g. crucial in evoked responses, aka event related potential (ERP) analysis)

Another definition:

A wavelet is an oscillating function whose energy is concentrated in time to better represent transient and nonstationary signals.

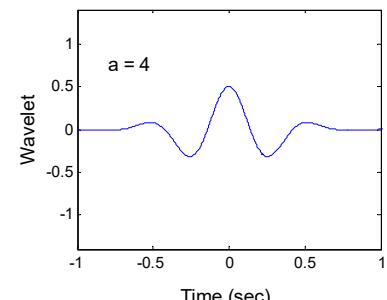
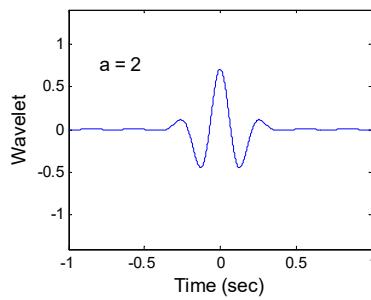
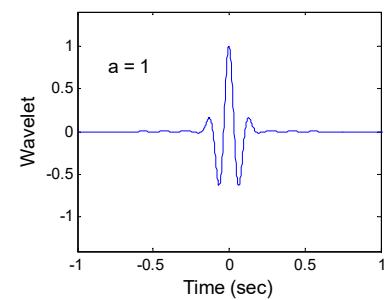
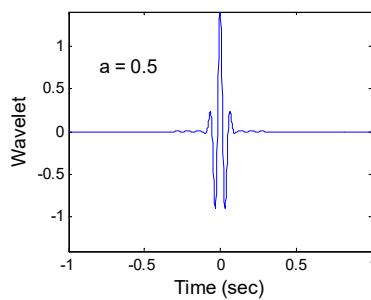
Wavelet Families - Morlet

The wavelet shown is the ‘Morlet Wavelet’ described by the equation:

$$\psi(t) = e^{-t^2} \cos\left(\pi\sqrt{\frac{2}{\ln 2}}t\right)$$

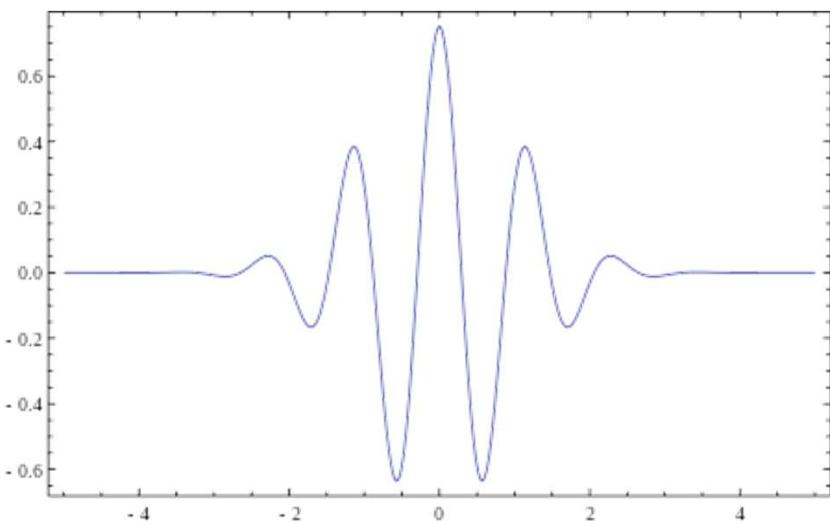
If $b = 0$, and $a = 1$, then the wavelet is in its basic form, the "mother wavelet."

If $a > 1$, the wavelet is stretched along the time axis, and if $a < 1$, the wavelet is contracted.

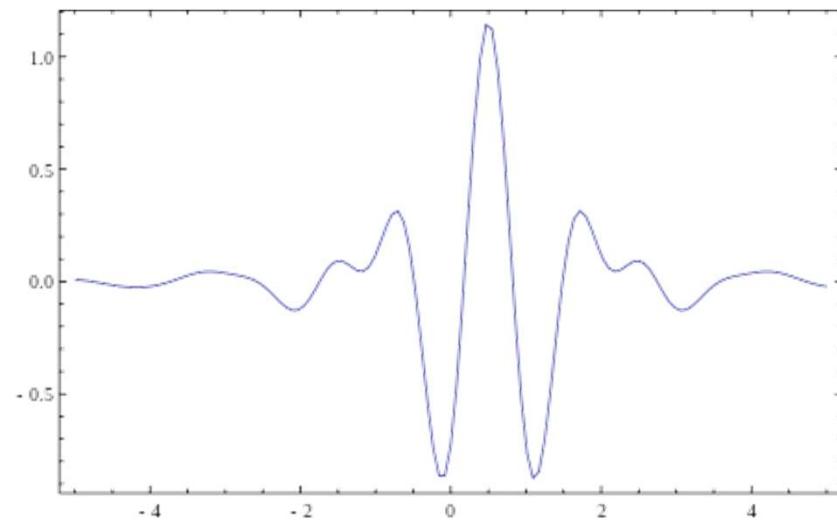


Wavelet Families - Meyer

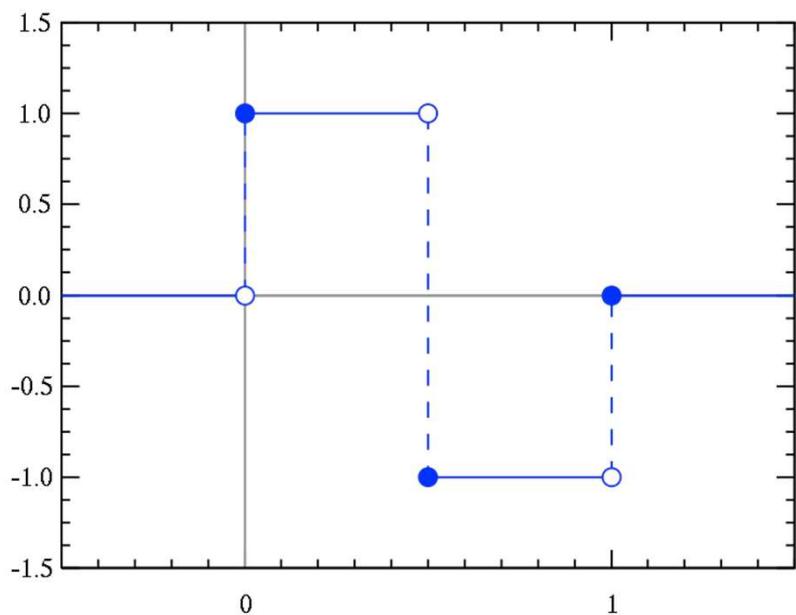
Morlet Wavelet(real valued)



Meyer Wavelet



Wavelet Families - Haar



Haar wavelet mother wavelet function:

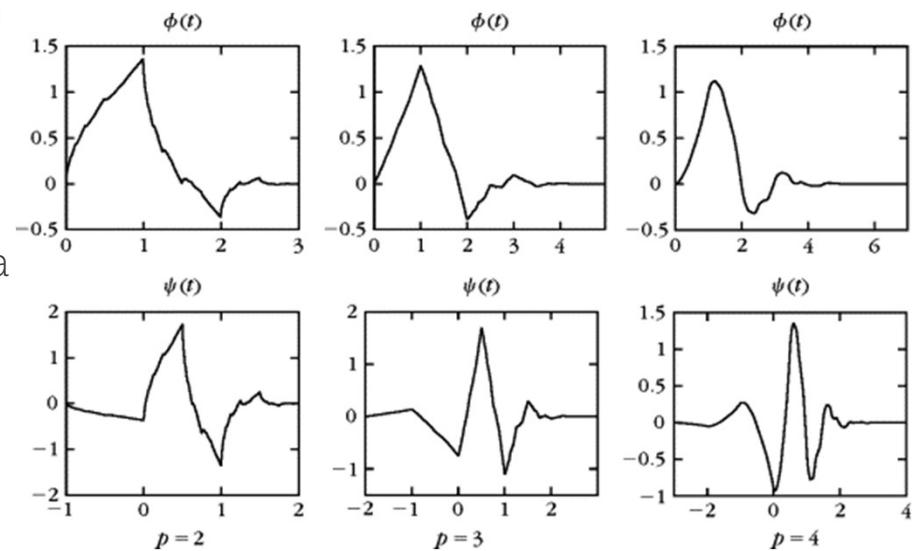
$$\Psi(t) = \begin{cases} 1 & 0 \leq t < 0.5 \\ -1 & 0.5 \leq t < 1 \\ 0 & otherwise \end{cases}$$

Haar wavelet scaling function:

$$\phi(t) = \begin{cases} 1 & 0 \leq t < 1 \\ 0 & otherwise \end{cases}$$

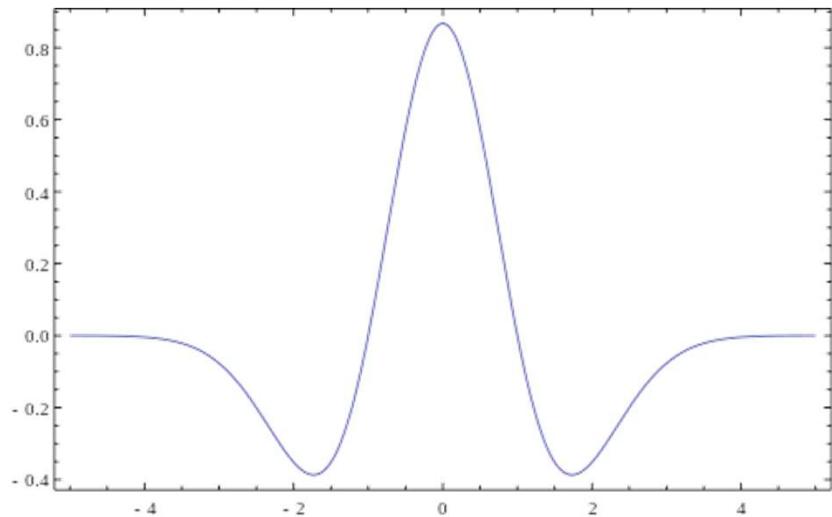
Wavelet Families - Daubechies

- Daubechies (dbX) wavelets are among the most popular mother wavelets that are commonly used in signal and image processing.
- The index "X" in dbX identifies the exact formulation of the function
- all dbX functions look more or less similar, db2 is a simpler mother wavelet than db3, db4, etc.
As a rule of thumb, more complex mother wavelets may be needed to analyze more complex signals.



NOTES

- every choice of mother wavelet gives a particular CWT
- Any choice of mother wavelet gives certain unique properties that make the resulting transformation a suitable choice for a particular task. E.g. “Mexican Hat”,
Infinite number of transformations

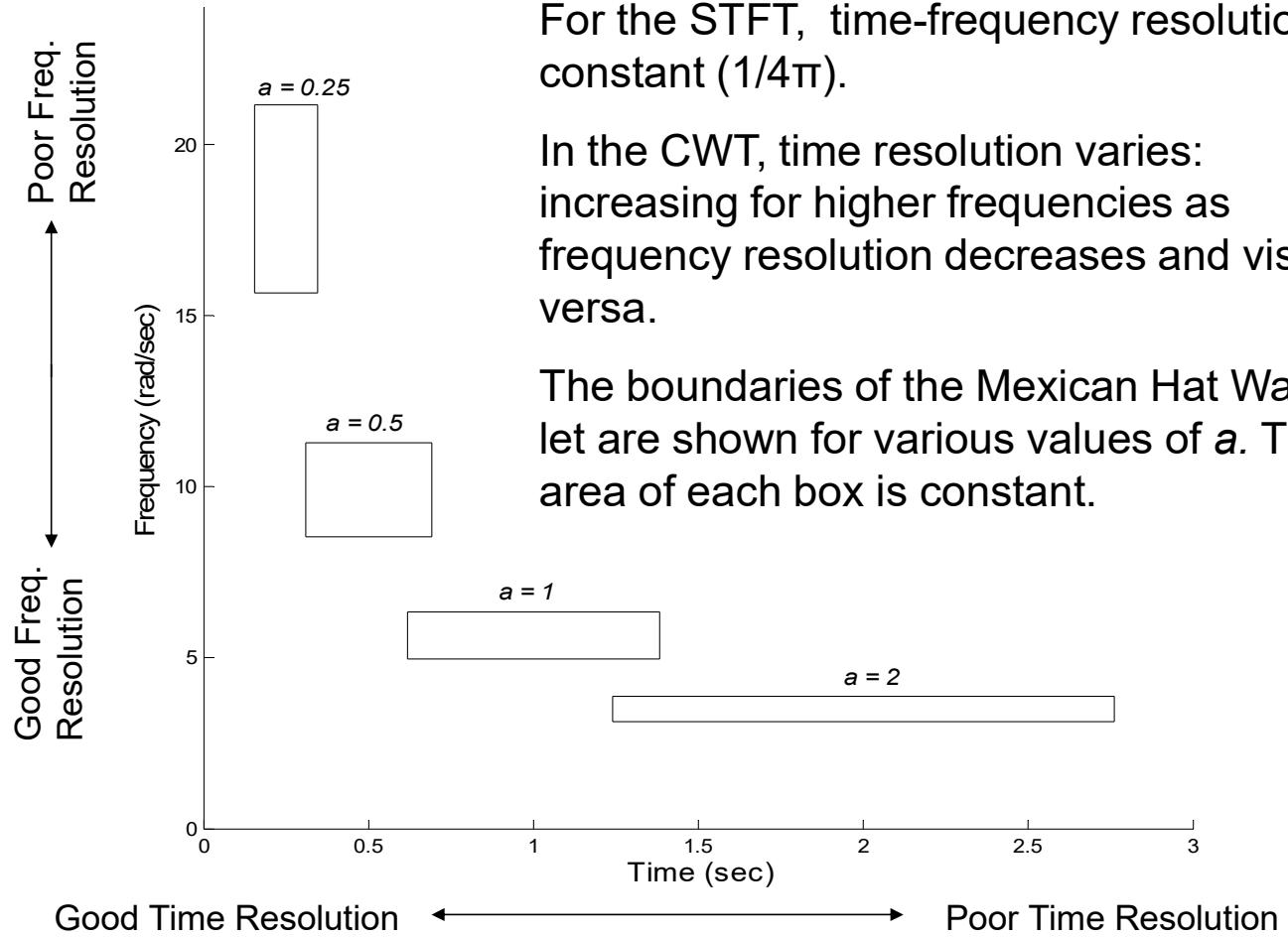


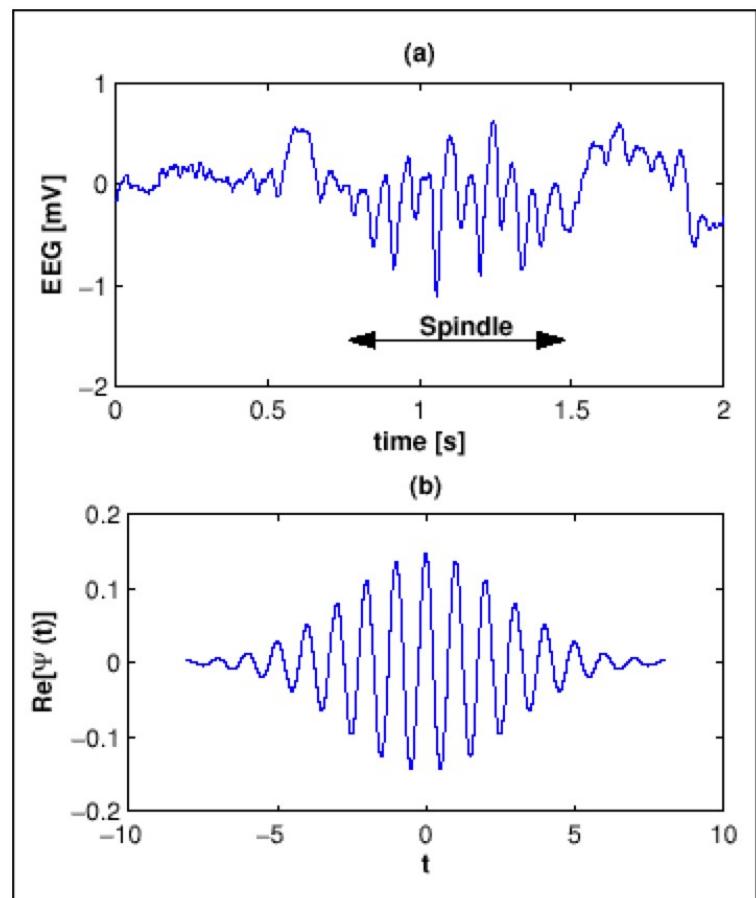
Wavelets: Time-Frequency Trade-off

Built in tradeoff between time and frequency:

$$\Delta\omega_\Psi(a) \Delta t_\Psi(a) = \Delta\omega_\Psi \Delta t_\Psi = \text{constant} \geq 0.5$$

This trade-off is illustrated in the next slide which shows the time-frequency boundaries for the Mexican hat wavelet for various values of a .





M. LATKA¹, A. KOZIK², J. JERNAJCZYK¹, B.J.WEST³, W. JERNAJCZYK⁴

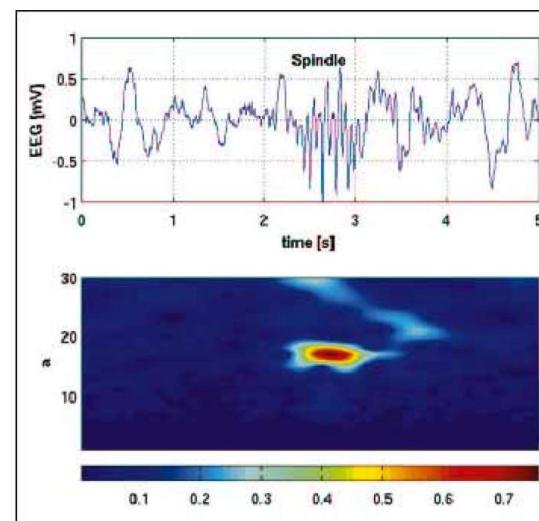
WAVELET MAPPING OF SLEEP SPINDLES IN YOUNG PATIENTS WITH EPILEPSY

¹Institute of Physics, Wroclaw University of Technology, Wroclaw, Poland;

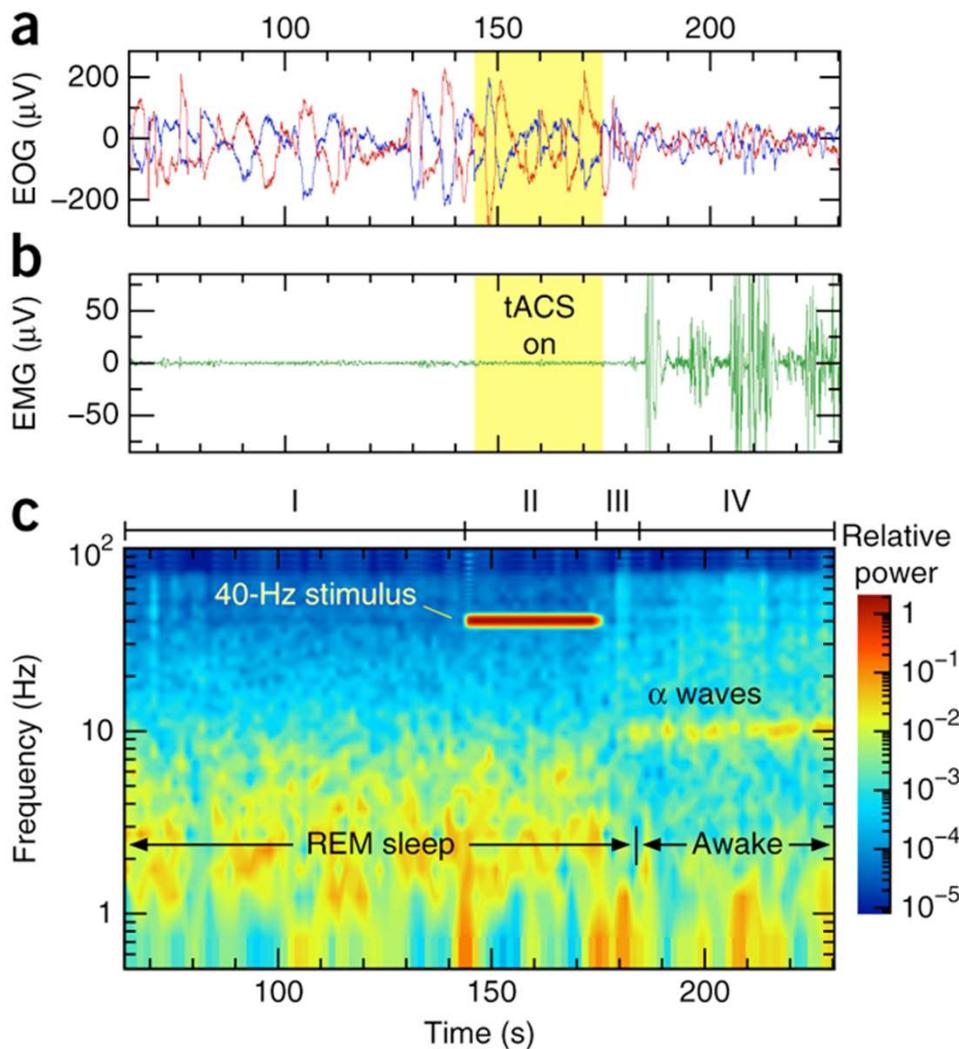
²Video EEG Laboratory, Department of Child Neurology, Marciniak Regional Medical Center, Wroclaw, Poland

³Mathematical & Information Science Directorate, Army Research Office, Research Triangle Park, NC, USA;

⁴Sleep Disorders Center, Department of Clinical Neurophysiology, Institute of Psychiatry and Neurology, Warsaw, Poland



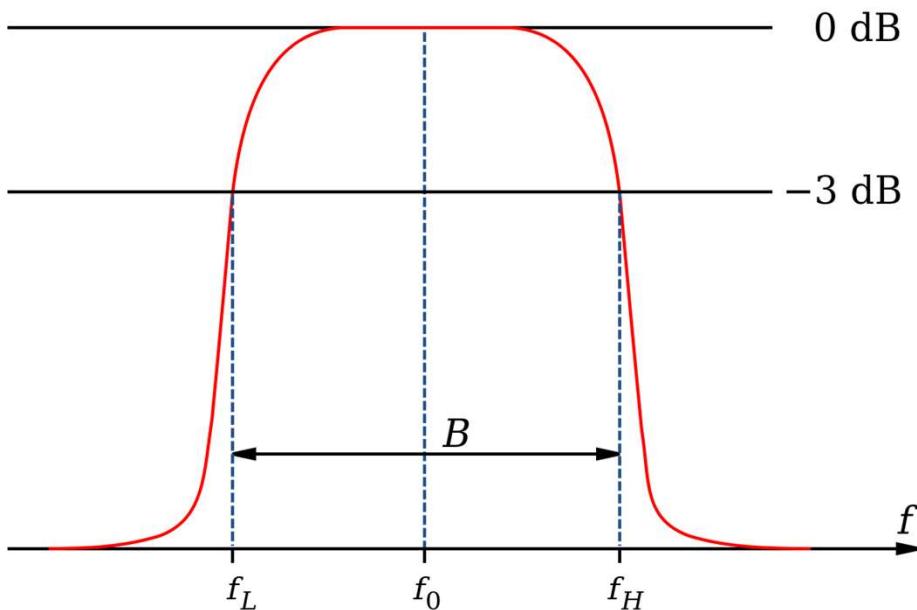
Morlet wavelet



(a) 2-channel EOG showing classical contralateral eye movements typical for REM sleep. Eye movements were synchronous before (phase I), during (phase II) and after stimulation (phase III), and only changed after awakening (phase IV).

(b) EMG activity unchanged until the subject awakened (phase IV), at which time it strongly increased signaling a loss of REM sleep atonia.
(c) Continuous wavelet transform of the recorded EEG signal at Fpz using the complex Morlet wavelet

Other ways to view the CWT



The CWT can be interpreted as a *linear filtering* operation

- convolution between the signal $x(t)$ and a filter with impulse response $\psi(-t/s)$

The CWT can be viewed as a type of bandpass analysis:

- where the scaling parameter (s) modifies the center frequency and the bandwidth of a bandpass filter

Wavelet rules

- By definition the mother wavelet must be limited in duration and looks like a decaying small wave.
- All other basis functions are the shifted and scaled version of the mother wavelet.
- In contrast to the STFT, instead of using sinusoidal functions in order to generate the basis functions, a mother wavelet is continuously shifted and scaled to create all basis functions in CWT.

ICWT

How to get back to our time domain signal?

→ The inverse continuous wavelet transform or ICWT:

$$x(t) = \frac{C_{\Psi}^{-1}}{a^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W_{\Psi,X}(a,b) \Psi\left(\frac{t-b}{a}\right) da db, \quad a \neq 0$$

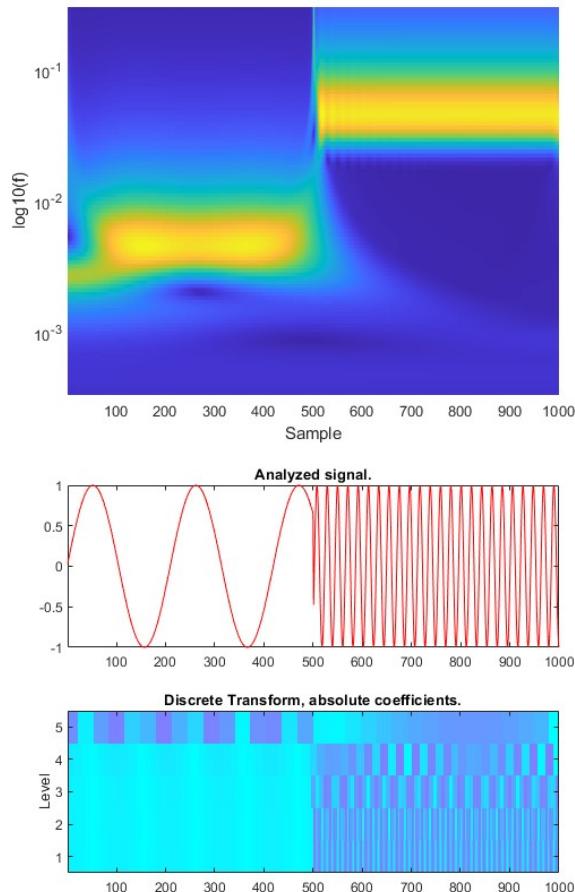
Here $C^{-1}\Psi$ is a constant depending on choice of mother wavelet $\Psi(t)$.

How to choose a mother wavelet?

-an open problem, without a definite answer.

However, 2 “intuitive” rules of thumb are widely followed:

- (1) complex mother wavelets are needed for complex signals
- (2) the mother wavelet that resembles the general shape of the signal to be analyzed would be the more suitable choice



Discrete Wavelet Transform

- concerns are the same as applicable to the CFT
- CFT (and ICFT) are computationally more expensive
- The DWT applies only discrete shifts and scales on signals

Discrete Wavelet Transform

The CWT is highly redundant:

- many more coefficients are generated than needed uniquely to specify the signal.

If the application calls for recovery of the original signal, all of the coefficients will be required and the computational effort could be excessive.

The “Discrete Wavelet Transform” (DWT):

- a) restricts the variation in translation and scale, usually to powers of 2;
- b) downsamples lower scaled data.

Discrete Wavelet Transform

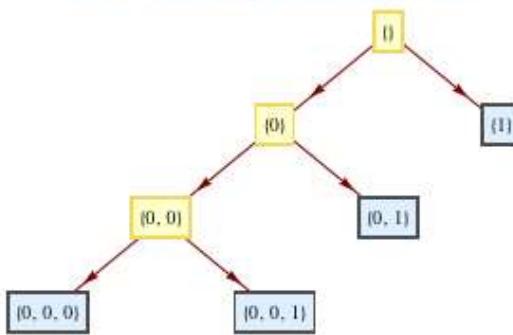
In the DWT, a new concept is introduced termed the “scaling function,” a function that facilitates computation of the DWT.

To implement the DWT efficiently

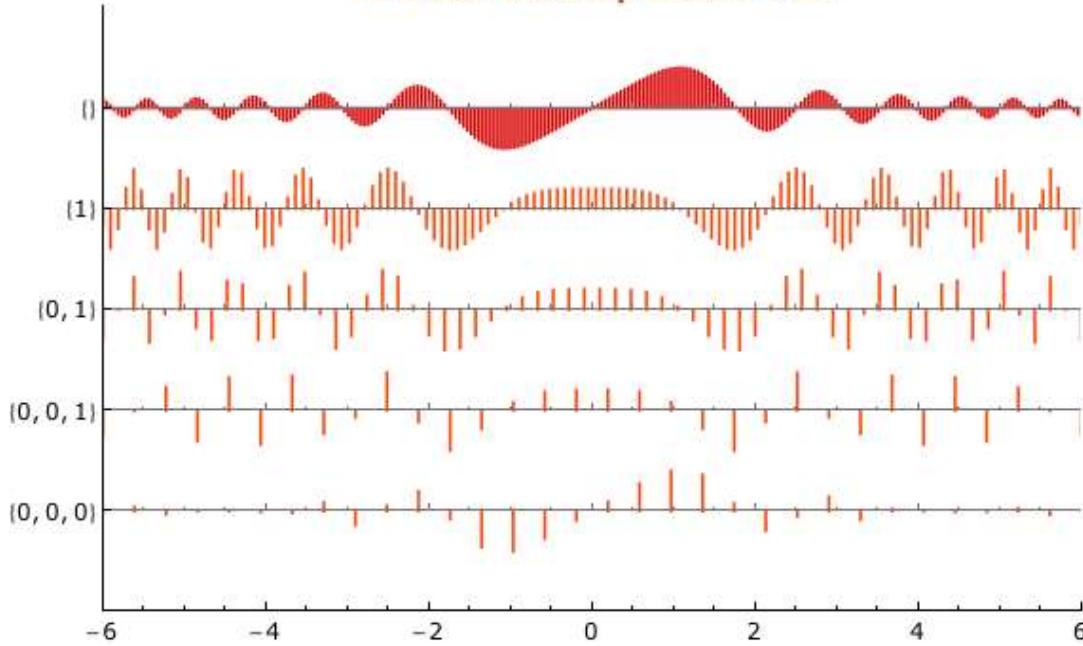
- the finest resolution is computed first.
- The computation then proceeds to courser resolutions,
- but rather than start over on the original waveform, the computation uses a smoothed version of the fine resolution waveform.

This smoothed version is obtained using a scaling function

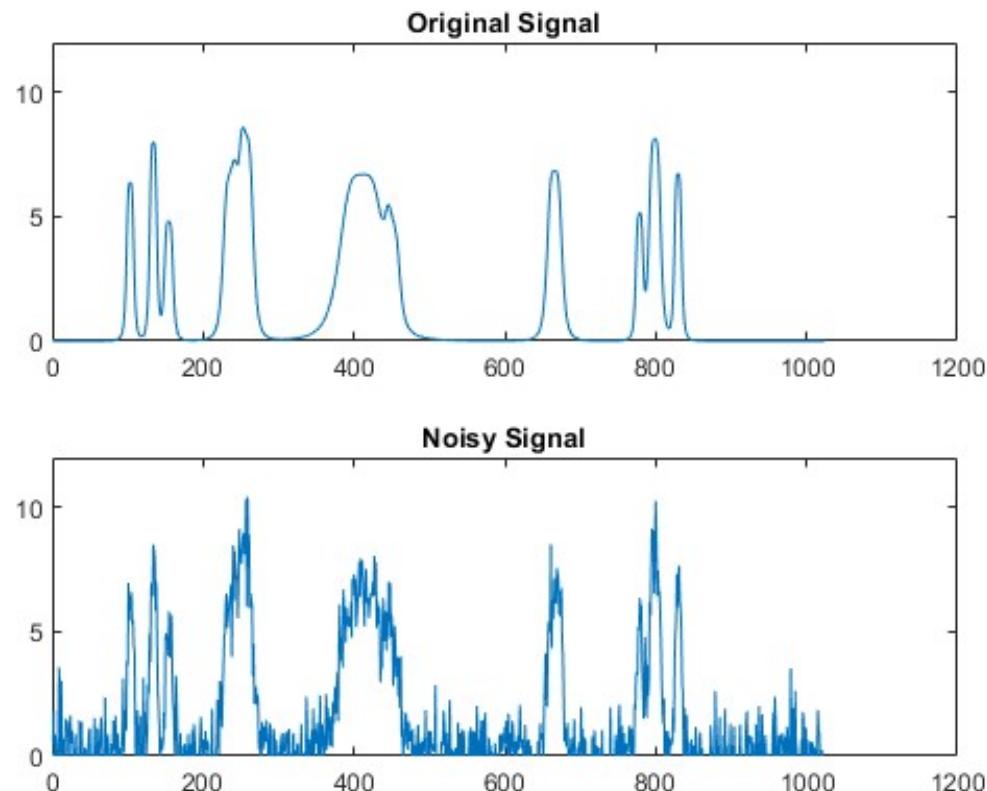
LWT Decomposition Tree



Wavelet Decomposition Plot



Wavelet Denoising



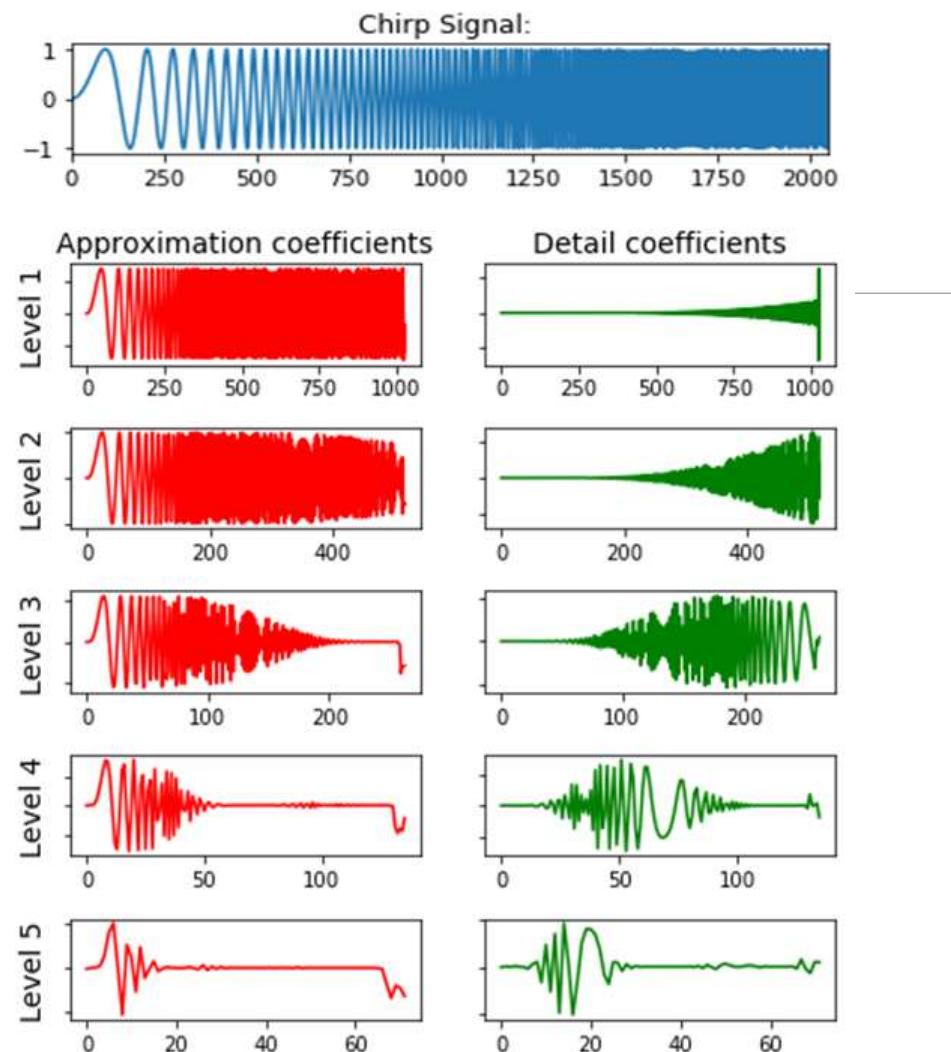
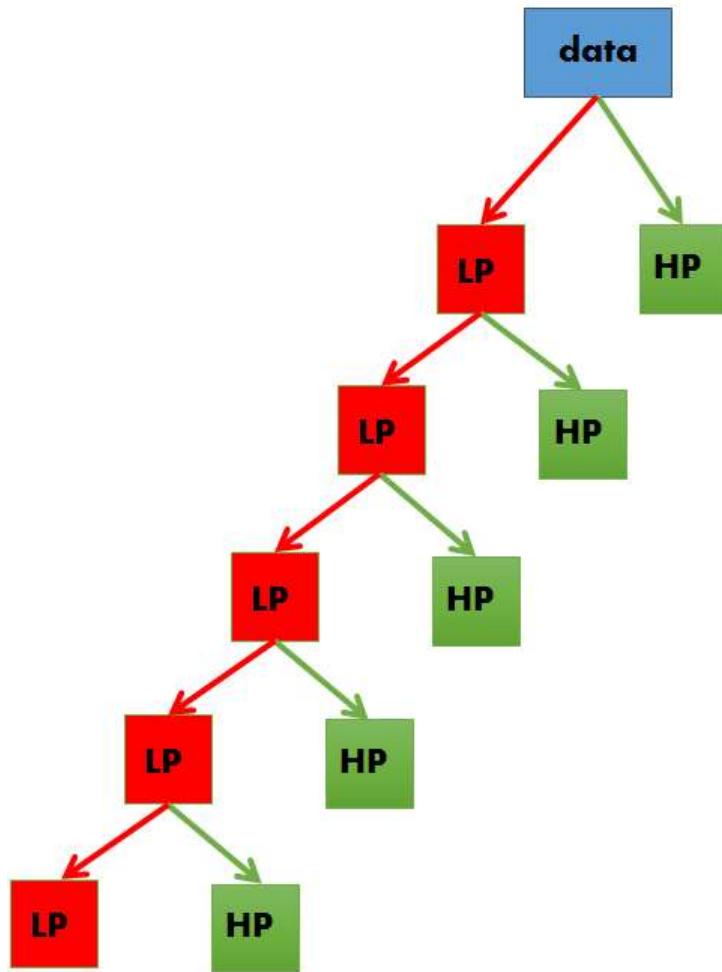
Discrete Wavelet Transform

The wavelet itself can be defined from the scaling function:

$$\Psi(t) = \sum_{n=-\infty}^{\infty} \sqrt{2} d(n) \phi(2t - n)$$

where $d(n)$ is a series of scalars that are related to the waveform $x(t)$ and that define the discrete wavelet in terms of the scaling function.

While the DWT can be implemented using the above equations, it is usually implemented using filter bank techniques.



Building the DWT:

$$\Psi_{jk}(t) = \frac{1}{\sqrt{a_{jk}}} \Psi\left(\frac{t-b_{jk}}{a_{jk}}\right) = a_0^{-j/2} \Psi(a_0^{-j}t - kT)$$

$$a_{jk} = a_0^j$$

$$b_{jk} = k a_0^j T$$

- T is the sampling time
- a_0 is a positive nonzero constant
- $\Psi(t)$ is continuous mother wavelet (a family of functions)
- $0 \leq j \leq (N-1)$, and $0 \leq k \leq (M-1)$

Building the DWT:

With the previous, the coefficients of the DWT can be calculated:

$$W_{jk} = \int_{-\infty}^{+\infty} x(t) \Psi_{jk}^*(t) dt$$

Which calculates a finite set of discrete coefficients directly from the signal.

$$x(t) = c \sum_{j=0}^{N-1} \sum_{k=0}^{M-1} W_{jk} \Psi_{jk}(t)$$

Building the DWT:

- c is a constant that depends on choice of mother wavelet.

- this equation can reconstruct the continuous signal directly from a set of discrete coefficients

BUT, now..... how to choose the number of basis functions for a given signal (i.e. how many shifted and scaled versions of the mother wavelet are needed to decompose a signal)??

$$x(t) = c \sum_{j=0}^{N-1} \sum_{k=0}^{M-1} W_{jk} \Psi_{jk}(t)$$

Discrete wavelet transform

CWT is highly redundant since a 1-dimensional function $x(t)$ is transformed into a 2-dimensional function. Therefore, it is Ok to discretize them to some suitably chosen sample grid.

The most popular is dyadic sampling:

$$s=2^{-j}, \tau = k2^{-j}$$

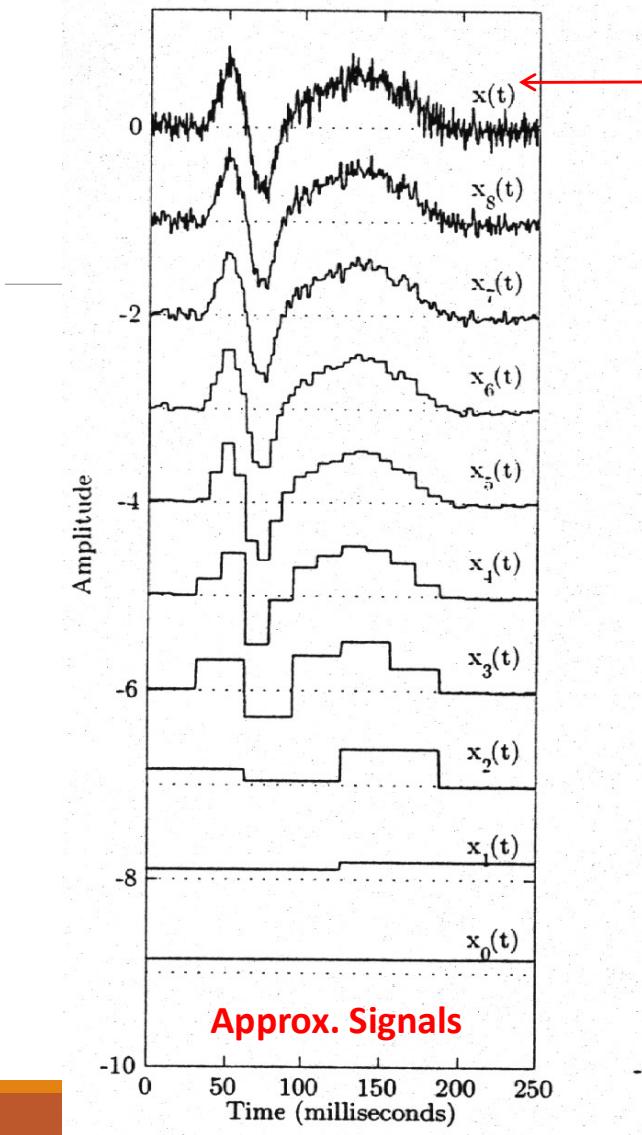
With this sampling it is still possible to reconstruct exactly the signal $x(t)$.

Multiresolution analysis

The signal can be viewed as the sum of:

1. a smooth (“coarse”) part – reflects main features of the signal (**approximation signal**);
2. a detailed (“fine”) part – faster fluctuations represent the details of the signal.

The separation of the signal into 2 parts is determined by the resolution.



Original evoked potential signal

Example of multi-resolution analysis
- using Haar

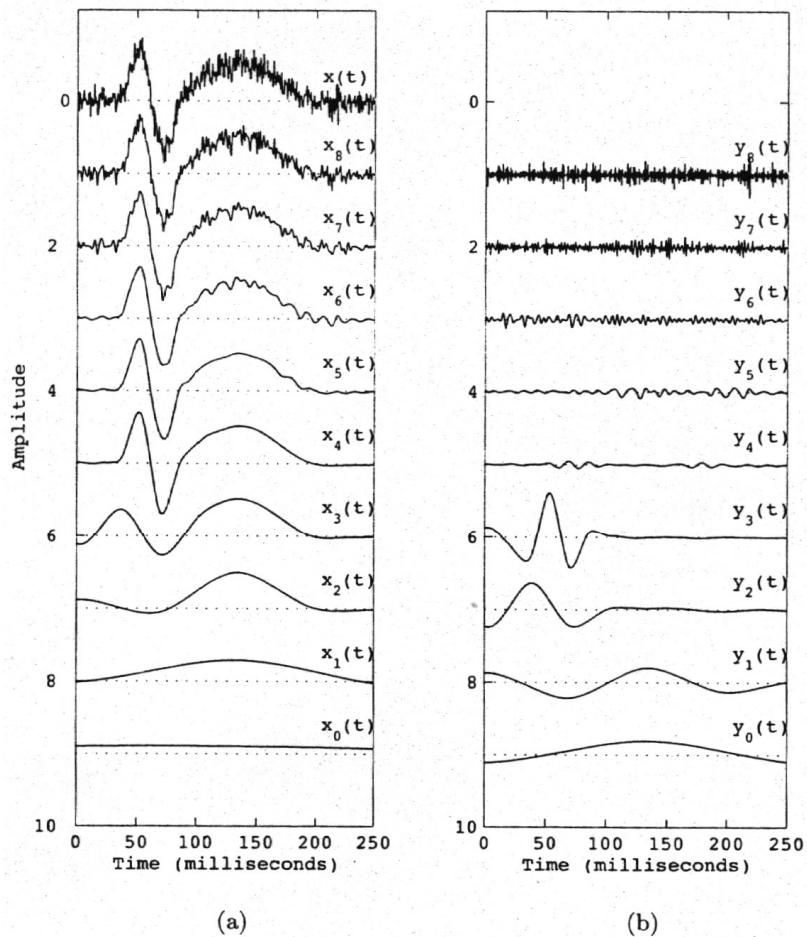


Figure 4.45: Multiresolution analysis of an evoked potential waveform using the Coiflet-4. (a) The approximation signals, and (b) the detail signals at different scales; the original signal is shown at the top left of the figure.

One more example but now with a smooth function
 Coiflet-4, you see, this one models the response somewhat better than Haar

What should you want from the scaling and wavelet function?

1. Orthonormality and compact support (concentrated in time, to give time resolution)
2. Smooth, if modeling or analyzing physiological responses (e.g., by requiring vanishing moments at certain scale): *Daubechies, Coiflets*.
3. Symmetric (hard to get, only Haar or sinc)

- Haar wavelet (square wave limited in time, superior time localization)
- Mexican hat (smooth)

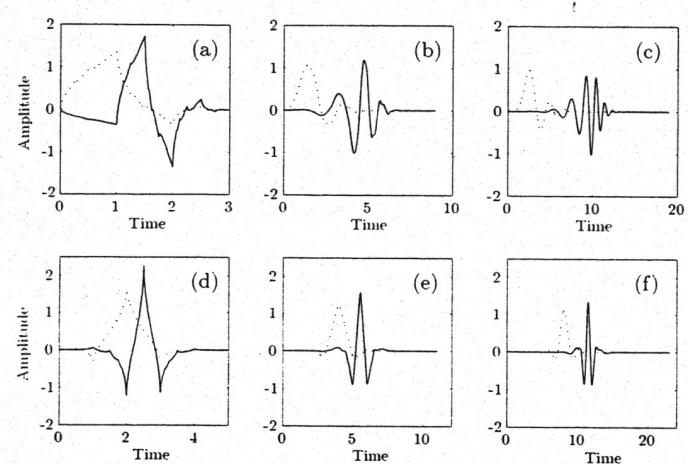
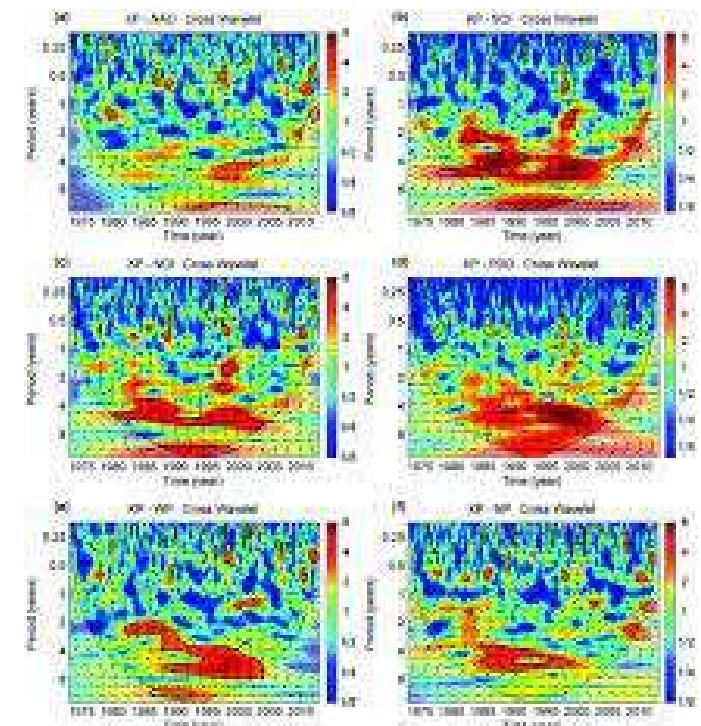
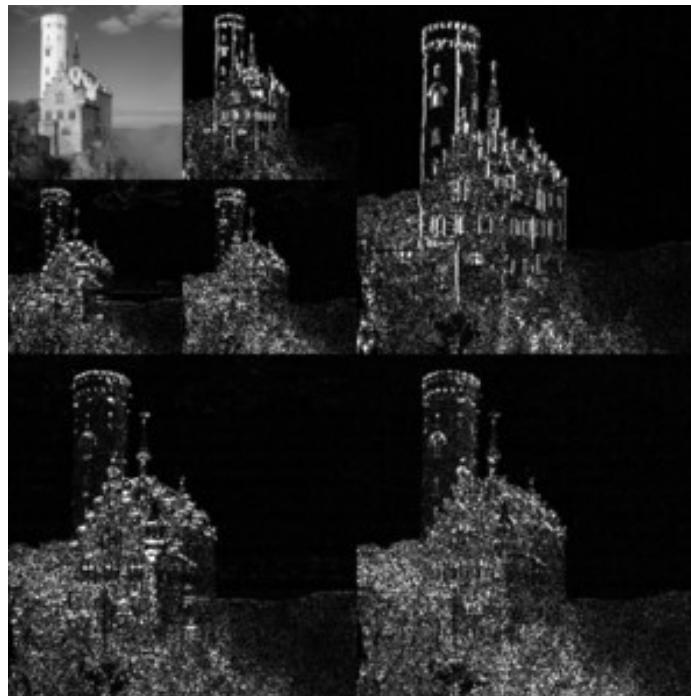


Figure 4.44: The scaling (dotted line) and wavelet function (solid line) for (a) Daubechies-2 (b) Daubechies-5, (c) Daubechies-10, (d) Coiflet-1, (e) Coiflet-2, and (f) Coiflet-4. Note that the time scales differ.

Applications of Wavelets

- Compression
- De-noising
- Feature Extraction
- Discontinuity Detection
- Distribution Estimation
- Data analysis
 - Biological data
 - NDE data
 - Financial data



Applications

Data Compression

Wavelet Shrinkage Denoising

Source and Channel Coding

Biomedical Engineering

- EEG, ECG, EMG, etc analysis
- MRI

Nondestructive Evaluation

- Ultrasonic data analysis for nuclear power plant pipe inspections
- Eddy current analysis for gas pipeline inspections

Numerical Solution of PDEs

Study of Distant Universes

- Galaxies form hierarchical structures at different scales

Applications

Wavelet Networks

- Real time learning of unknown functions
- Learning from sparse data

Turbulence Analysis

- Analysis of turbulent flow of low viscosity fluids flowing at high speeds

Topographic Data Analysis

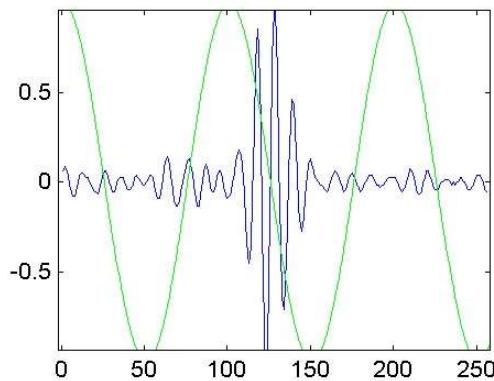
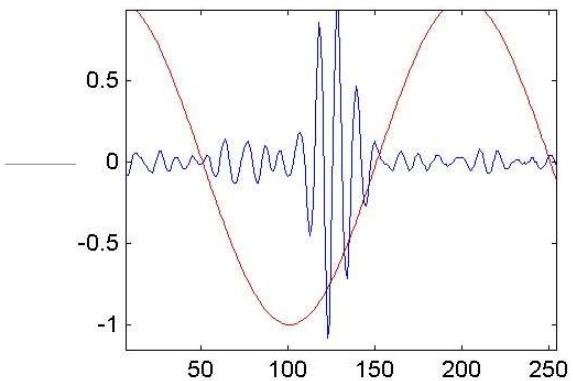
- Analysis of geo-topographic data for reconnaissance / object identification

Fractals

- Daubechies wavelets: Perfect fit for analyzing fractals

Financial Analysis

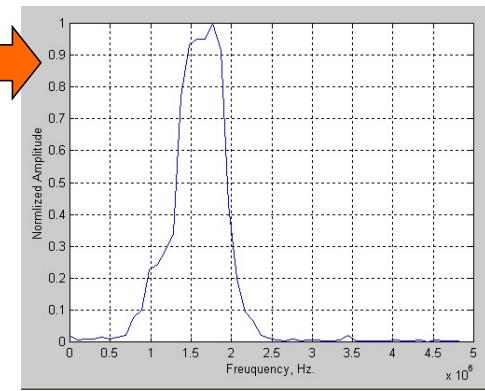
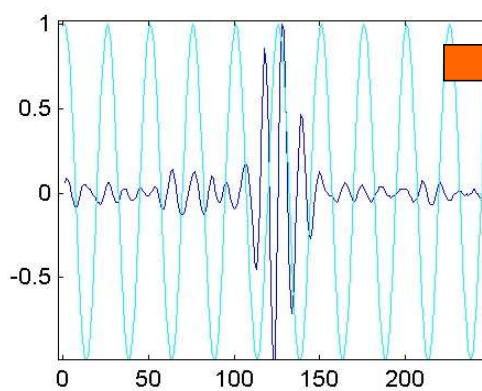
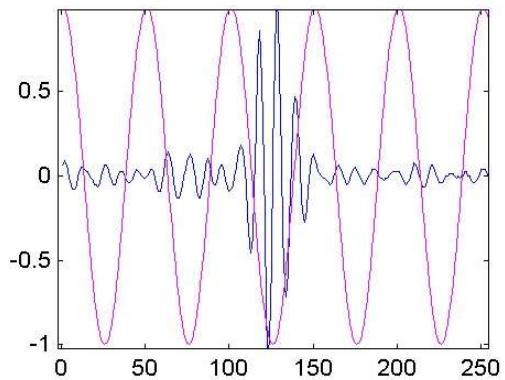
- Time series analysis for stock market predictions



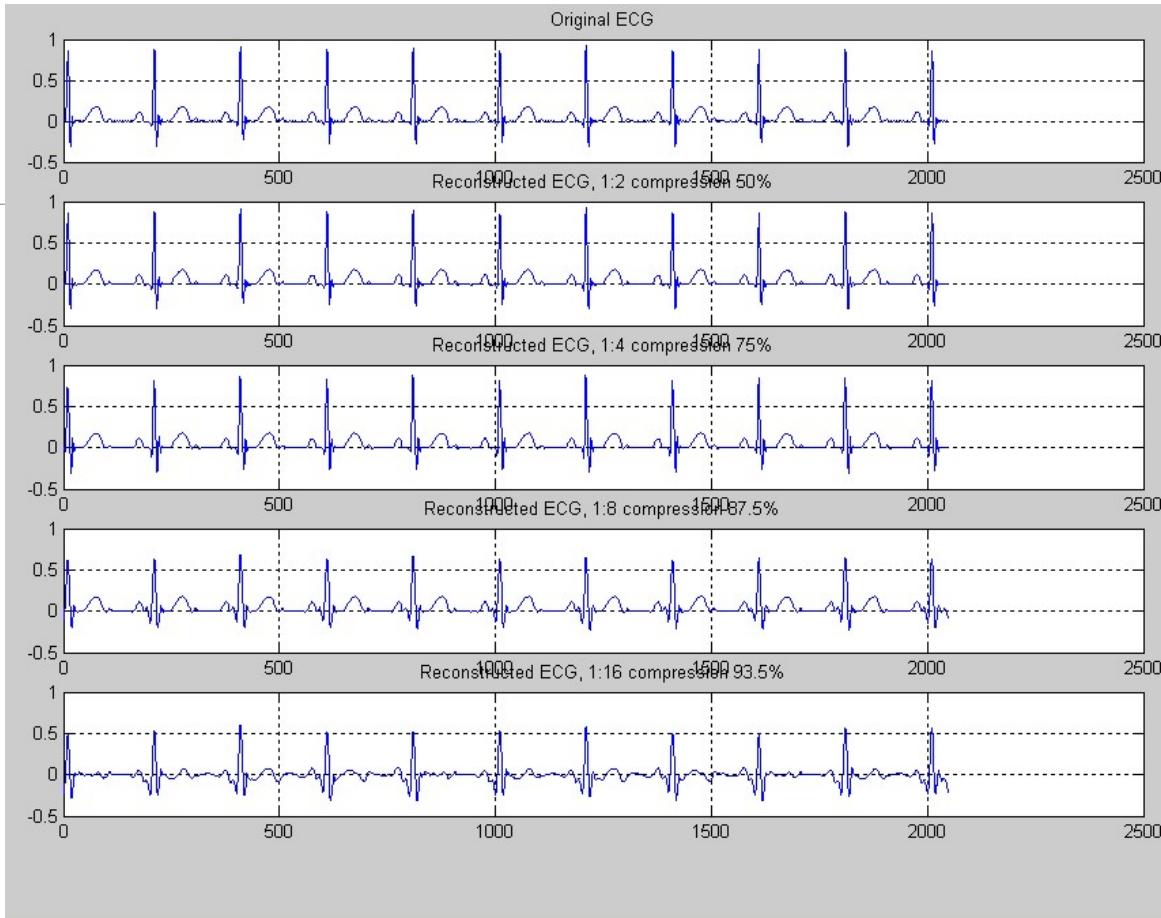
Complex exponentials
(sinusoids) as basis
functions:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \cdot e^{j\omega t} dt$$

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) \cdot e^{j\omega t} dt$$



An ultrasonic A-scan using 1.5 MHz transducer, sampled at 10 MHz



Lecture 9

TAYLOR DEVET MASC.

PHD. CANDIDATE BIOLOGICAL AND BIOMEDICAL ENGINEERING

MCGILL UNIVERSITY

SHRINERS HOSPITAL FOR CHILDREN

Todays Aims...



Scaling Allometry

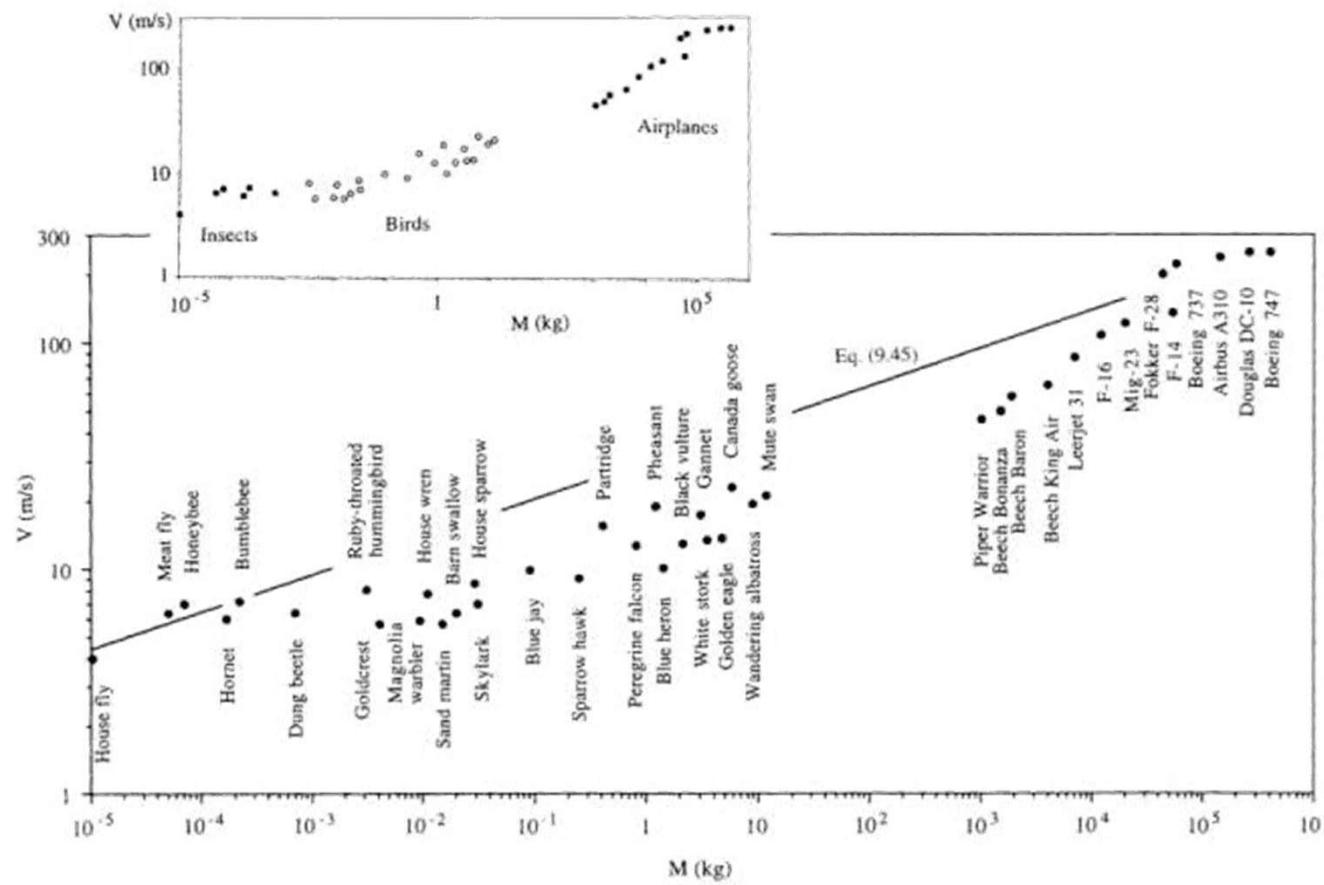


Complexity of Biology



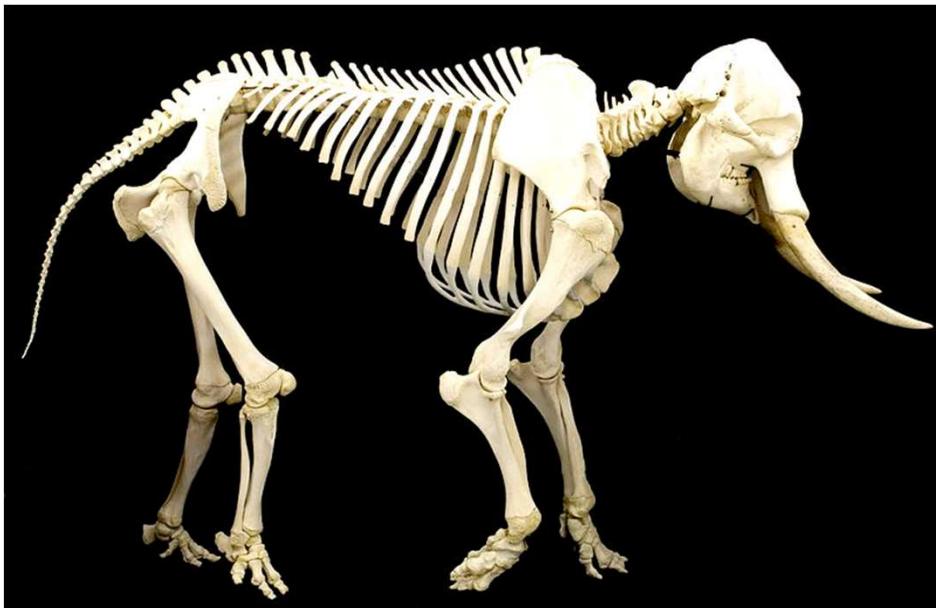
Chaos and Fractals

Power Laws in Biology: Science of Allometry



<http://en.wikipedia.org/wiki/Allometry>

Bone thickness related to size

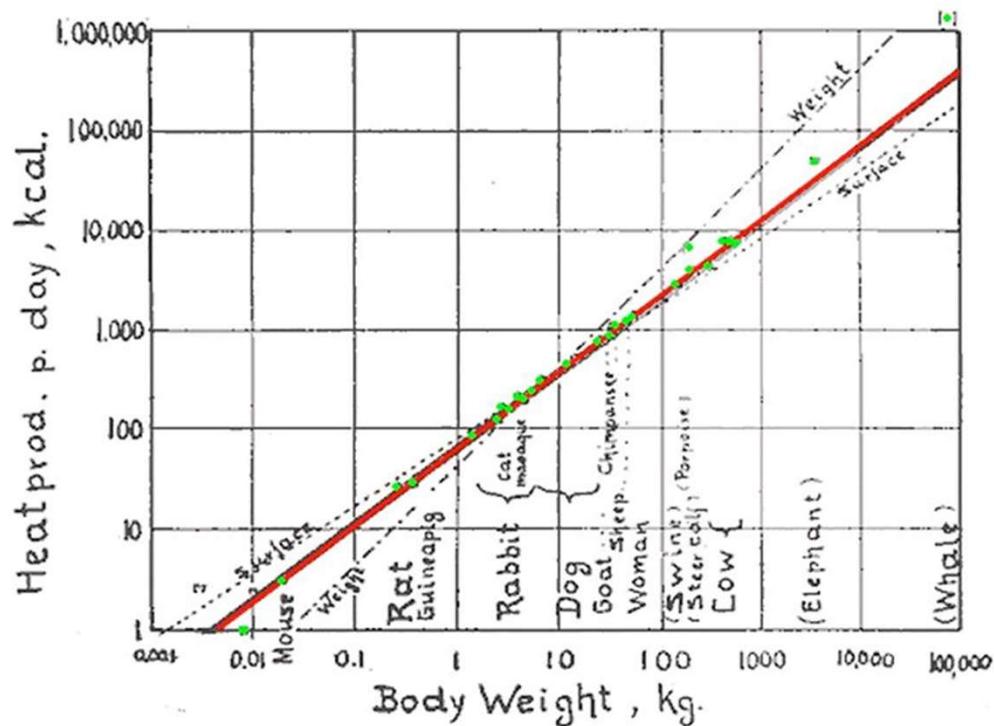




Rhinoceros Beetle (*Dynastes hercules*)



Kleiber's law



If q_0 = metabolic rate, and M =mass, then Kleiber's law states that:

$$q_0 = 70 M^{3/4}$$

Thus a cat (mass=100x mouse), will have a metabolism $\sim 31x$ greater

Power Laws

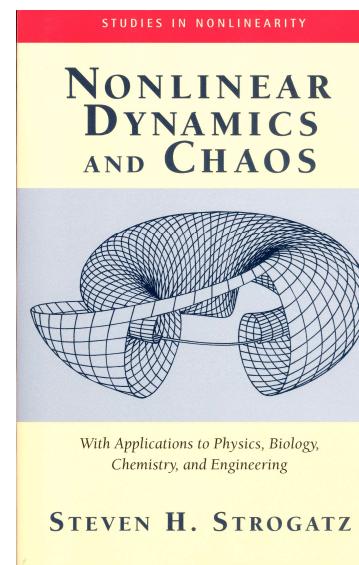
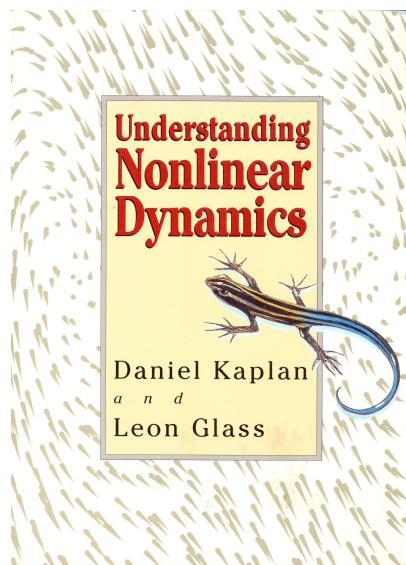
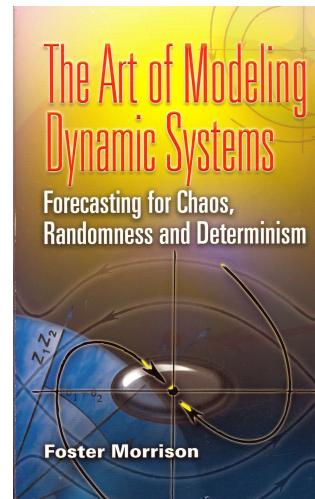
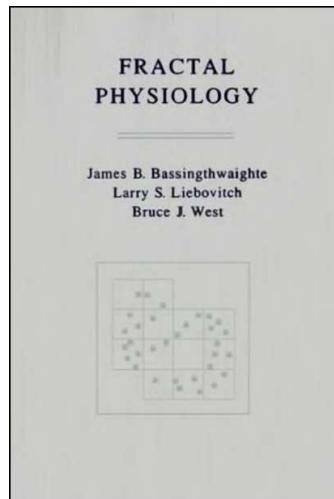
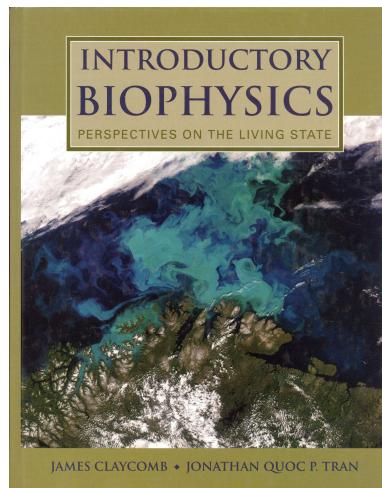
Many lower laws in biology follow the same form

$$y=kx^a$$

$$\log(y)=a \cdot \log(x) + \log(k)$$

Observable, y	Power Law exponent, a
Genome Length	$\frac{1}{4}$
Concentration of RNA	$-\frac{1}{4}$
Total Mitochondrial Mass Relative to Body Mass	$-\frac{1}{4}$
Basal Metabolic Rate	$\frac{3}{4}$
Heart Rate	$-\frac{1}{4}$
Life Span	$\frac{1}{4}$
Radii of Aorta and Tree Trunks	$\frac{3}{8}$

Unfortunately most things in biology don't follow nice linear models...



Why nonlinear processing?

The real world is nonlinear

- Often we can make linear assumptions, but we lose information that way.

Nonlinear analysis can provide greater information regarding dynamics of the measured system

- Dynamics refers to the governing rules of the system from which the signal originated

Why nonlinear processing?

- All other analyses have come up short
 - Not a very good reason
- We suspect our signal comes from a nonlinear system and we want to know more about it
 - A better reason
- We know our system is nonlinear and we want to differentiate measurements from different system conditions
 - Even better reason

Nonlinear Vs. Linear

Nonlinear signals come from nonlinear systems

Nonlinear systems have governing equations with nonlinear terms

- Sine, exponentials, logarithms, etc...

The input-output relationships do not follow super-position

A hallmark of a nonlinear system is that small changes can be amplified, or vice versa

Simple Nonlinear System

- $y = x^2$
 - Here y and x are variables representing the output and input of a signal
- Inputs of a and b to this system result in a^2 and b^2
- But an input of $(a + b)$ results in an output of

A More Complicated Nonlinear System

A damped driven pendulum can be modeled as:

Where θ and ω are the angular position and velocity respectively

B is the damping constant and k is the driving force

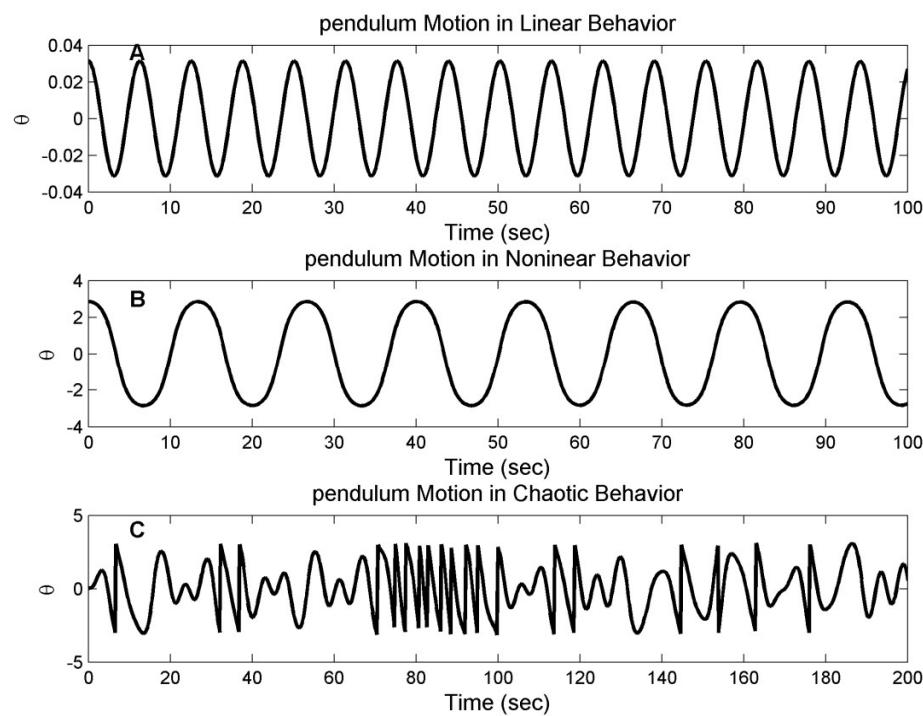
G is gravity

L is the string length

This system has altered modes of behaviour depending on what values b and k take, and the initial conditions

This is NOT linear due to sine terms

Pendulum in 3 modes of behavior



Initial conditions and parameters alter the pendulum behavior

Initial θ (rads)	Force Parameter k	Behavior
$\pi/100$	0	Linear
$\pi/1.1$	0	Nonlinear (but still regular)
$\pi/100$	0.75	Chaotic



MOVIECLIPS.COM

<http://www.youtube.com/watch?v=HH2jPq9g6CI>

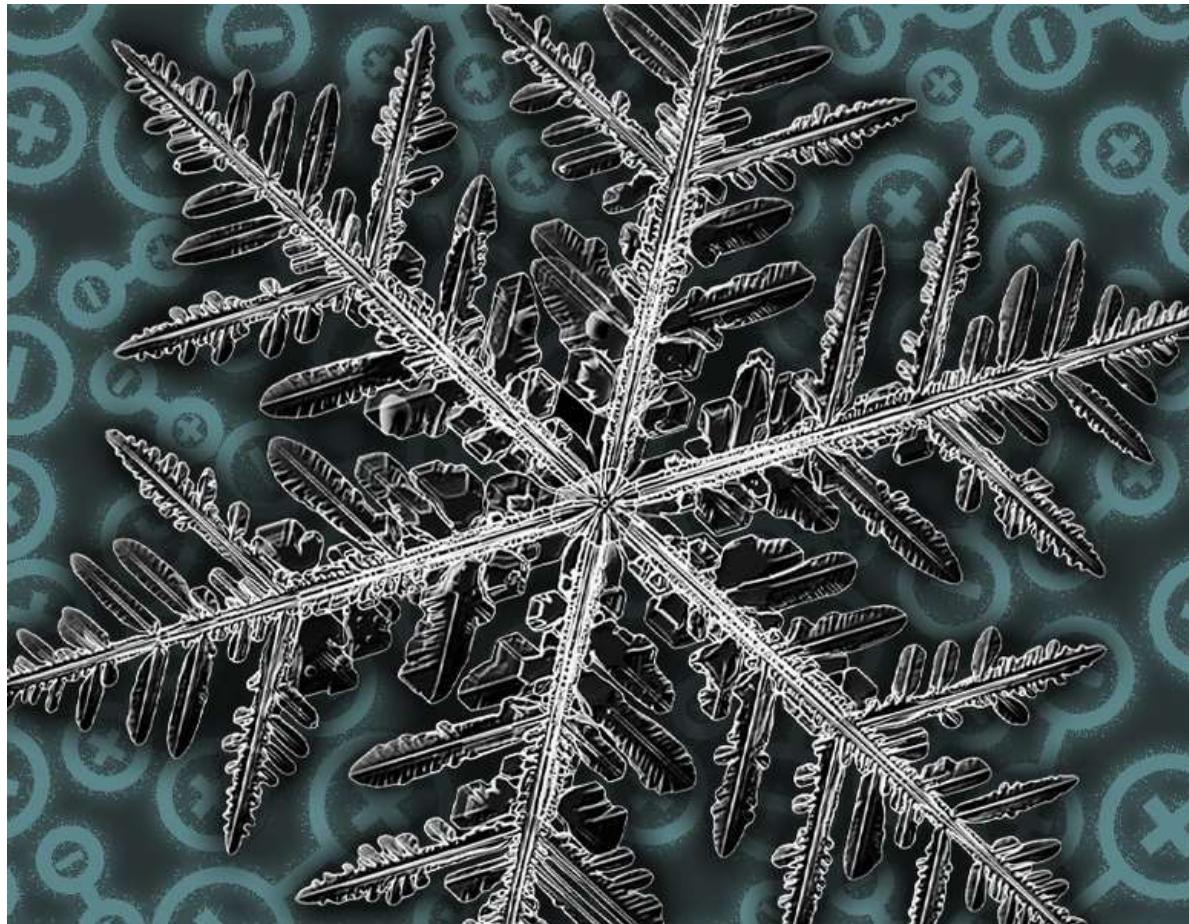


Complex Systems: Biology

Number of variables					
	$n = 1$	$n = 2$	$n \geq 3$	$n \gg 1$	Continuum
Linear	<i>Growth, decay, or equilibrium</i>	<i>Oscillations</i>		<i>Collective phenomena</i>	<i>Waves and patterns</i>
	Exponential growth	Linear oscillator	Civil engineering, structures	Coupled harmonic oscillators	Elasticity
	RC circuit	Mass and spring		Solid-state physics	Wave equations
	Radioactive decay	RLC circuit	Electrical engineering	Molecular dynamics	Electromagnetism (Maxwell)
Nonlinear		2-body problem (Kepler, Newton)		Equilibrium statistical mechanics	Quantum mechanics (Schrödinger, Heisenberg, Dirac)
					Heat and diffusion
					Acoustics
					Viscous fluids

Complex Systems: Biology

Number of variables →					
	$n = 1$	$n = 2$	$n \geq 3$	$n \gg 1$	
Nonlinear	Fixed points	Pendulum	Strange attractors (Lorenz)	Coupled nonlinear oscillators	Spatio-temporal complexity
	Bifurcations	Anharmonic oscillators		Lasers, nonlinear optics	Nonlinear waves (shocks, solitons)
	Overdamped systems, relaxational dynamics	Limit cycles	3-body problem (Poincaré)	Nonequilibrium statistical mechanics	Plasmas
	Logistic equation for single species	Biological oscillators (neurons, heart cells)	Chemical kinetics		Earthquakes
		Predator-prey cycles	Iterated maps (Feigenbaum)	Nonlinear solid-state physics (semiconductors)	General relativity (Einstein)
		Nonlinear electronics (van der Pol, Josephson)	Fractals (Mandelbrot)	Josephson arrays	Quantum field theory
			Forced nonlinear oscillators (Levinson, Smale)	Heart cell synchronization	Reaction-diffusion, biological and chemical waves
				Neural networks	Fibrillation
				Immune system	Epilepsy
			Practical uses of chaos	Ecosystems	Turbulent fluids (Navier-Stokes)
			Quantum chaos ?	Economics	Life



Fractals

a never-ending pattern.

infinitely complex patterns

repeating a process over and over
in an ongoing feedback loop.

fractals are images of dynamic
systems

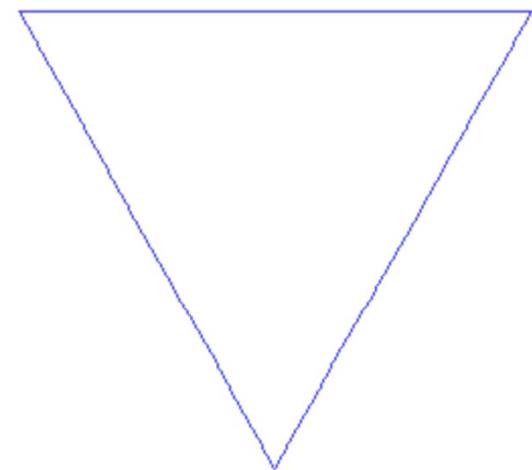
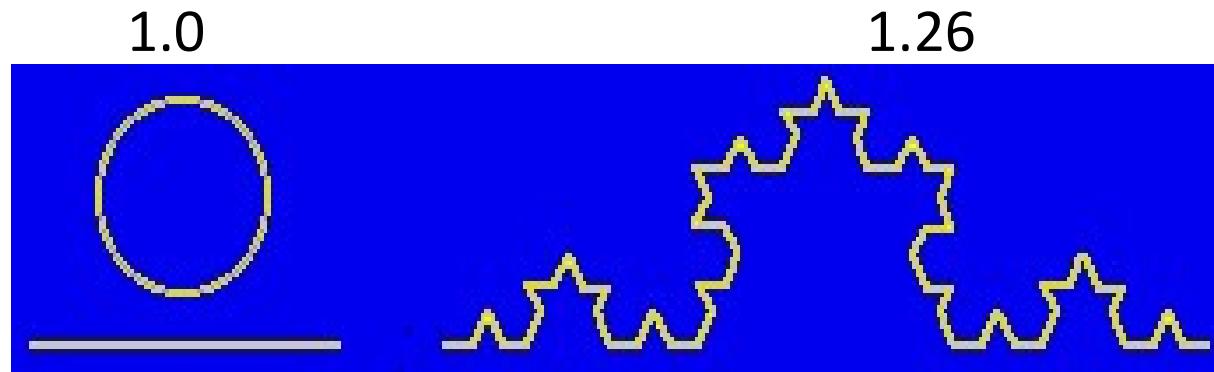
Enter the Fractal Dimension

Assigns a number that assesses the extent of “chaotic” or oscillatory behavior.

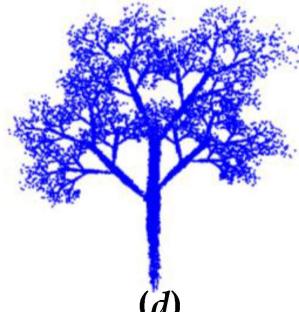
Measure of complexity

Measure of self similarity

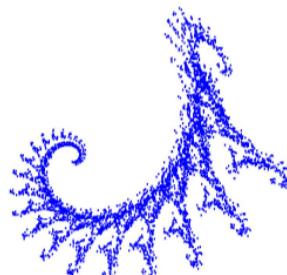
Self-similar across different scales.



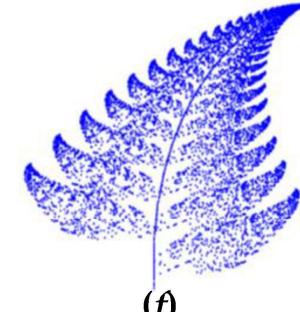
(a)



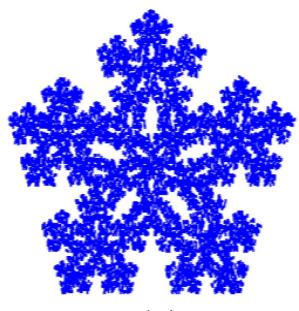
(b)



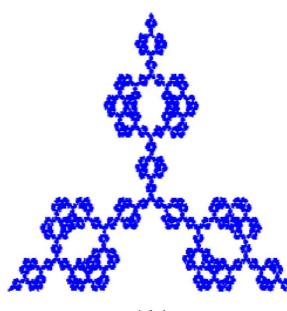
(c)



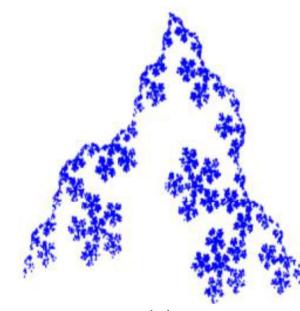
(d)



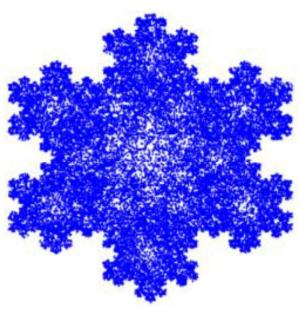
(e)



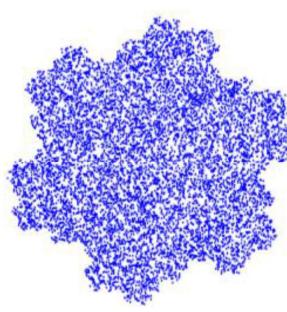
(f)



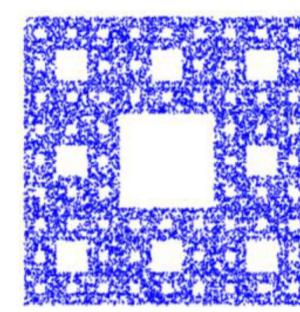
(g)



(h)



(i)



Fractally generated
world.
e.g. minecraft



Fractals in biology

- Fractal structures provide an advantageous architecture for living organisms seeking to cover a large area while conserving the amount of building material.
- Examples are branching networks in circulatory systems, lungs, nerves, and plant structures.
- Barnsley's fern is an example of a mathematical fractal that resembles natural plant structures. The fern is generated by the iterative transformation

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} a_j & b_j \\ c_j & d_j \end{pmatrix} \begin{pmatrix} x_n \\ y_n \end{pmatrix} + \begin{pmatrix} e_j \\ f_j \end{pmatrix}$$

where the coefficients

$$\begin{pmatrix} a_j & b_j \\ c_j & d_j \end{pmatrix} \text{ and } \begin{pmatrix} e_j \\ f_j \end{pmatrix}$$

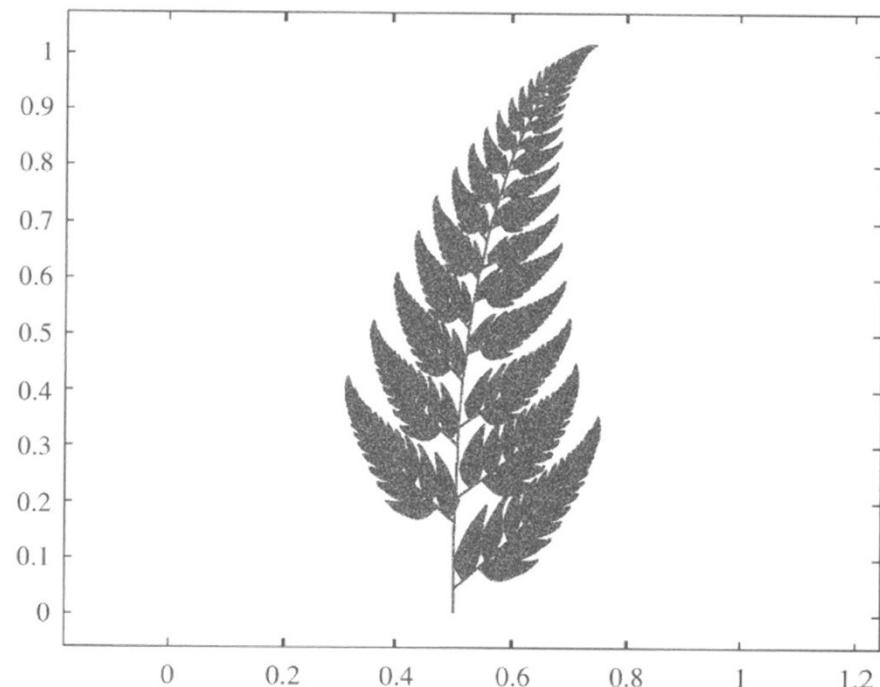
are chosen with probability P_j resulting in a rotation and translation of the vector (x_n, y_n) to (x_{n+1}, y_{n+1})

Fractal Ferns

This can be re-written in succinct notation as:

$$\mathbf{X}_{n+1} = A_j \mathbf{X}_n + B_j$$

where A_j are rotation matrices and B_j are translation vectors

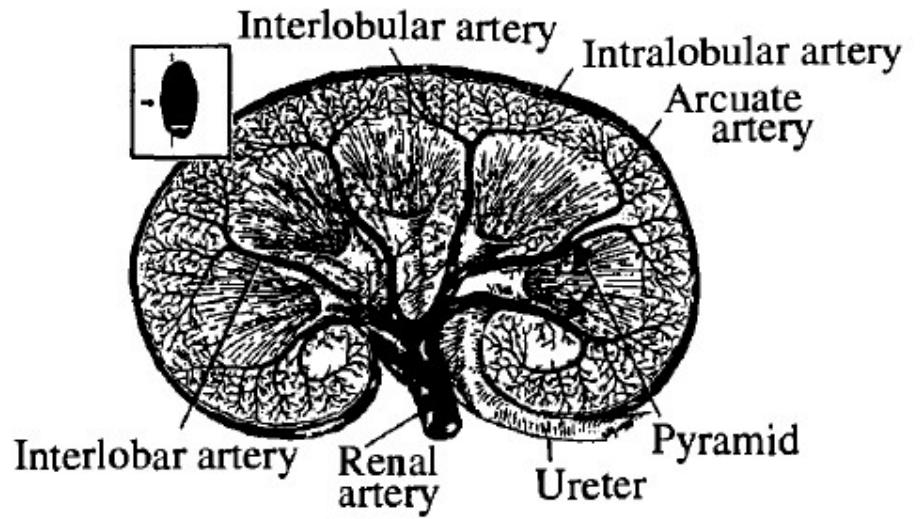
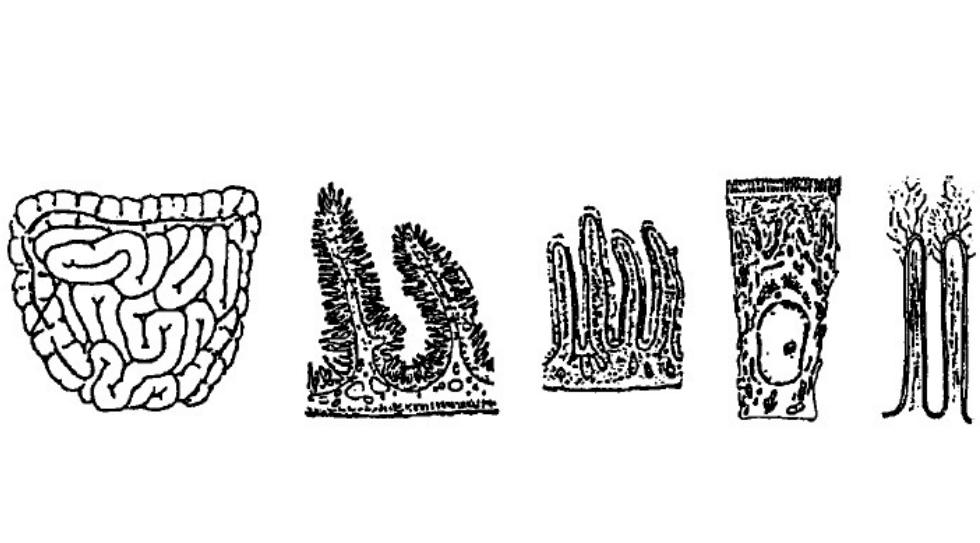


Fractal Ferns



The ultimate fractal vegetable (Wired Magazine)

- natural representation of the Fibonacci or golden spiral, a logarithmic spiral where every quarter turn is farther from the origin by a factor of ϕ , the golden ratio.



Fractals in the human body

Fractals can be Visualized
in Space and Time or
BOTH

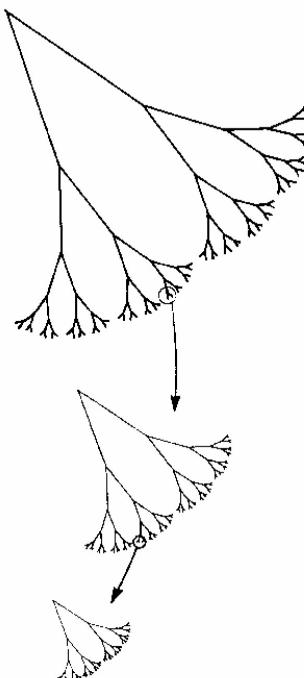
Fractal Dimension

Assigns a number between 1 - 1.5 that assesses the extent of “chaotic” or oscillatory behaviour.

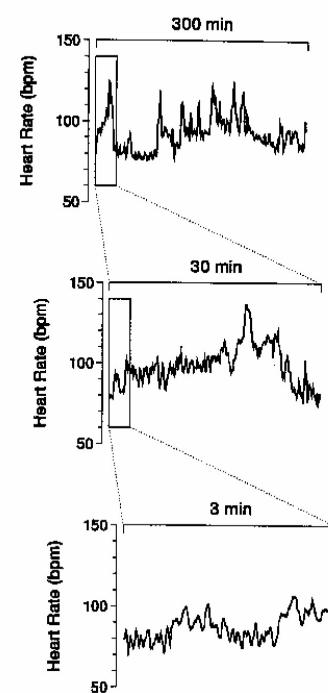
1.0 → Uniform, correlated and distinctly periodic signal.

1.5 → random noise, approaching ‘chaos’.

Self-Similar Structure



Self-Similar Dynamics



Fractal Dimension of Norway

The Fractal Dimension of
the Norwegian coastline= 1.52



Fractal Dimension (d)

- The fractal dimension can be thought of as a measure of geometric complexity. If we measure the length d of a one-dimensional curve by placing rulers end to end, then we will require a number of rulers N that is inversely proportional to the length of the rulers L since $d = N \times L$.

Fractal Dimension (d)

- If we halve the length of the rulers $L \rightarrow L/2$, then we must double their number $N \rightarrow 2N$ to cover the same d .
- A plot of N vs. $1/L$ will be a straight line with slope equal to one, the same as the dimensionality of the curve.

Fractal Dimension (d)

- For a square of side L , the number of boxes of length $\epsilon = L/2$ is given by $L^2/\epsilon^2 = 4$ so that $d = 2$.
- similarly use $d = 1$ for a line and $d = 3$ for a cube.

L	$L/2$	$L/3$

Box counting (Spatial domain)

Simple method to describe fractal dimension

Determine length of a one dimensional curve by placing boxes (or rulers) end to end

In traditional geometry a measuring stick 1/3 the original's size will give a total length 3 times as many "sticks" long

If one measures the area of a square then measures again with a box of side length 1/3 the size of the original, one will find 9 times as many squares as with the first measure. Holds in 2 dimensions with boxes.

Koch Snowflake



$N = 1 \quad \varepsilon = L$



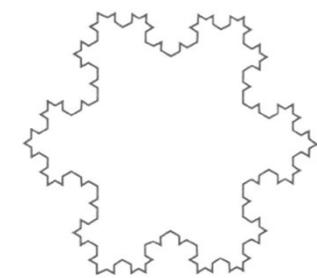
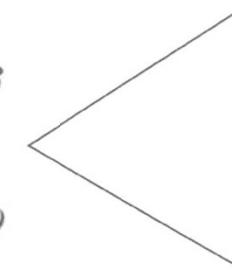
$N = 2 \quad \varepsilon = L/3$



$N = 4 \quad \varepsilon = L/9$



$N = 8 \quad \varepsilon = L/27$

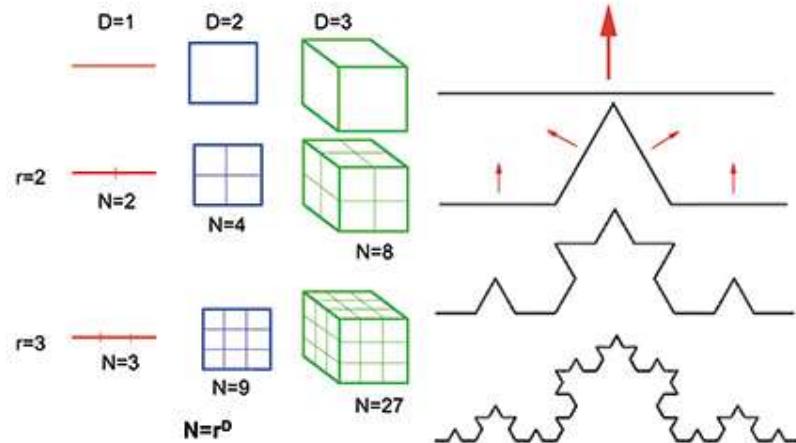


Koch Snowflake after
numerous iterations

$$d_c = \lim_{\varepsilon \rightarrow 0} \frac{\ln N}{\ln(1/\varepsilon)} = \lim_{n \rightarrow \infty} \frac{\log 2^n}{\log L + \log 3^n} = \frac{\log 2}{\log 3}$$

Fractal Dimension (d)

- Instead of rulers, imagine covering the curve with boxes.
- Then count the # of boxes required to span the curve as a function of box size.
- The capacity dimension (d) is calculated by covering the curve with d -dimensional boxes where the # of boxes of length ϵ is given by:



Fractal Dimension (d)

- An equation for the capacity dimension is obtained by taking the logarithm of both sides of :

Which can be rearranged to:

Fractal dimension

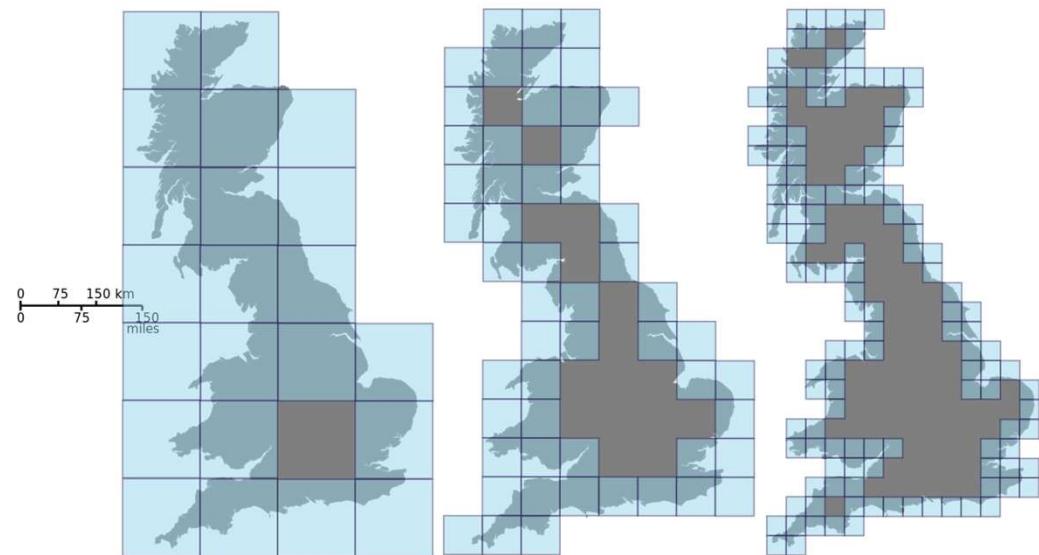
$$d = \frac{\log N(L)}{\log(1/L)}$$

Once again:

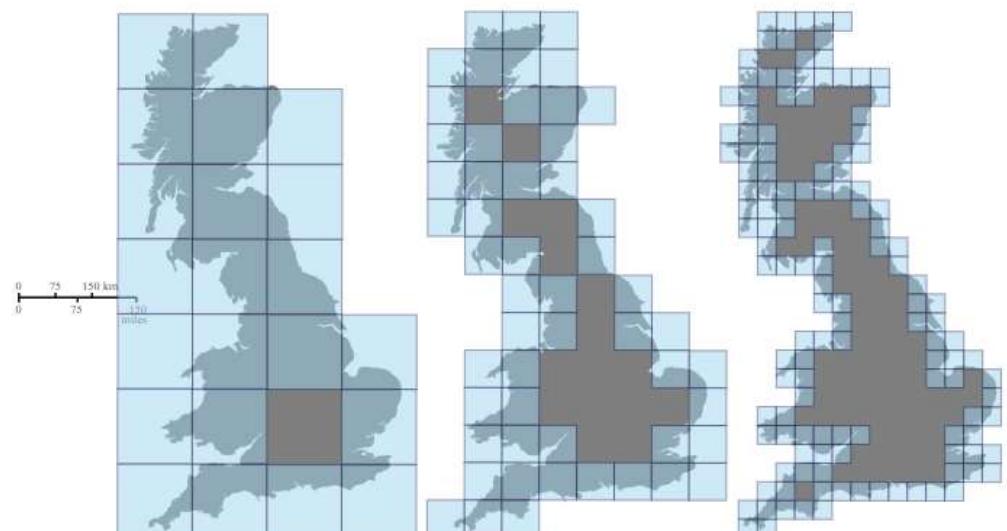
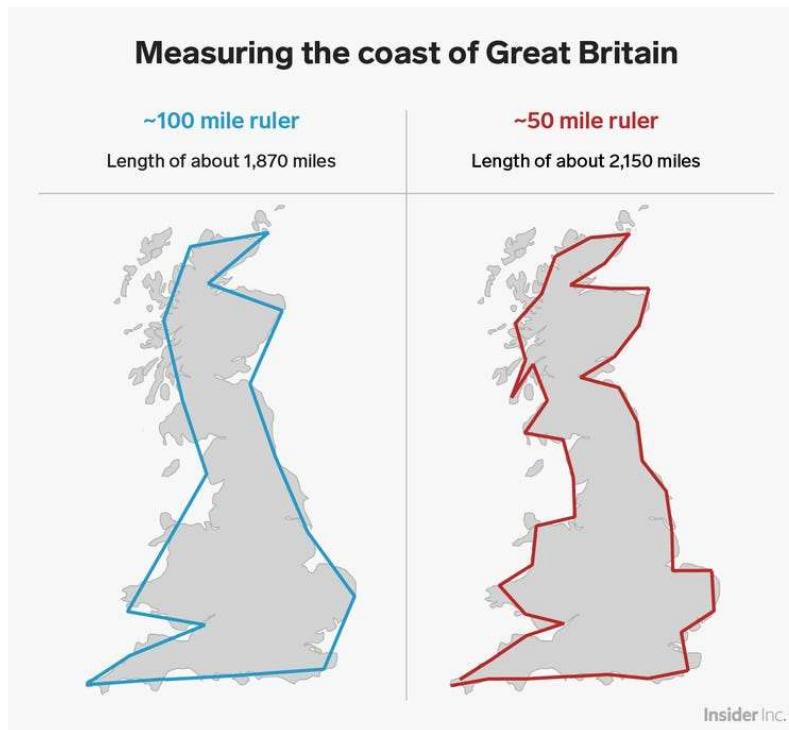
N is number of measuring devices

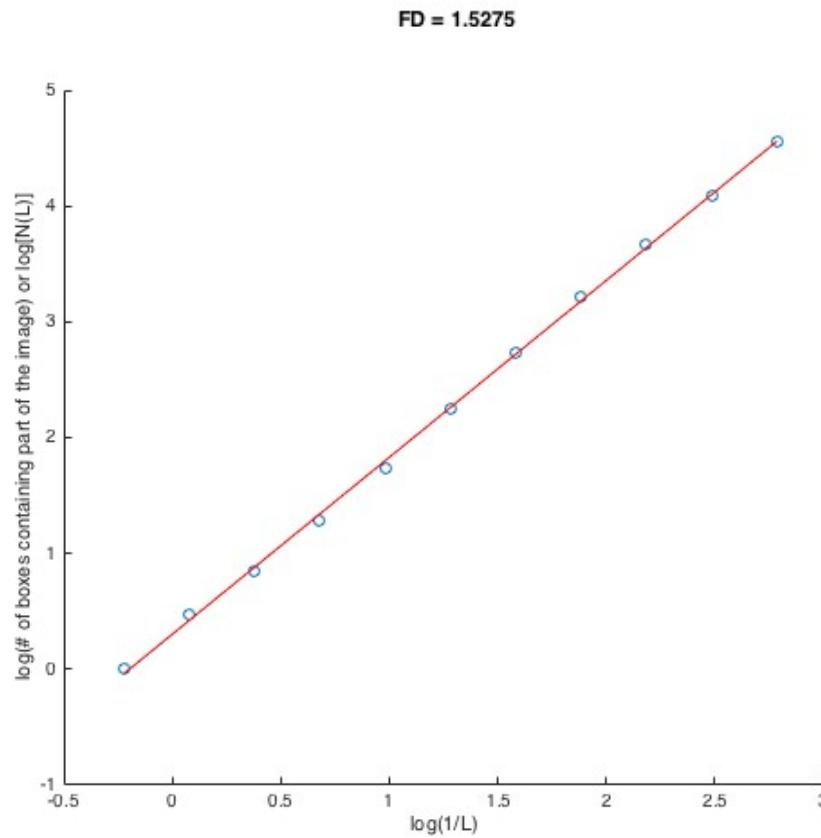
L is scaling factor

D is fractal dimension



Fractal Dimension of Great Britain





Boxcounting

Find the number of boxes you can use for an $m \times n$ image

- # of boxes should be equal to 2^i (might have to crop or zeropad image)
- # of scaling factors that can be used will be equal to i

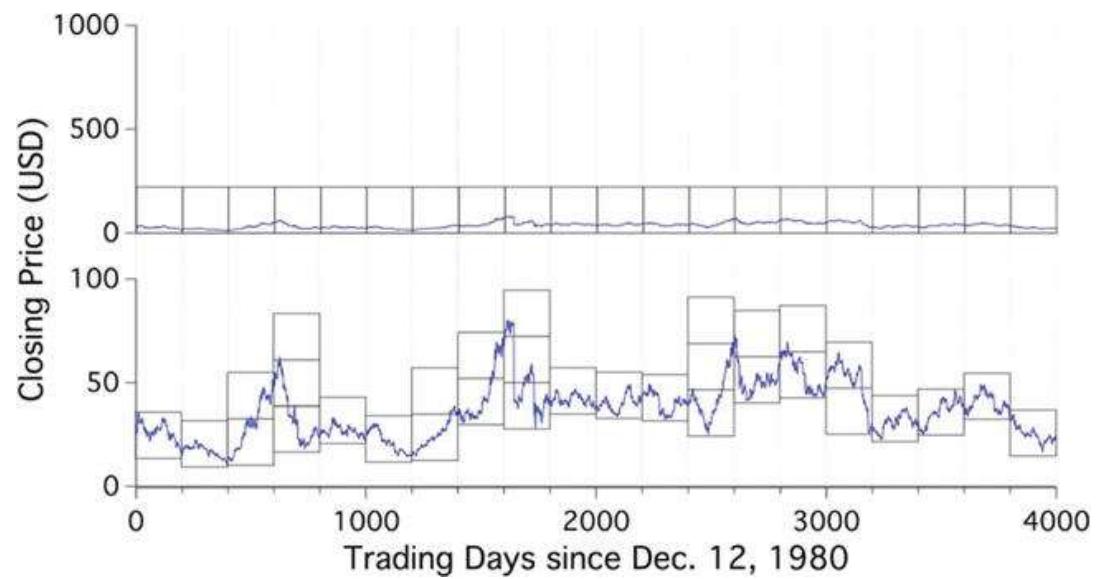
$\log(\# \text{ of boxes including edges of image})$ vs $\log(\# \text{ of boxes in grid} = 2^i)$

Find slope using polyfit function

Fractal Dimension (Time Domain)

Assigns a number between 1 -2 that assesses the extent of “chaotic” or oscillatory behaviour.

- 1.0 - Uniform, correlated and distinctly periodic signal.
- 1.5 - random noise, approaching ‘chaos’.



Relative Dispersion (time domain)

Can estimate F.D. using relative dispersion

Empirically determined relation in 1 dimension (ex. time series):

$$RD(m)/RD(m_0) = (m/m_0)^{1-D}$$

Where,

- RD is the relative dispersion (standard deviation divided by mean)
- m is the sample size
- m_0 is the reference sample size

Fractal Dimension: RD Calculation

- Look at temporal data over a “time-scale” m , calculate mean and standard deviation
- i.e. for 2048 points it would look like this:

$$m, \text{scale} \quad RD = SD/\text{mean}$$

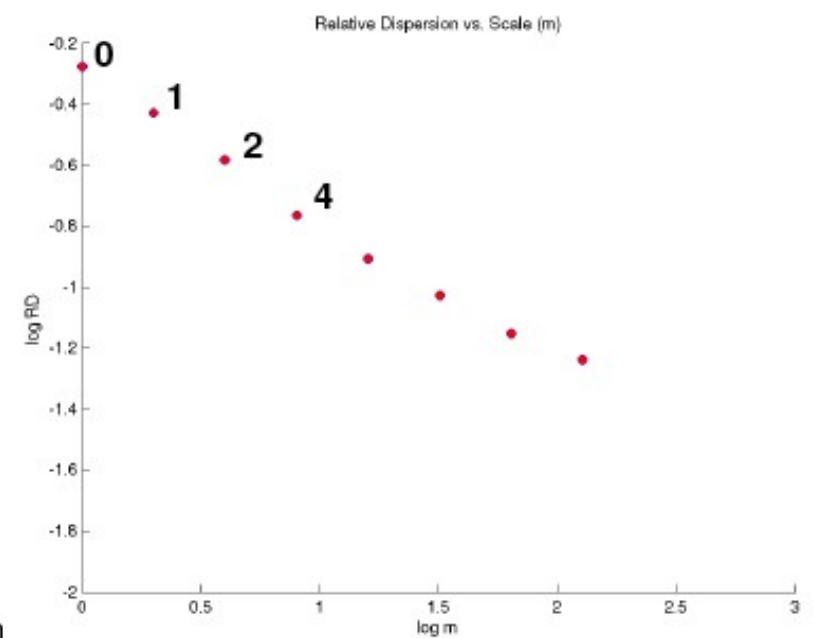
1	x_1
2	x_2
4	x_4

$$1024 \quad x_{1024}$$

slope log(RD) vs. log (m)

slope = H – 1: H, Hurst Exp.

slope = 1 – D: D, Fractal Dim



Relative dispersion

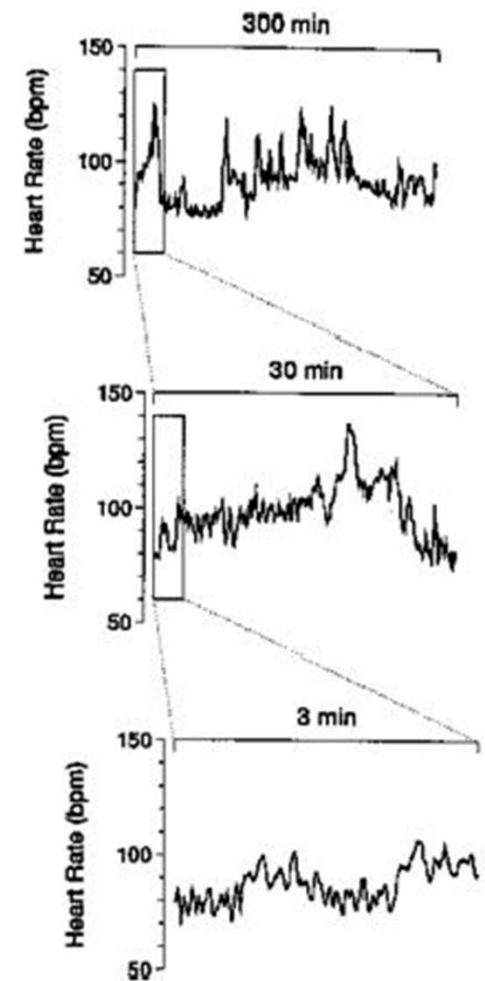
- 1) Calculate RD for bin size of m_0 samples ($RD = SD/\text{mean}$)
- 2) Rinse and repeat until $m_0 = \text{length of signal}$
- 3) Now with the array of RD values, plot on a $\log(m)$ vs $\log(RD)$ plot and determine the slope

$$D = 1 - \text{slope}$$

$$H = \text{slope} - 1 \quad (\text{Hurst Exponent})$$

[dim_rd, hurst_rd] = fractal_RD (sig , 1);

The values of the Hurst exponent vary between 0 and 1, with higher values indicating a smoother trend, less volatility, and less roughness.



Power Spectral Method (frequency domain)

Find the power spectrum of the time series, $|A|^2$

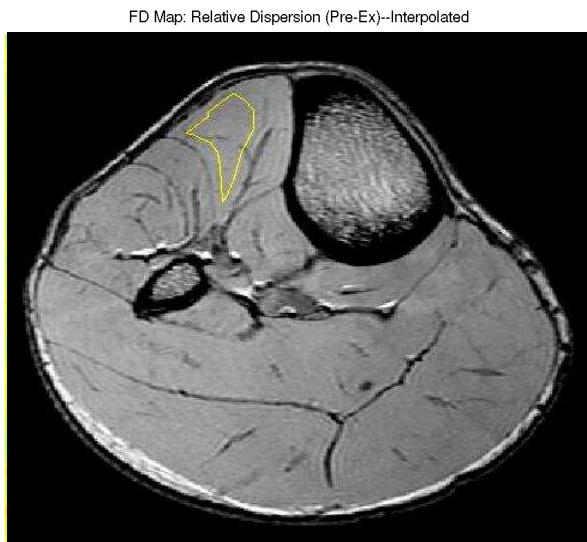
Plot $\log(f)$ vs $\log(|A|^2)$ and estimate the slope r of the best-fit line

$$\log |A|^2 = -r * \log f$$

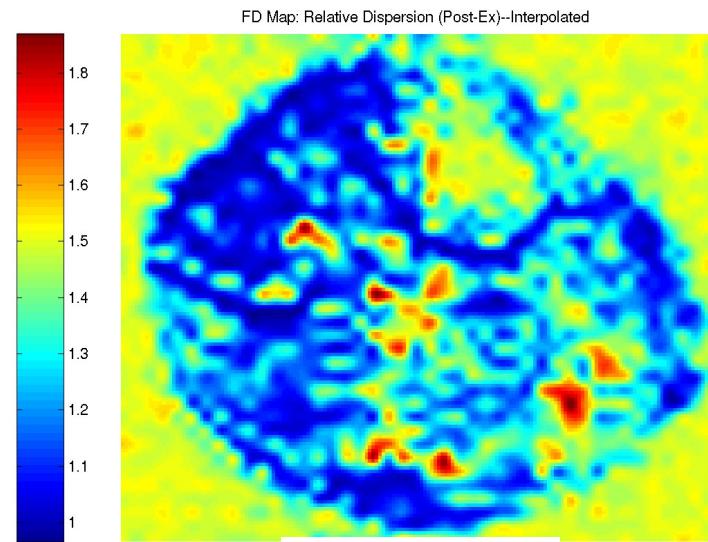
From the relation $D = 2 - (r+1)/2$

[dim_ps, hurst_ps] = fractal_PS (sig , 2)

Relative Dispersion



FD map (**PRE**)



FD Map (**POST**)

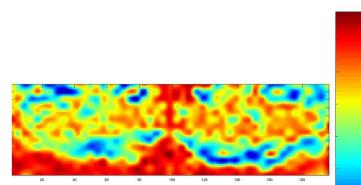
NOTE

Bin/average BOLD data over time

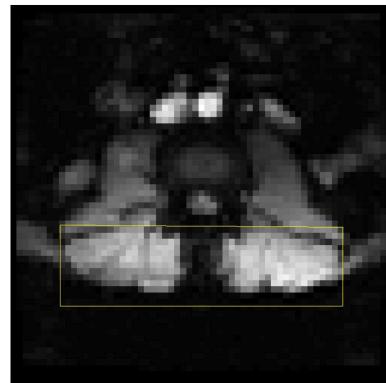
- noise characteristics stay the same pre vs. post exercise.

Visualizing the Effect of a [Therapeutic] Swedish Massage

PRE

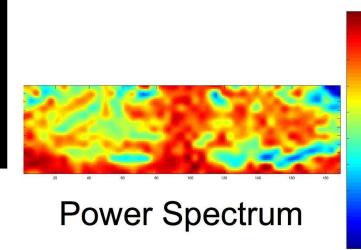


Power Spectrum

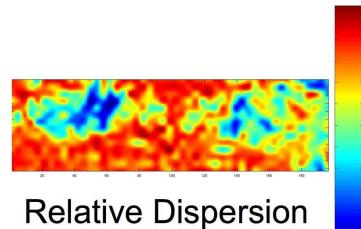


Axial Image
- box shows area
of analysis

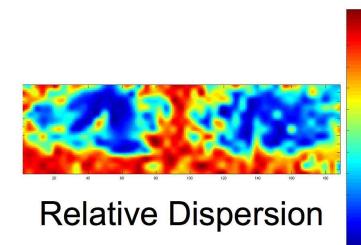
POST



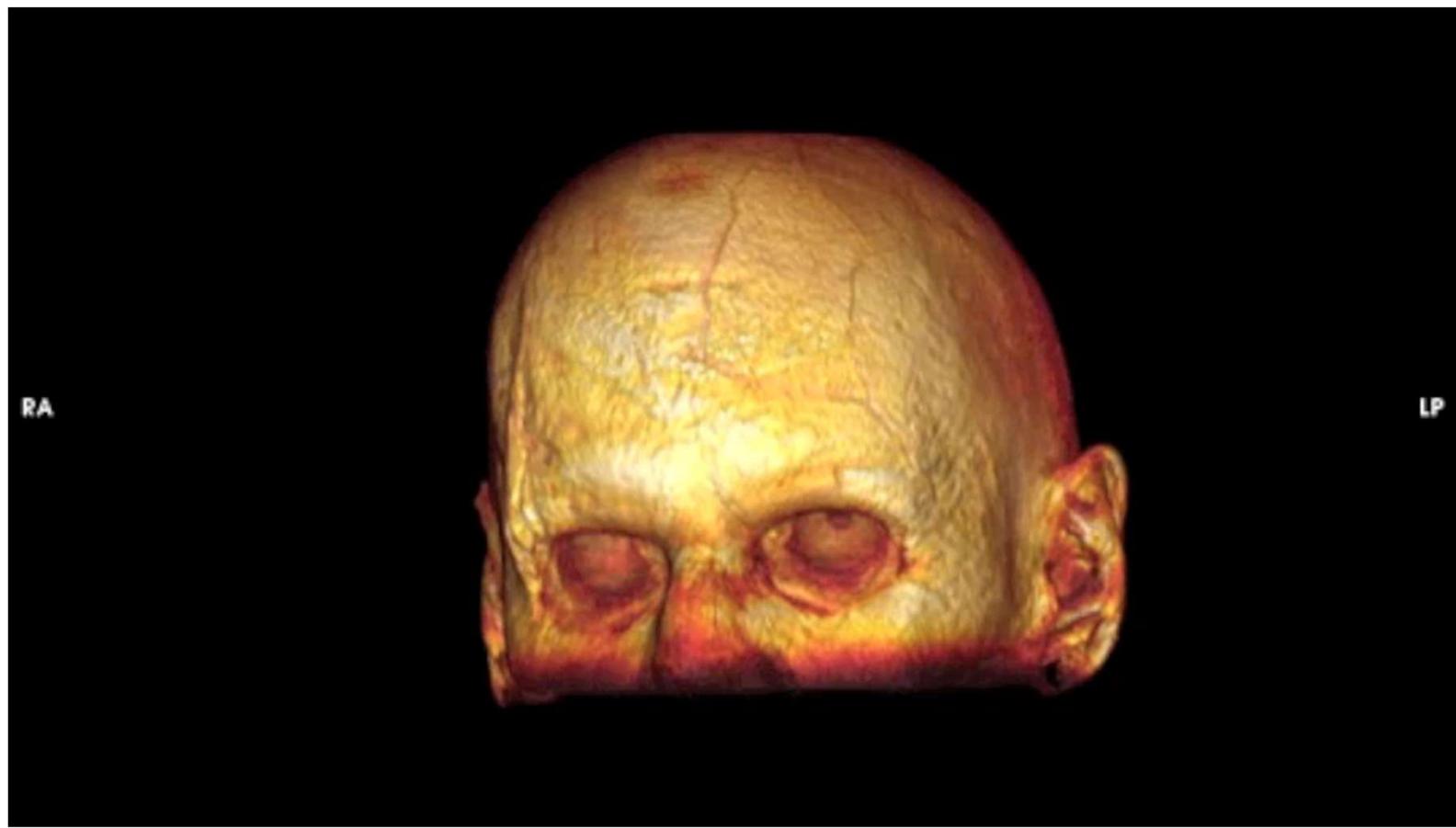
Power Spectrum



Relative Dispersion



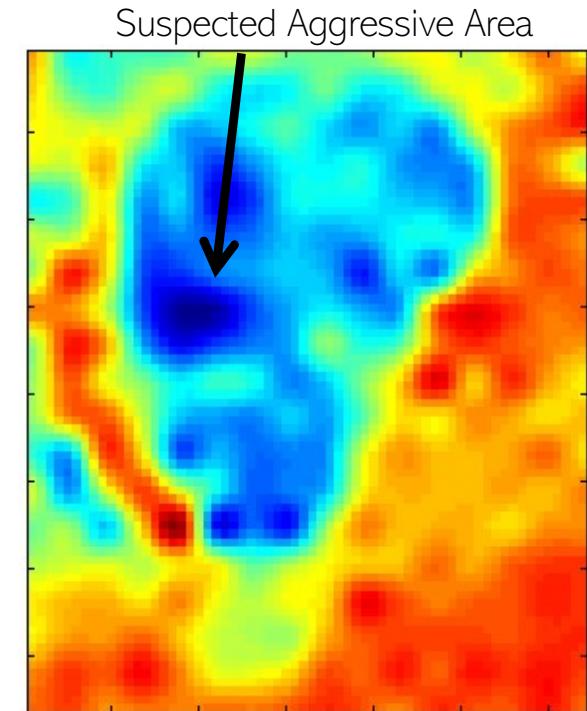
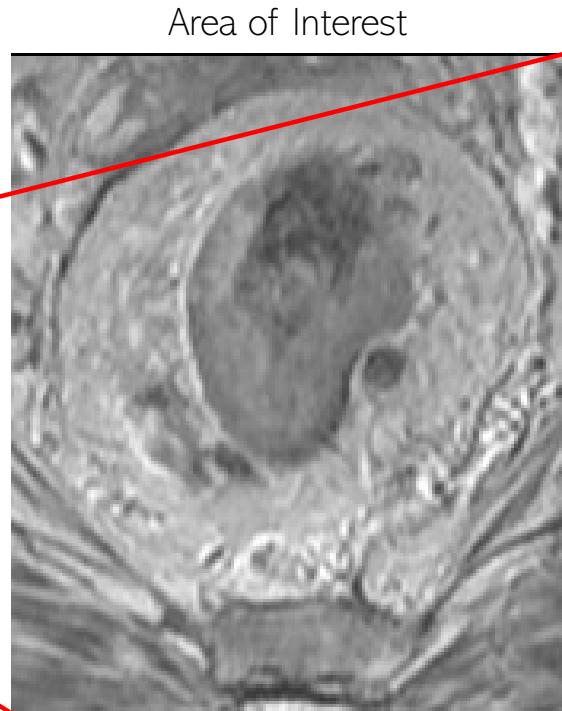
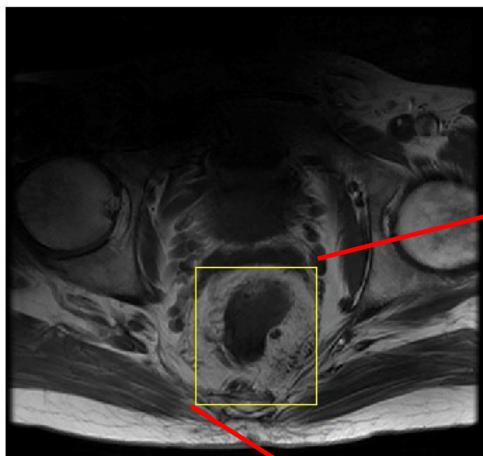
Relative Dispersion



FD signal mapping: Alzheimer's disease

Warsi, Molloy, Noseworthy (2012) MAGMA. 25:335-344.

Fractal Dimension (FD) Mapping: Oncology



Fractals and Chaos

- The studies of fractals and chaos are linked in many ways. They share common ideas and methods of analysis. Each field is used within the other.
- a phase space set can be formed from a time series. When this is fractal the system that generated the time series is chaotic.
- The fractal dimension of the phase space set tells us the minimum number of independent variables needed to generate the time series.
- Chaotic systems can be designed that generate a phase space set of a given fractal form.

Fractal vs Chaos

It is important to remember that the objects and processes studied by fractals and chaos are essentially different.

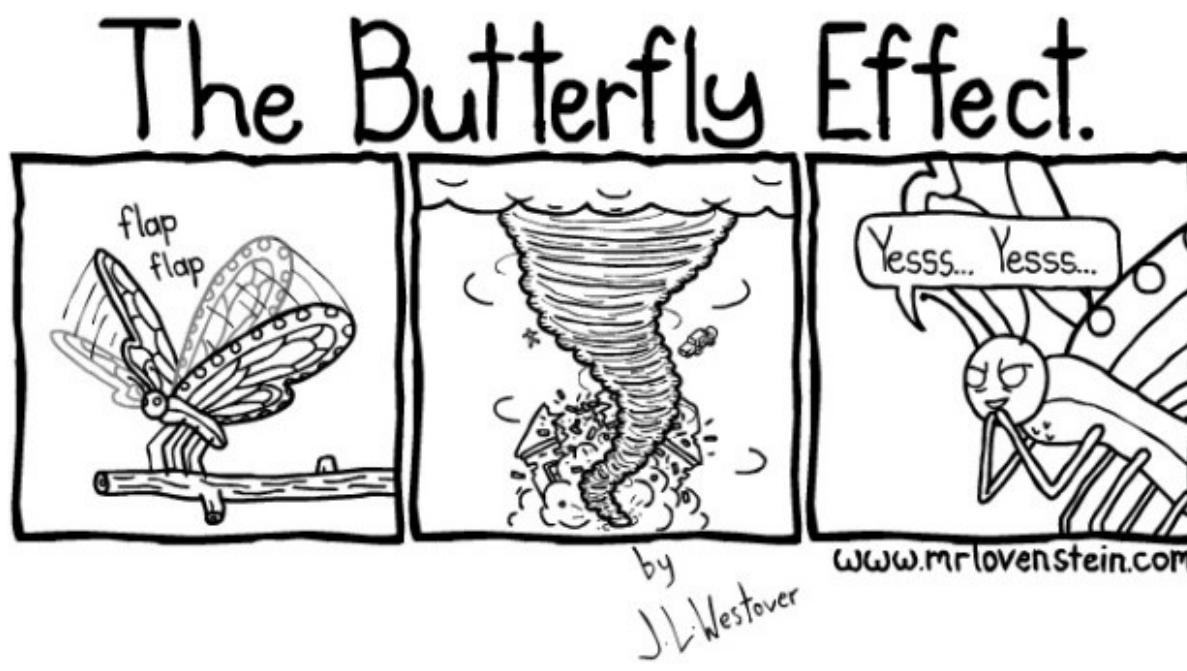
Fractals are objects or processes whose small pieces resemble the whole.

The goal of fractal analysis is to determine if experimental data contain self-similar features, and if so, to use fractal methods to characterize the data set.

Chaos means that the output of a deterministic nonlinear system is so complex that it mimics random behavior.

The goal of chaos analysis is to determine if experimental data are due to a deterministic process, and if so, to determine the mathematical form of that process.

Properties of Chaos

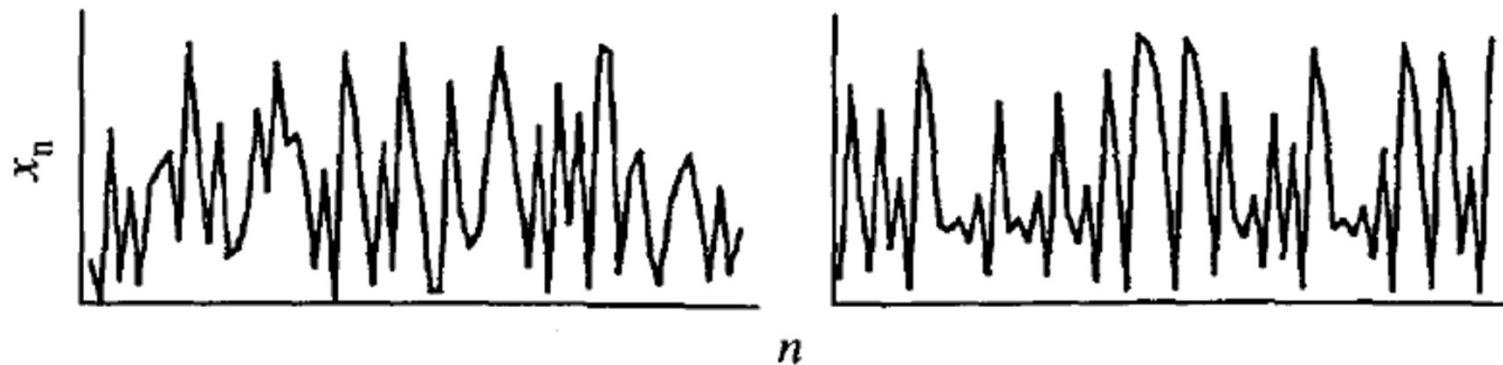


- Exponential Divergence
 - Divergence is the phenomenon of trajectories of a system that begin with similar initial conditions ending up with very different trajectories.
 - This is the opposite of convergence, in which systems tend towards the same value over long periods of time
 - While non-chaotic systems may show divergence, only chaotic systems have trajectories that diverge exponentially.

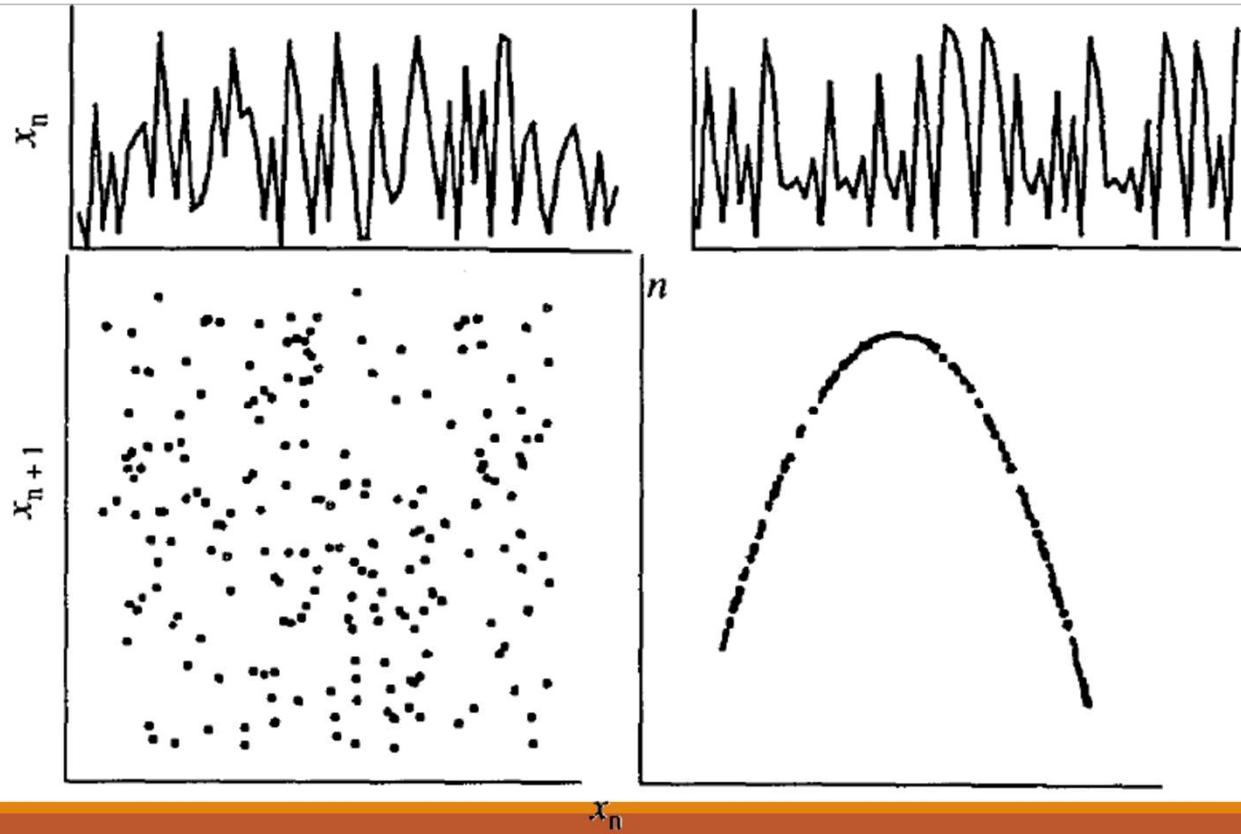
Properties of Chaotic Phenomena

Consider 2 time series. They have approximately the same statistical properties (i.e. similar means and variances)

- both look random



Plot X_n vs X_{n+1}



Chaos

- some systems are deterministic but the output is so complex that it mimics random behaviour
- jargon word is “chaos” which is unfortunate as day-to-day usage this means disordered!
- in mathematical systems chaos means ordered (but complex)
- need to differentiate between “chaos” and “noise”

Defining Properties of Chaos

- 1) Aperiodic
 - system is never repeated, ever
- 2) Bounded
 - system stays within a finite range and does not approach $\pm\infty$
- 3) Deterministic
 - in principle, x_0 can be used to calculate all future values of x_t
- 4) Sensitivity to initial conditions
 - two points that are initially close at $t=0$ will drift apart as time proceeds

1. A chaotic system is a deterministic dynamical system

(i.e. values of the variables that describe the system in the future are determined by the present values.

- e.g. third-order, single-variable equation for a nonlinear damped spring with sinusoidal forcing (Duffing equation):

$$\frac{d^2x}{dt^2} + \gamma \frac{dx}{dt} + \alpha x + \beta x^3 = B \cos(\omega t)$$

CONSTANTS

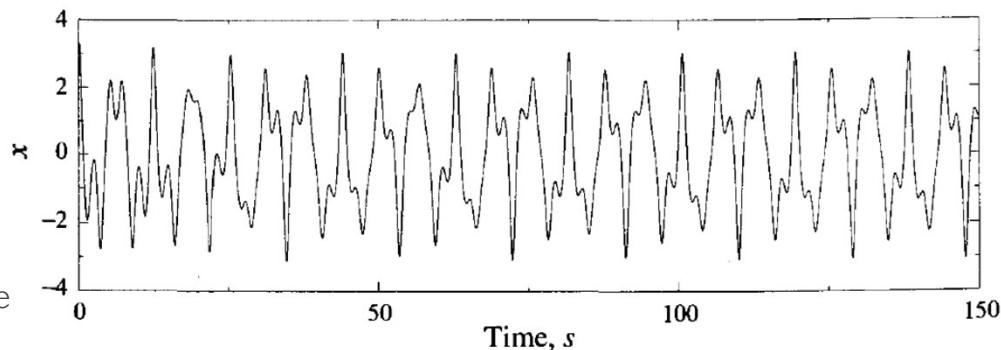
$\gamma = 0.05$: magnitude of damping (i.e. Friction)

$\alpha = 1$: magnitude of restoring force

$\beta = 1$: Amount of non-linearity in restoring force. If $\beta=0$ describes a damped and driven simple harmonic oscillator.

$B=7.5$: controls amplitude of periodic driving force. If $\gamma=0$ there is no driving force

$\omega = 1$: frequency of periodic driving force



2. A chaotic system is described by either difference or differential equations

- A difference equation has values where variables at the next time step are a function of their current values.
 - Values of variables computed at discrete steps (i.e. logistic equations are one-variable difference equations that can demonstrate chaotic behavior).
- A differential equation, has values of variables changing continuously in time. The values of future variables depend on current values and the derivatives of their current values.
- A chaotic system can consist of a single equation with one variable if it is discrete, or a set of coupled equations with more than three variables if it is continuous (e.g. Duffing equation)

3. A chaotic system has sensitivity to initial conditions.

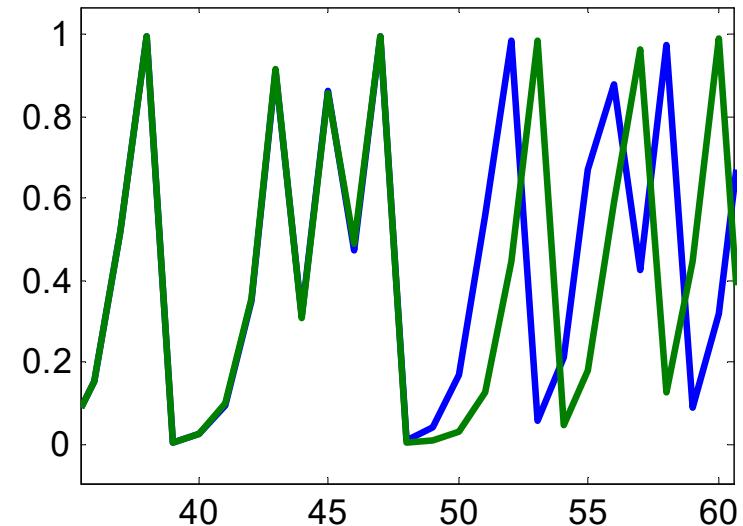
- values of the variables after a given time depend on their initial values.
- very small changes in these initial values produce very large changes in later values. i.e. if initial values at time $t=0$ were $x_1(t = 0)$ and $x_2(t = 0)$, then after a time t :

where λ is called the Lyapunov exponent. Thus, the difference in the values diverges exponentially quickly in time.

- A system displaying this sensitivity to initial conditions is said to be "chaotic."

Sensitivity to Initial Conditions

- Since the divergence is exponential, a plot of the log divergence should give a straight line
- The slope of the line gives an estimate of the Lyapunov exponent, a measurement of how quickly the divergence happens



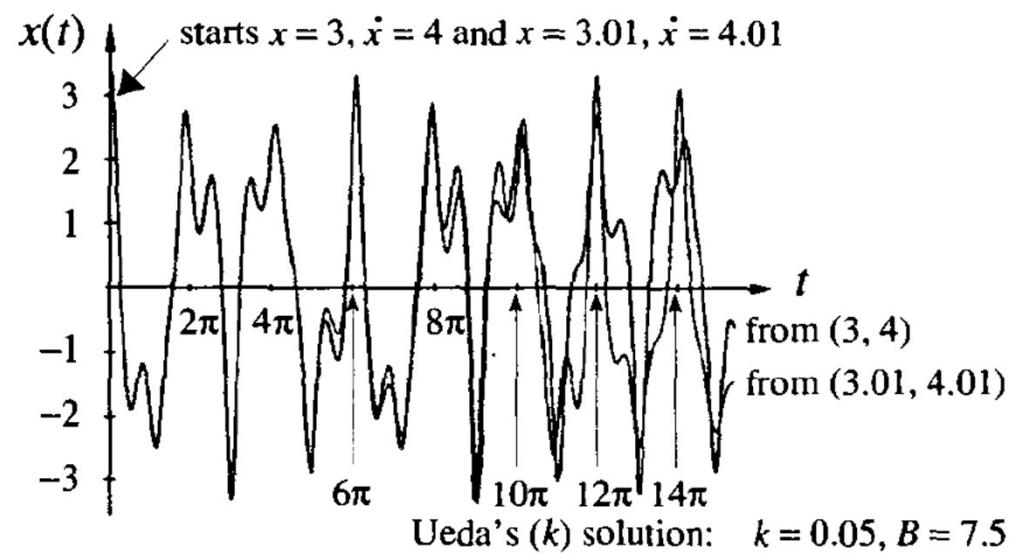
Sensitivity to Initial Conditions

We can only specify values of initial conditions to finite precision

- accuracy is continually lost.
- we cannot predict exactly their values over a long time.

Therefore:

- although the chaotic system is fully deterministic,
it is not predictable in the long run, because
of the sensitivity to initial conditions.



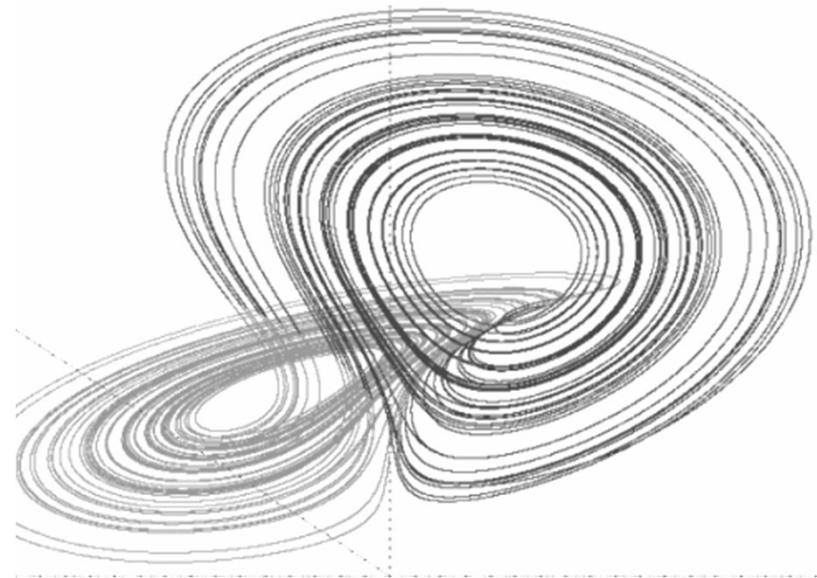
4. The values of the variables are not predictable in the long run.

- Because of the sensitivity to initial conditions, values computed [as time goes by] will diverge ever further from their true values based on their exact initial values.

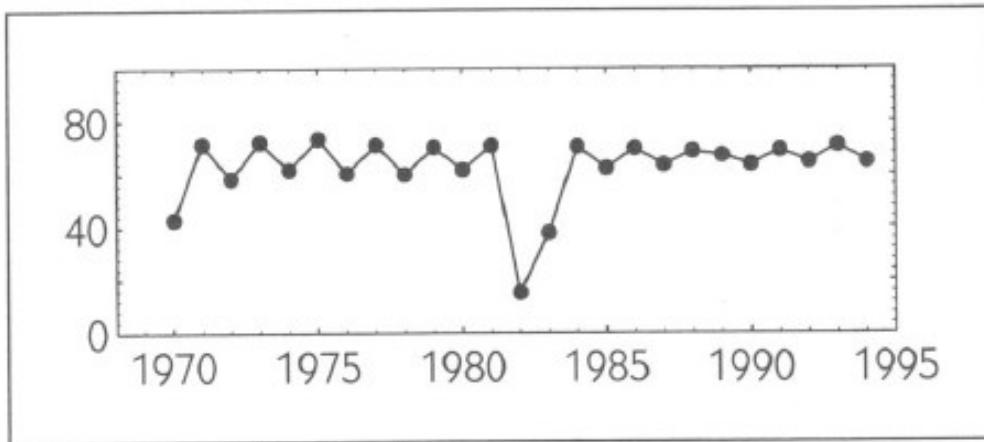
e.g. if values of initial conditions are specified with 5 digits, in time the accuracy of their values will fall to 4, then 3, then 2, digits.

5. The values of the variables do not take on all possible values.

- In the long run, although the values of the variables seem to fluctuate widely, they do not take on all combinations of values.
- This restricted set of possible values is called the attractor.



Finite Difference Equations



$$N_{t+1} = f(N_t)$$

Equations of this form, which relate values at discrete times (e.g., each May), are called finite-difference equations.

$$N_1 = RN_0,$$

$$N_2 = RN_1 = R^2 N_0,$$

$$N_3 = RN_2 = R^2 N_1 = R^3 N_0,$$

⋮

$$N_t = R^t N_0$$

Verification:

$$N_{t+1} = R^{t+1} N_0 = RR^t N_0 = RN_t$$

Example: Different Constants

Figure 1.2
The solution to
 $N_{t+1} = 0.90N_t$.

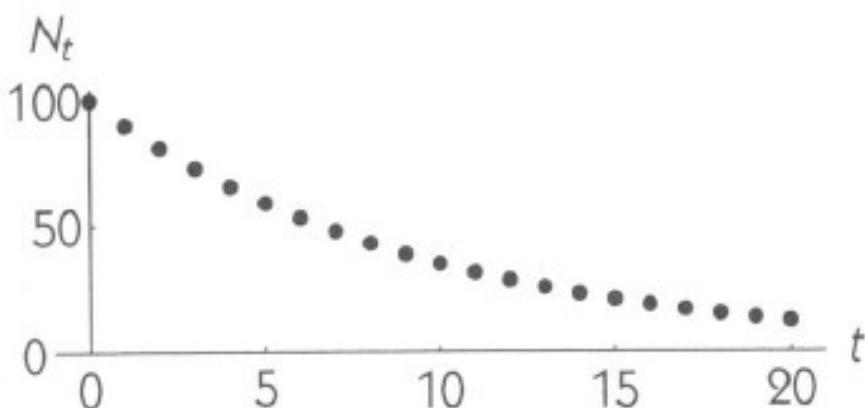
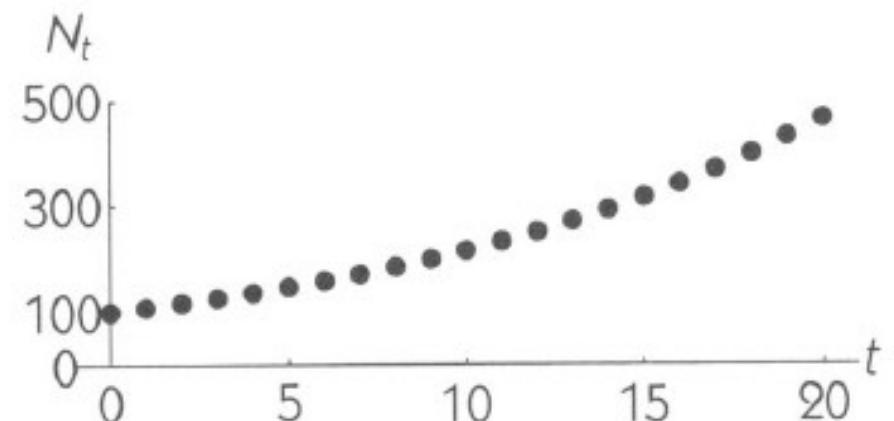


Figure 1.3
The solution to
 $N_{t+1} = 1.08N_t$.



Example: Different Constants

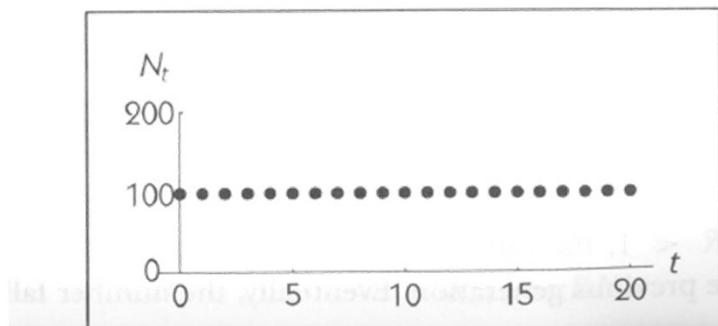


Figure 1.4
The solution to
 $N_{t+1} = 1.00N_t$.

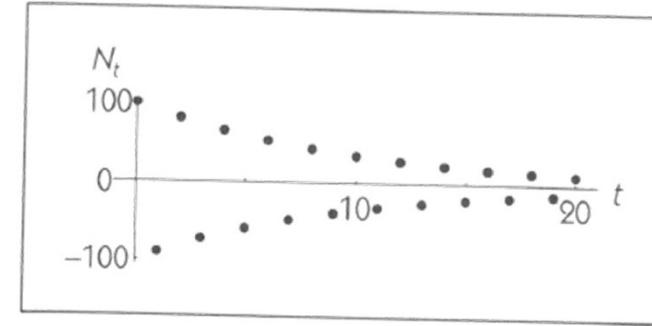


Figure 1.5
The solution to
 $N_{t+1} = -0.90N_t$.

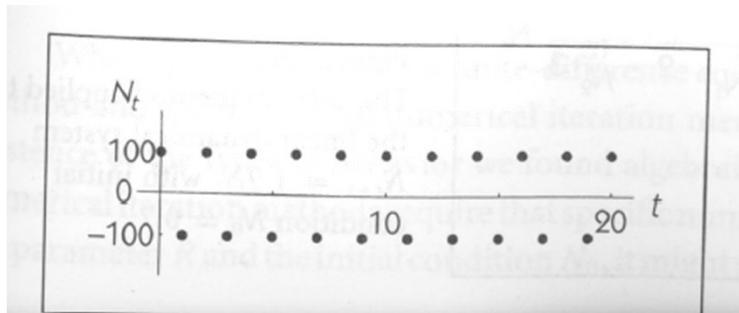


Figure 1.7
The solution to
 $N_{t+1} = -1.00N_t$.

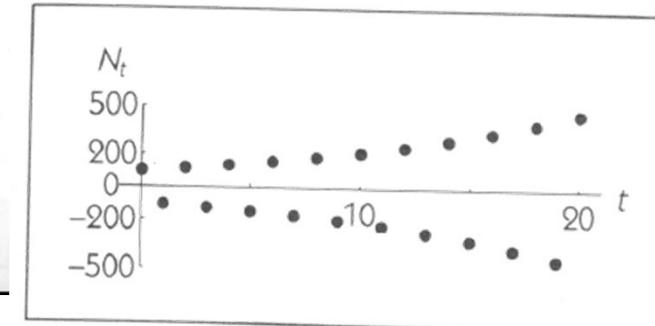


Figure 1.6
The solution to
 $N_{t+1} = -1.08N_t$.

Diverging

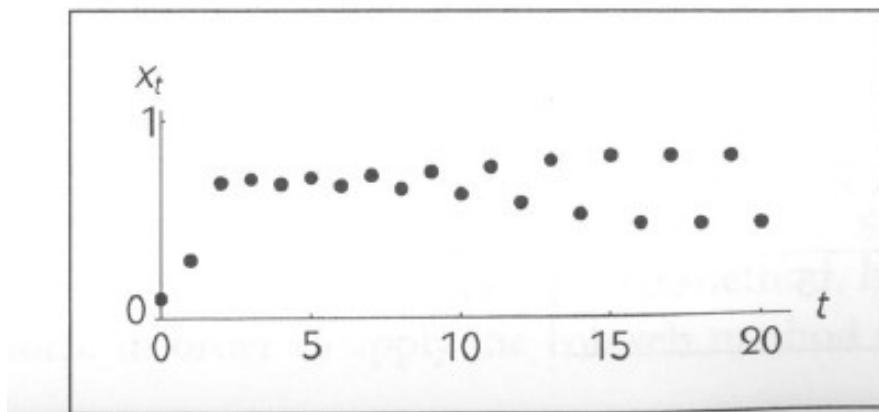


Figure 1.13
The solution to
 $x_{t+1} = 3.3(1 - x_t)x_t$.

Approaching Chaos

Figure 1.14
The solution to
 $x_{t+1} = 3.52(1 - x_t)x_t$.

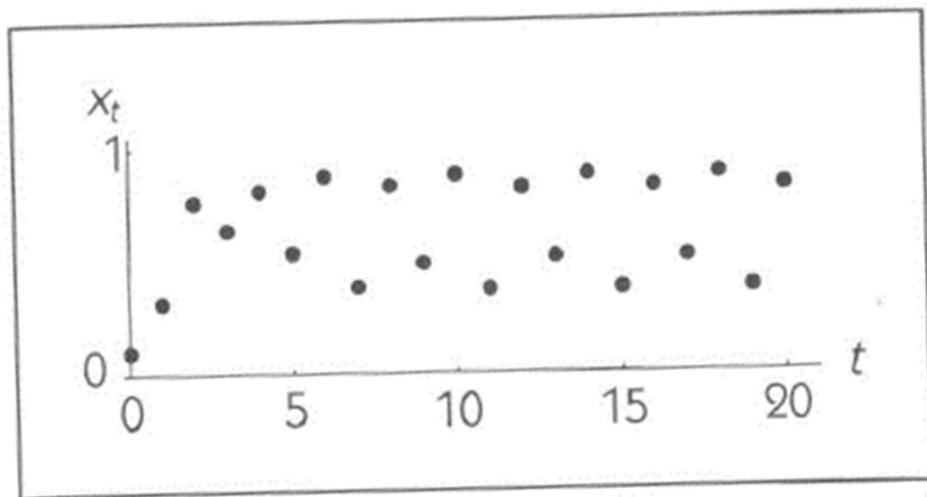
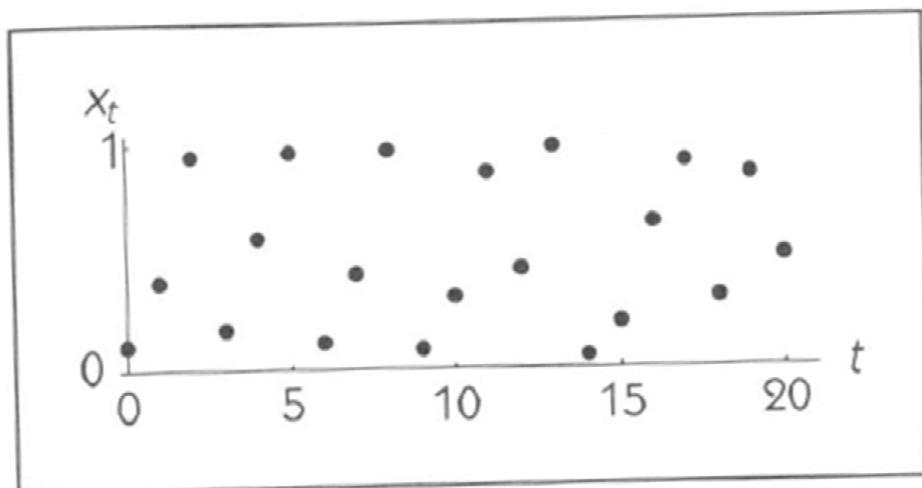


Figure 1.15
The solution to
 $x_{t+1} = 4(1 - x_t)x_t$.



Aperiodic Behaviour

– now we have reached chaos

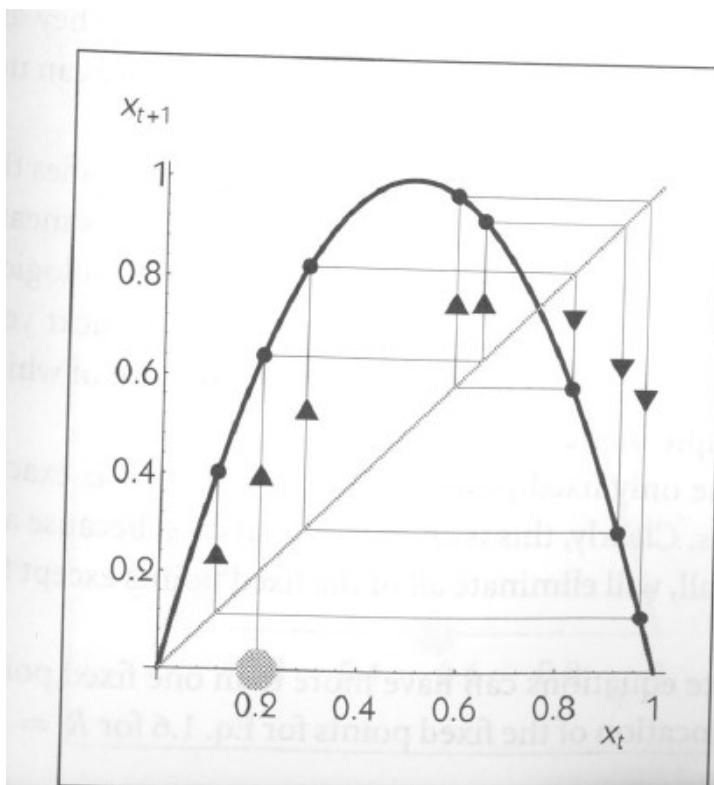


Figure 1.16
Cobweb iteration of
 $x_{t+1} = 4(1 - x_t)x_t$.

Methods of Iteration: Numerical

$N_{t+1} = 0.9N_t$ with $N_0 = 100$

$$N_0 = 100,$$

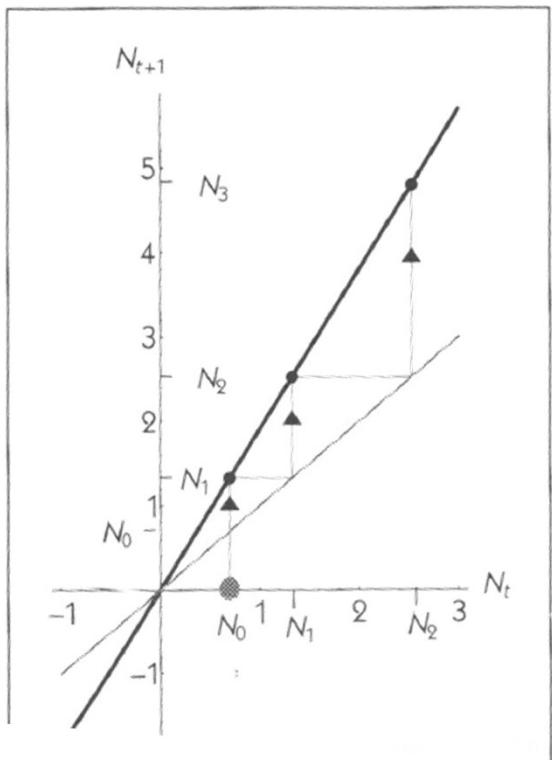
$$N_1 = f(N_0) = 0.9 \times 100 = 90,$$

$$N_2 = f(N_1) = 0.9 \times 90 = 81,$$

$$N_3 = f(N_2) = 0.9 \times 81 = 72.9,$$

⋮

Methods of Iteration: Cobweb Plot



$$N_{t+1} = 1.9N_t \text{ with } N_0 = 0.7$$

Nonlinear Finite-Difference Equations

In the linear equation, R is the number in each generation t

- to make the number of subsequent generations decrease as N_t gets larger, make the growth rate a function of N_t
- A simple solution is the function $(R - bN_t)$
- b governs how growth rate decreases as the total gets larger
- R is the growth rate when the population is very, very small.

This assumption that the number is $(R - bN_t)$ gives a new finite-difference equation:

Nonlinear Finite-Difference Equations

Here there are 2 parameters, R and b, that can vary independently.

- However, a simple change of variables shows that there is only one parameter that affects the dynamics:

- Define a new variable $xt = bNt/R$ which is just a way of scaling the total number by b/R
- Substituting xt and $xt+1$ results in:

- A solution cannot generally be found using algebra. Hence, numerical iteration and the cobweb method, is needed to find solutions.

Steady States and their Stability

A simple type of dynamical behavior is when the system stays at a steady state. A steady state is a state of the system that remains fixed, that is, where:

Steady states in finite-difference equations are associated with the mathematical concept of a fixed point.

Fixed Points

There are 3 important questions to ask about fixed points in finite difference equations:

1. Are there any fixed points?
2. If the initial condition happens to be near a fixed point, will subsequent iterates approach the fixed point? If subsequent iterates approach the fixed point, we say the fixed point is locally stable.
3. Will the system approach a given fixed point regardless of the initial conditions? If the fixed point is approached for all initial conditions, we say that the fixed point is globally stable.

Finding Fixed Points

- A fixed point of a function $f(xt)$ is a value xt^* that satisfies $xt^*=f(xt^*)$
- From the graph of $xt+1=f(xt)$ it is easy to locate fixed points: These are the points where the graph intersects the line $xt+1=xt$
- Or, can solve the equation $xt=f(xt)$
- For a linear finite-difference equation, xt^* is a fixed point if it satisfies $xt^*=Rxt^*$
- One solution is always $xt^*=0$ (i.e. the origin is a fixed point for a linear system).

Finding Fixed Points - Example

- The solution $x_t = 0$ is the only fixed point, unless $R = 1$
- If R is exactly 1, then all points are fixed points
- Clearly, this defines an exceptional case, because any change in R will eliminate all fixed points (except the one at the origin)
- non-linear finite difference equations can have >1 fixed points.

$$x_{t+1} = x_t$$

Cobweb Plot

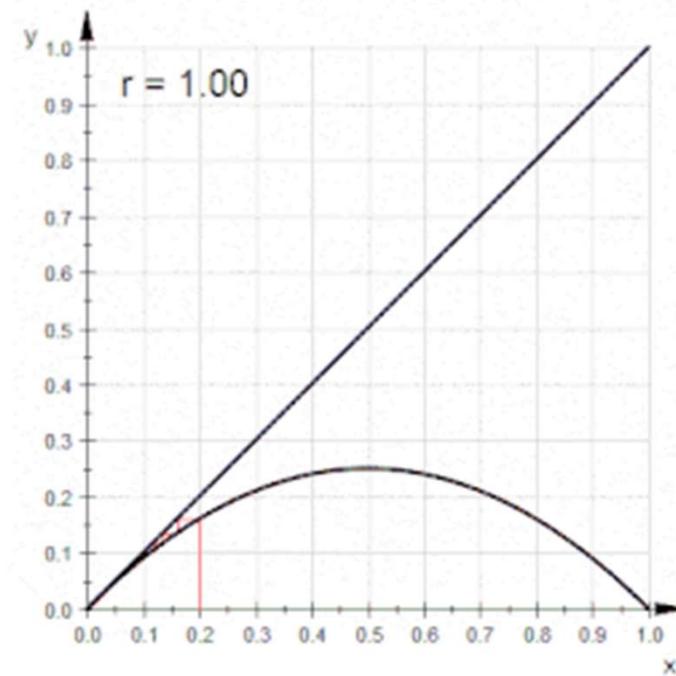
used to investigate the qualitative behaviour of 1D iterated functions

possible to infer the long-term status of an initial condition under repeated application

Assess Stability

a stable fixed point → inward spiral

an unstable fixed point → outward spiral

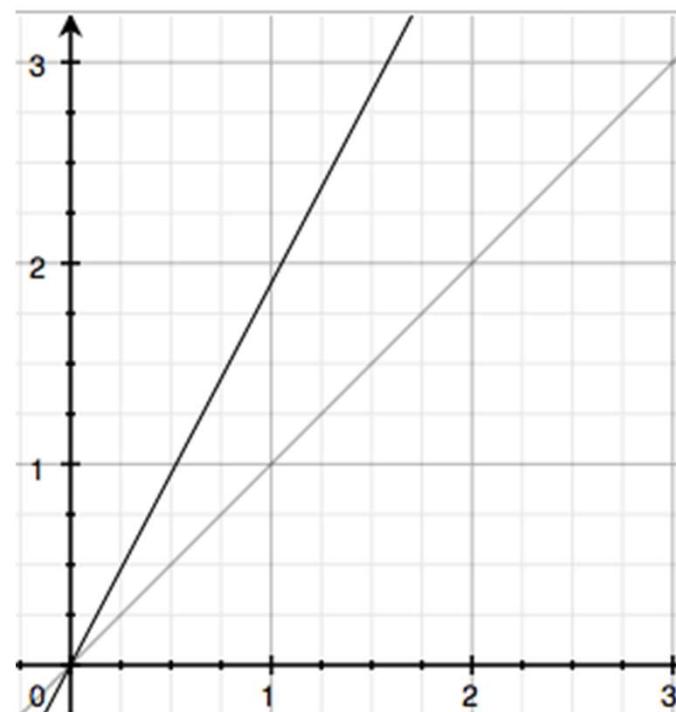


The Cobweb Method

- a graphical method for iterating finite-difference equations

e.g. linear system of: $N_{t+1} = RN_t$

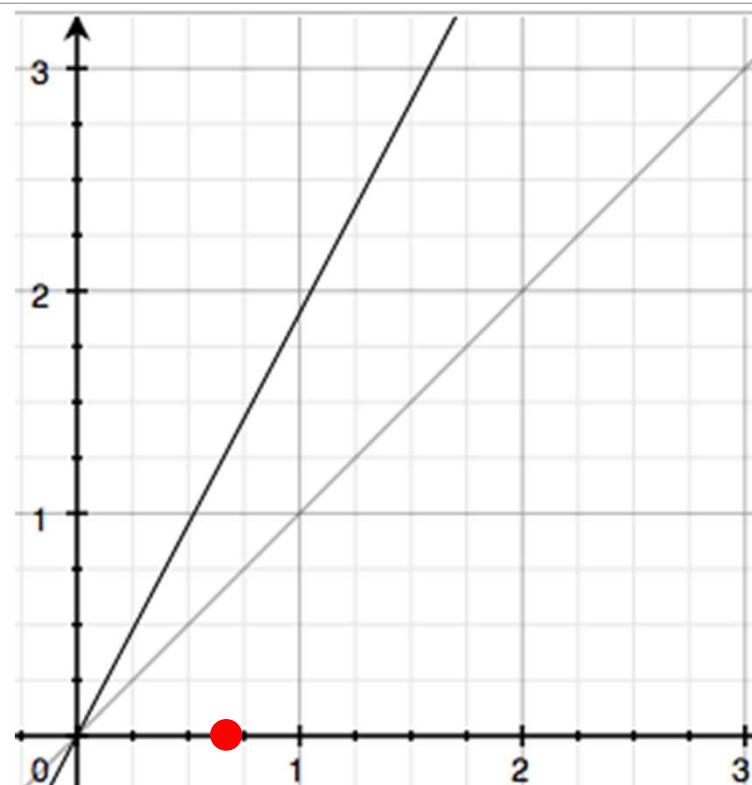
1. Graph the function. Here it is $f(N_t) = RN_t$. Pick a specific value for R (e.g. 1.9), so that the finite difference equation is $N_{t+1} = 1.9N_t$



The Cobweb Method

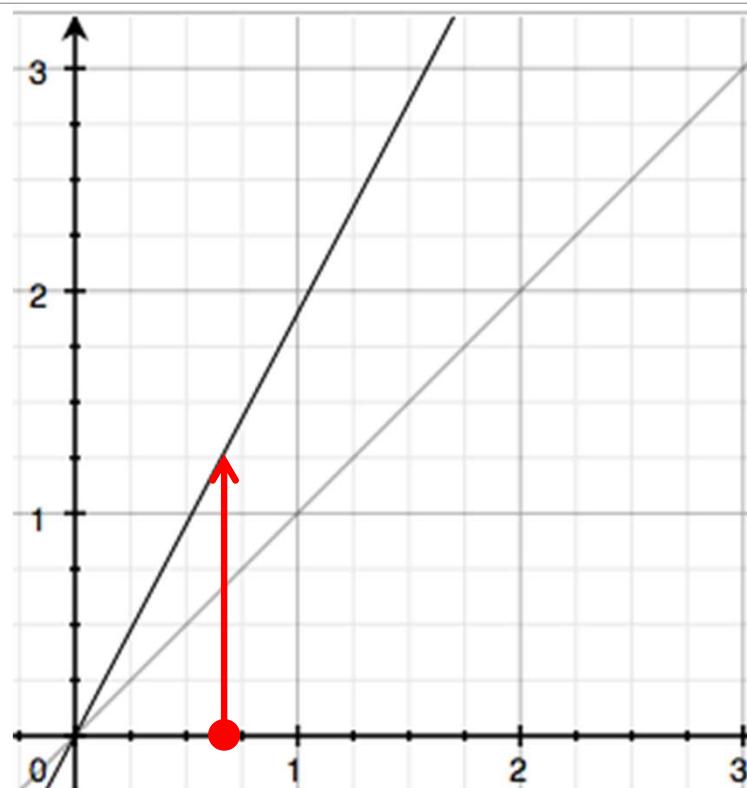
2. Pick an initial condition N_0)

In this example = 0.7, shown
as the red dot on the x-axis
below



The Cobweb Method

3. Draw a vertical line from N_0 on the x-axis up to the function. The position where this vertical line hits the function (shown as a solid dot at the end of the arrow) tells us the value of N_1

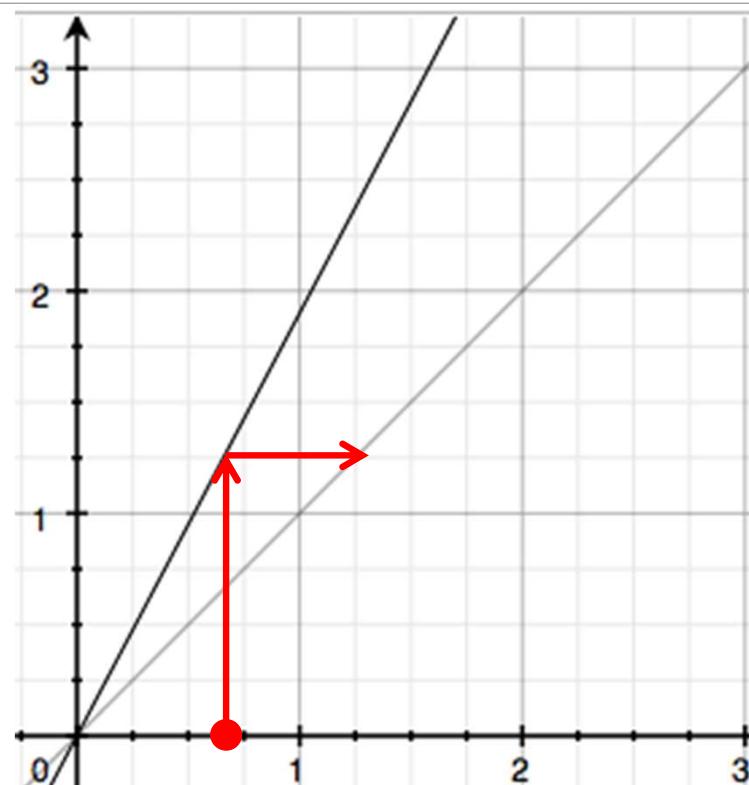


The Cobweb Method

4. Take this value of N_1 , plot it again on the x-axis, and again draw a vertical line to find the value of N_2 .

Simple shortcut:

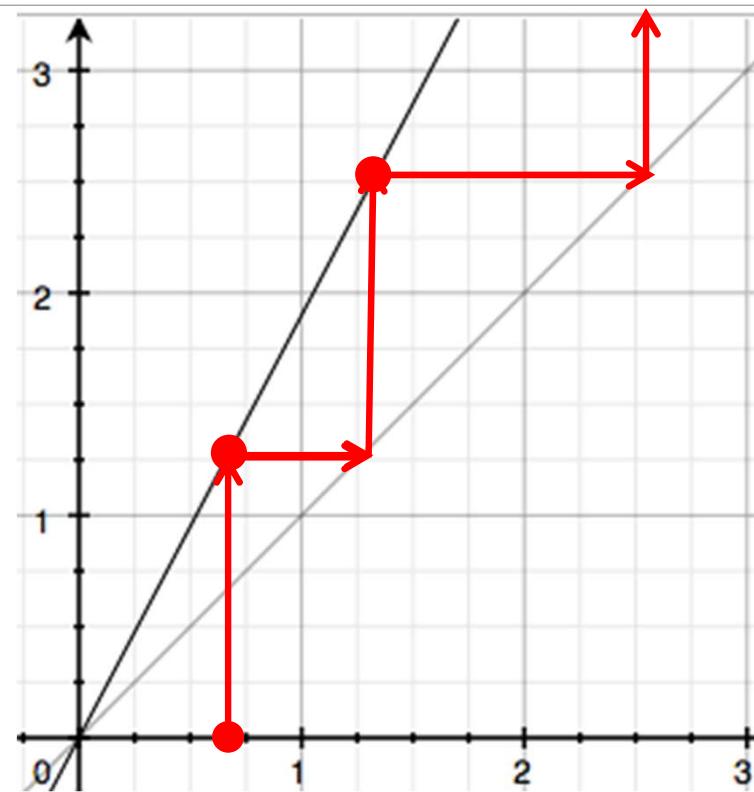
- Draw a horizontal line to the $N_{t+1} = N_t$ line.
- The place where the horizontal line intersects the 45-degree line is the point from which to draw the next vertical line to find N_2



The Cobweb Method

5. To find N_3, N_4 , etc., repeat the process of drawing vertical lines to the function and horizontal lines to the line of $N_{t+1} = N_t$

- The result of iterating $N_{t+1} = 1.9N_t$ is growth towards ∞
- This is consistent with the algebraic solution found previously for $R > 1$



Cobweb Method - Nonlinear

In order to apply the cobweb method to we first must draw a graph of the function:

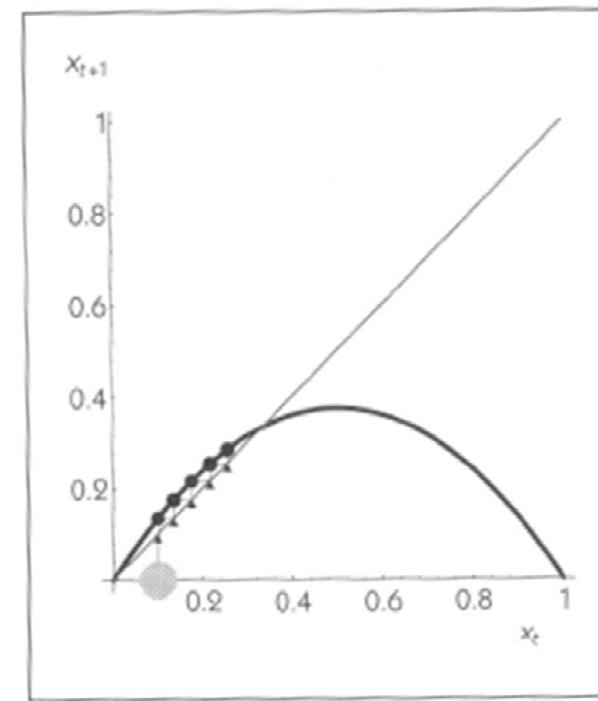
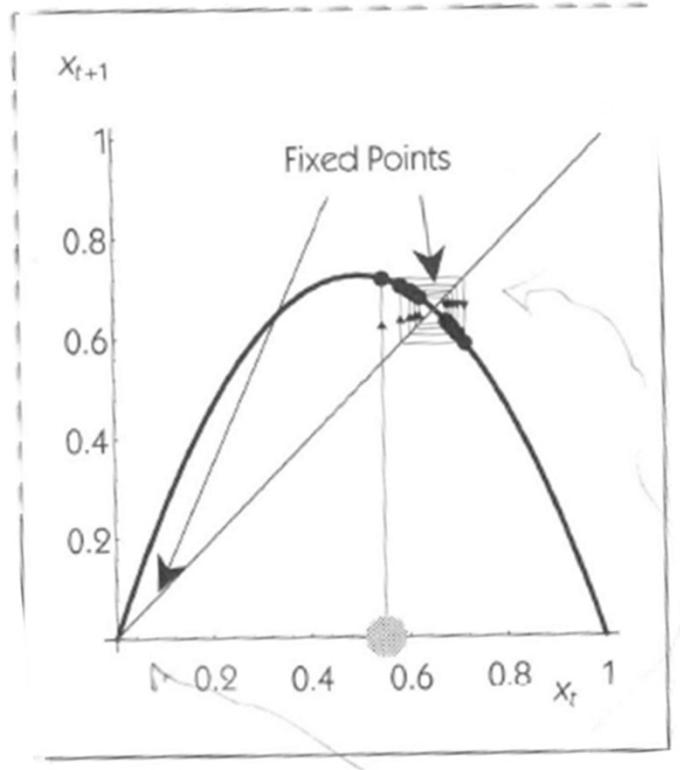
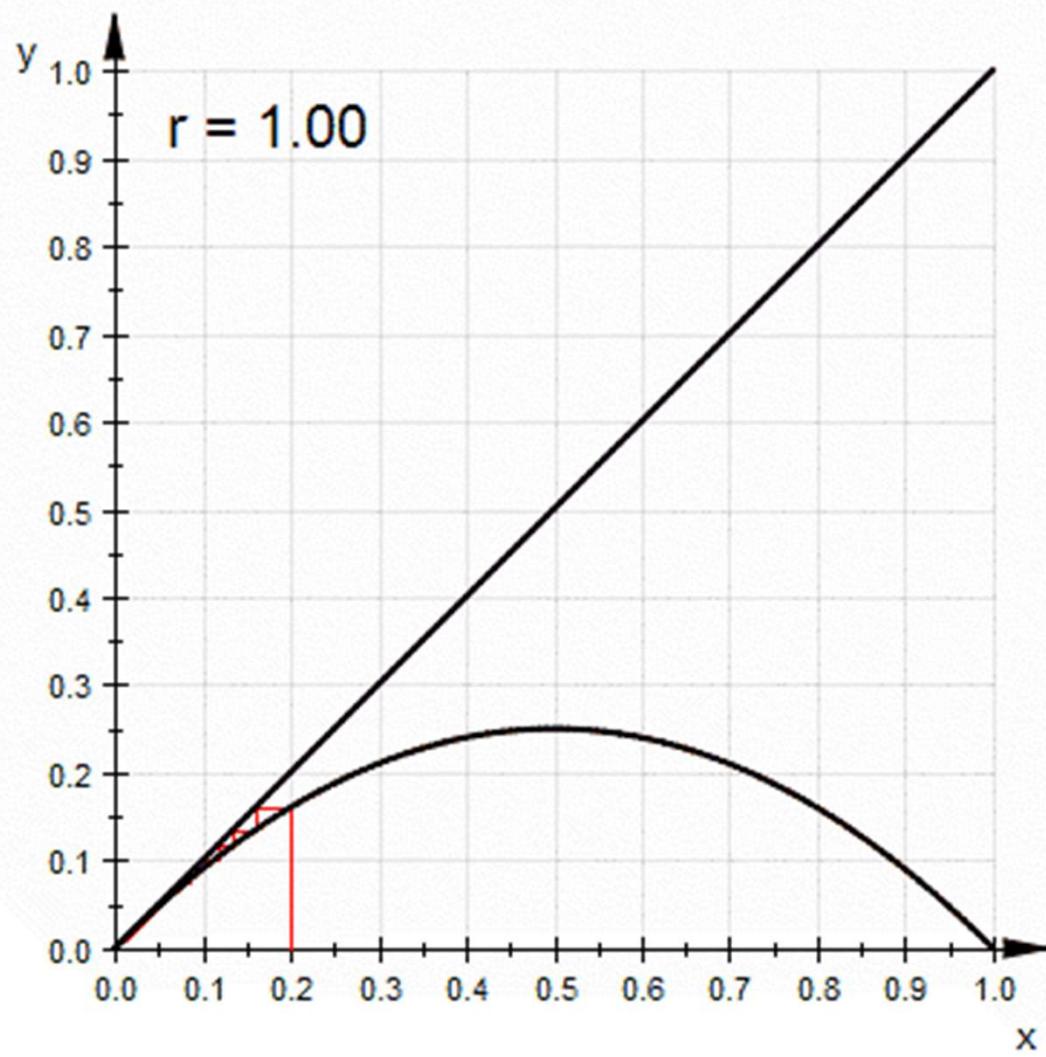


Figure 1.9
Cobweb iteration of
 $x_{t+1} = 1.5(1 - x_t)x_t$.

Finding Fixed Points



$$x_{t+1} = 2.9(1 - x_t)x_t$$



Non-linear Finite-Difference Equations

$$N_{t+1} = (R - bN_t)N_t = RN_t - bN_t^2$$

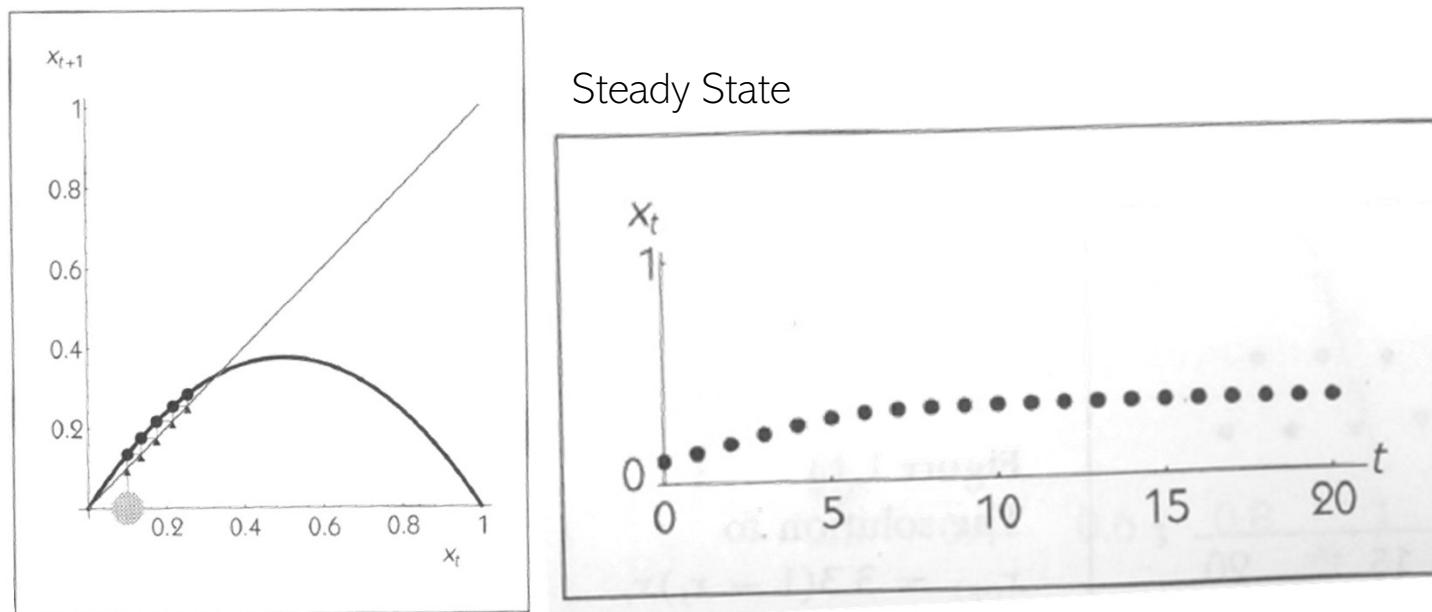


Figure 1.10
The solution to
 $x_{t+1} = 1.5(1 - x_t)x_t$

Periodic Cycles

Figure 1.11
The solution to
 $x_{t+1} = 2.9(1 - x_t)x_t$.

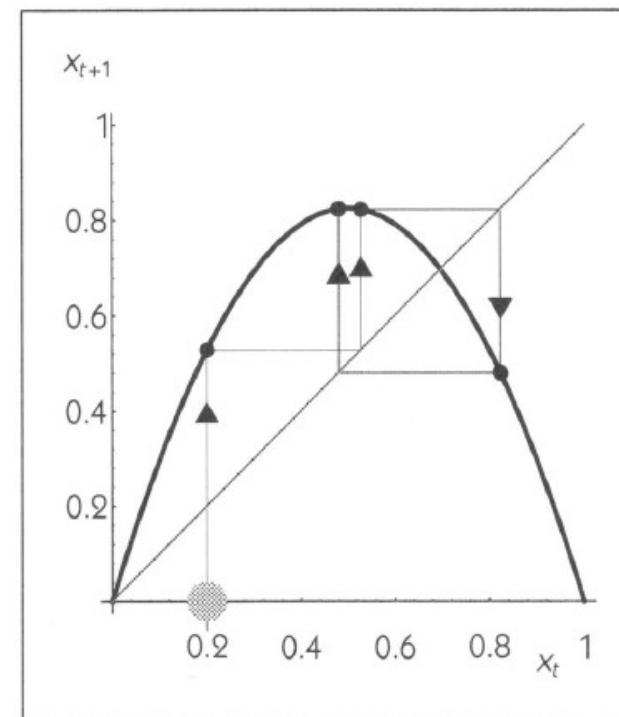
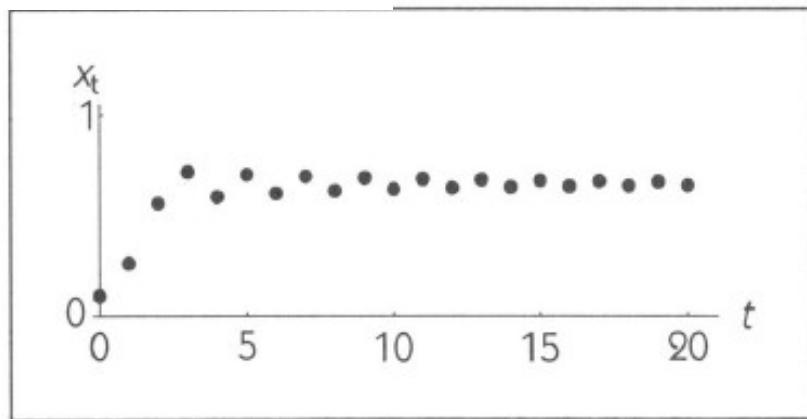


Figure 1.12
Cobweb iteration of
 $x_{t+1} = 3.3(1 - x_t)x_t$.

Example

- Cells reproduce by division
- One way to regulate the rate of reproduction of cells is by regulating mitosis.
- There is biochemical evidence that there are compounds, called chalones, that are tissue-specific inhibitors of mitosis
- For simplicity, assume that the generations of cells are distinct and that the number of cells in each generation is given by N_t

Example

- Following the same logic as before, assume that for each cell in generation t , there are R cells in generation $t+1$. (If every cell divided in half every time step, then R would equal 2.)
- The finite difference equation describing this situation is the linear equation $N_{t+1} = RN_t$, which leads either to exponential growth or to decay to zero.
- A possible role of chalones is to make R depend on the number of cells.
- Assume that the amount of chalones produced is proportional to the number of cells. The more chalone there is, the greater the inhibitory effect on mitosis

Example

The biochemical action of chalones is to bind to a protein involved in mitosis, rendering the protein inactive.

Binding of molecules to proteins is often modeled by a Hill function

We will assume that $n \geq 2$ and $R=2$, $\theta=5$, and $n=3$.

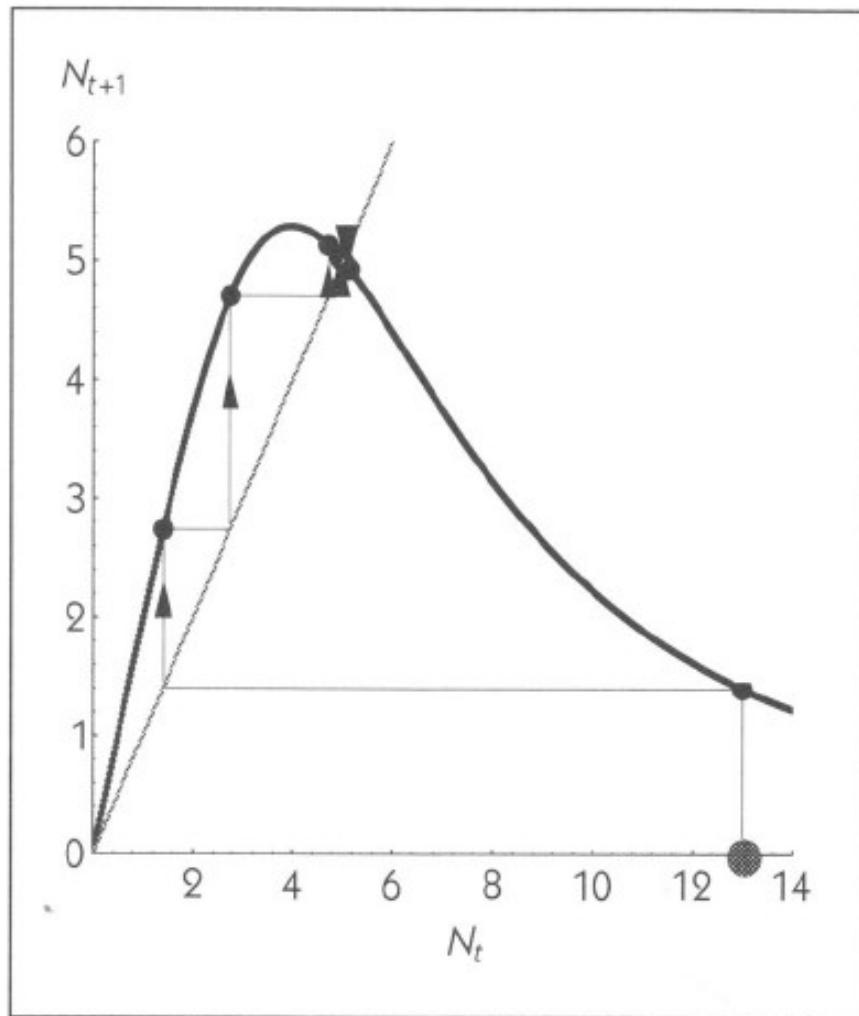
Find the fixed points of this system and determine their stability

Example

1. To determine the fixed points we solve the equation

$$N^* = \frac{RN^*}{1 + \left(\frac{N^*}{\theta}\right)^n}$$

These are the only fixed points. There are also imaginary solutions that can be ignored in this case because we are only concerned with biologically meaningful solutions, and the number of cells in each generation must be real numbers



Example – Fixed point stability

2. To determine the stability of the fixed points you have to computer the slope at the fixed points. If we take the derivative we find:

3. From the above equation we find that the slope at the fixed point $N^* = 0$ is just R .

- If $R > 1$, the fixed point at the origin is always unstable. (To be a plausible model of the regulation of cell reproduction, we must have $R > 1$.)
- Otherwise, the population would always fall to zero even in the complete absence of the mitosis-inhibiting chalones.)

Example – Fixed point stability

The slope at the fixed point $N^* = \theta(R - 1)^{\frac{1}{n}}$ is

$$\left. \frac{df}{dN_t} \right|_{N^*} = 1 + n \left(\frac{1}{R} - 1 \right).$$

For $R = 2$, the fixed point will be unstable when $n > 4$ and stable otherwise.

- Local stability tells us whether the fixed point is approached if the initial condition is sufficiently close to the fixed point.
- can be assessed simply by looking at the slope of the function at the fixed point.

A more difficult-question is whether a locally stable fixed point is globally stable.

Fixed Point Stability

For linear finite-difference equations a locally stable fixed point is also globally stable:

i.e. regardless of initial conditions, the iterates will eventually reach the locally stable point (i.e., the origin) from any initial condition.

Stability and dynamics at fixed points

- $|df/dX| > 1$: Unstable (doesn't converge at fixed point)
- $0 < df/dX < 1$: Stable, **monotonically** approach to fixed point (converges)
- $-1 < df/dX < 0$: Stable, **oscillatory** approach to fixed point (converges)

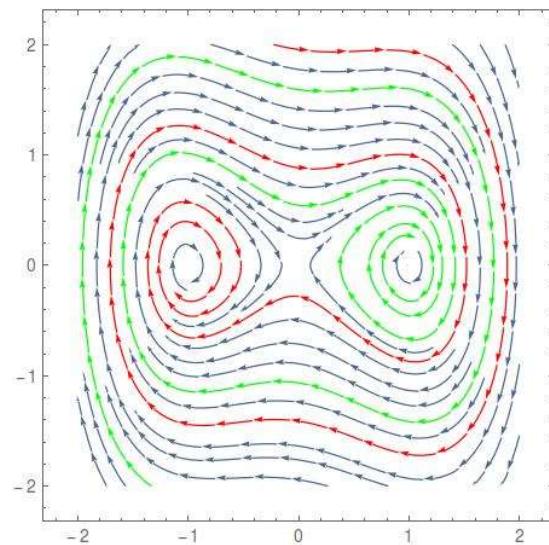
Fixed Point Stability

- For nonlinear finite-difference equations, there can be more than one fixed point. AND when multiple fixed points are present, none of the fixed points can be globally stable.

The set of initial conditions that eventually leads to a fixed point is called the basin of attraction of the fixed point.

Often, the basin of attraction for fixed points in nonlinear systems can have a very complicated geometry

If multiple fixed points are locally stable we say there is [multistability](#).



The Period-Doubling Route to Chaos

We have seen that the simple finite-difference equation:

$$N_{t+1} = (R - bN_t)N_t = RN_t - bN_t^2$$

can display various qualitative types of behavior for different values of R:

- 1) steady states; 2) periodic cycles of different lengths; 3) chaos.

The change from one form of qualitative behavior to another as a parameter is changed is called a bifurcation.

An important goal in studying nonlinear finite-difference equations is to understand the bifurcations that can occur as a parameter is changed!

Feigenbaum

- For $3.0000 < R < 3.4495$, there is a stable cycle of period 2.
- For $3.4495 < R < 3.5441$, there is a stable cycle of period 4.
- For $3.5441 < R < 3.5644$, there is a stable cycle of period 8.
- For $3.5644 < R < 3.5688$, there is a stable cycle of period 16.
- As R is increased closer to 3.570, there are stable cycles of period 2^n , where the period of the cycles increases as 3.570 is approached.
- For values of $R > 3.570$, there are narrow ranges of periodic solutions as well as aperiodic behavior.

Feigenbaum's Number

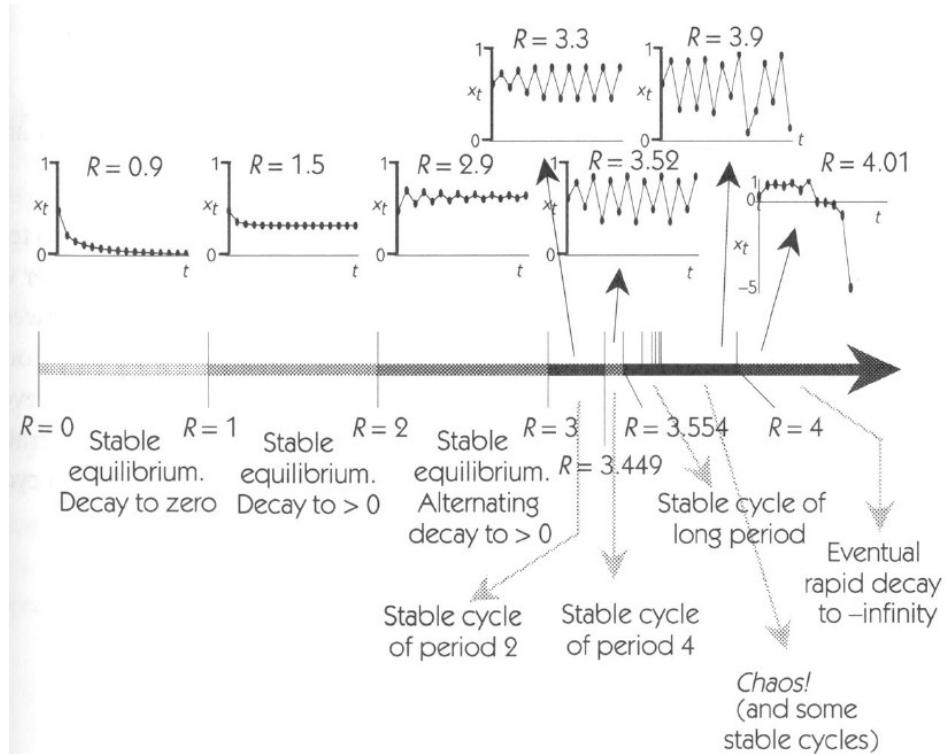
This illustrates a sequence of period-doubling bifurcations at $R = 3.0000$, $R = 3.4495$, $R = 3.5441$, $R = 3.5644$, with additional period-doubling bifurcations as R increases.

- This transition from the stable periodic cycles to the chaotic behavior at $R = 3.570$ is called the period-doubling route to chaos.

$$\lim_{n \rightarrow \infty} \frac{\Delta_n}{\Delta_{2n}} = 4.6692 \dots$$

The constant, $4.6692 \dots$ is now called Feigenbaum's number, appearing not only in the simple theoretical models but also in other theoretical models and in experimental systems in which there is a period-doubling route to chaos.

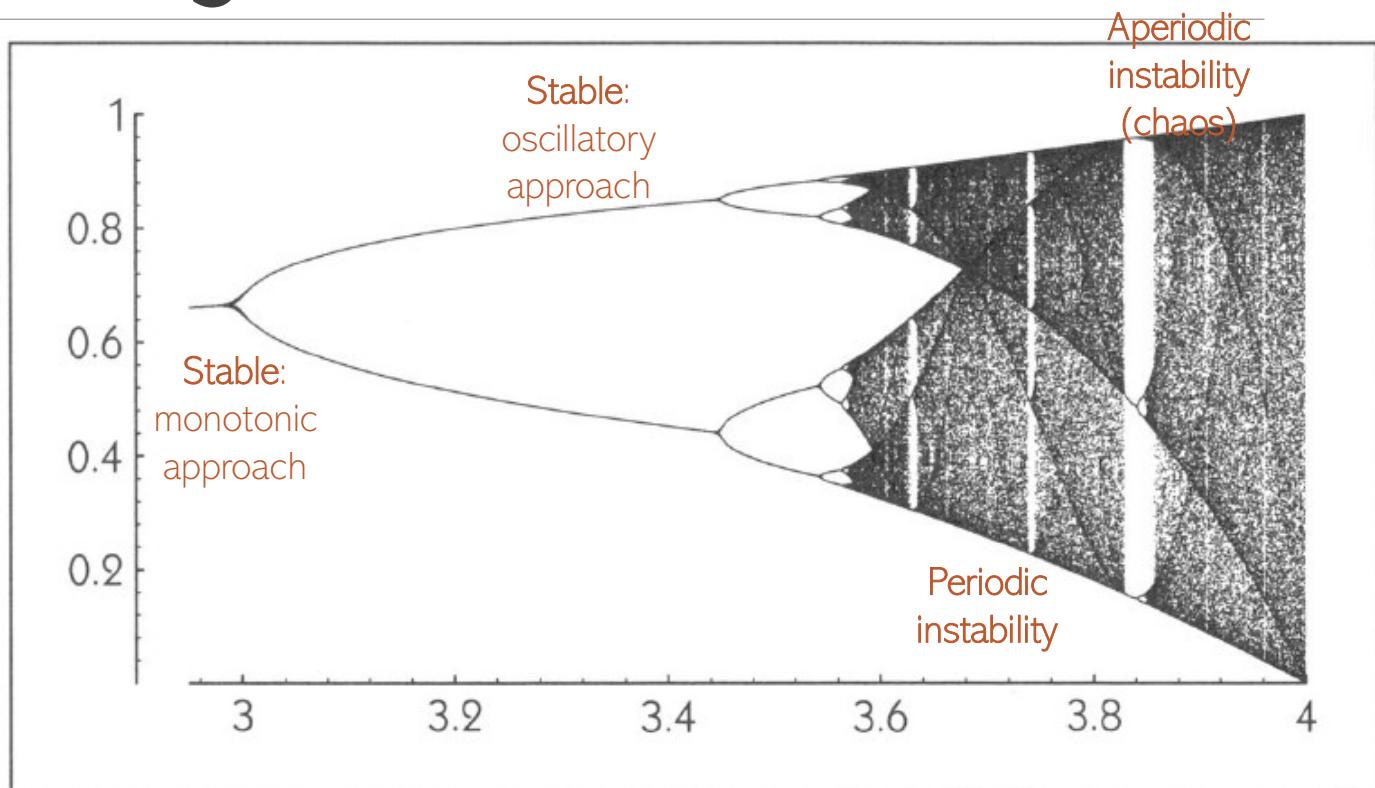
Chaotic Behaviour



Bifurcation Diagram

try matlab code:
feigen(0.5,100,100);

A bifurcation diagram
of nonlinear finite
difference equation
with asymptotic
values of X_t plotted
vs. R



Non-linear Dynamics

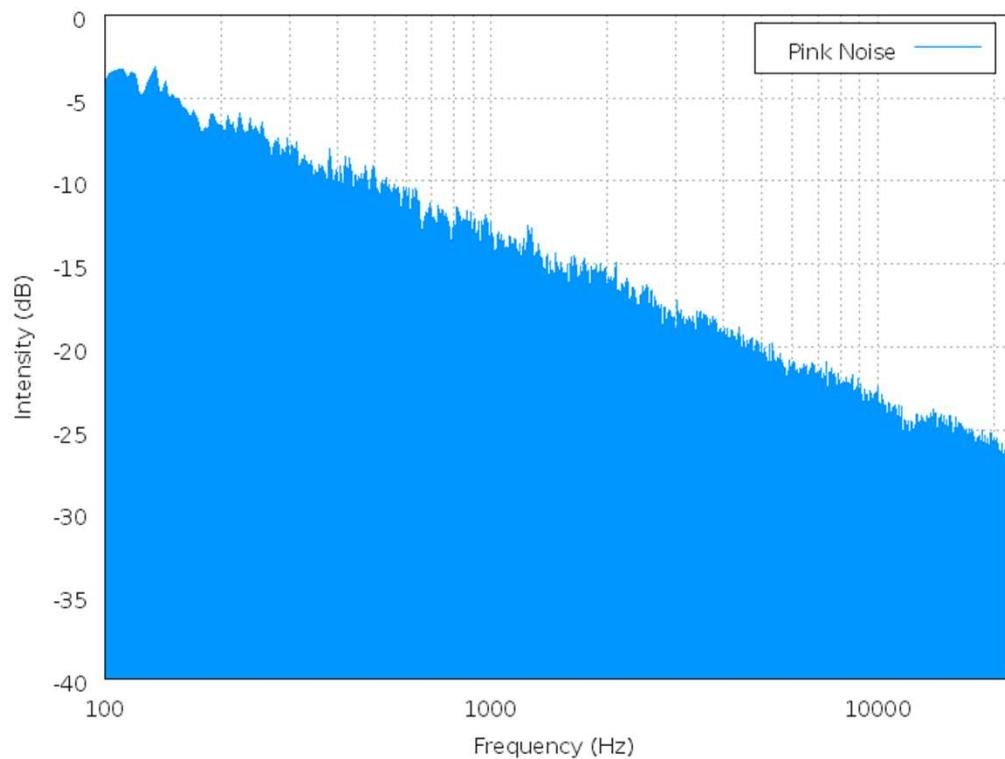
Self-organized criticality (SOC)

- refers to a type of dynamical behavior exhibited by systems with many interacting degrees of freedom where minor fluctuations lead to system rearrangements without characteristic size or time scale.
- rearrangements are referred to as avalanches with power laws describing their size, lifetime, and power spectrum.

SOC-like behavior is often associated with the formation of fractals in nature

Non Linear Dynamics

- Seen in biological systems, is present in heart beat rhythms, neural activity, mental states (modeling in psychology) and the statistics of DNA sequences.
- Also describes the statistical structure of many natural images (images from the natural environment).
- SOC model was developed by Bak, Tang, and Wiesenfeld (1987), as a possible explanation of noise spectra with $1/f^\alpha$ frequency dependence

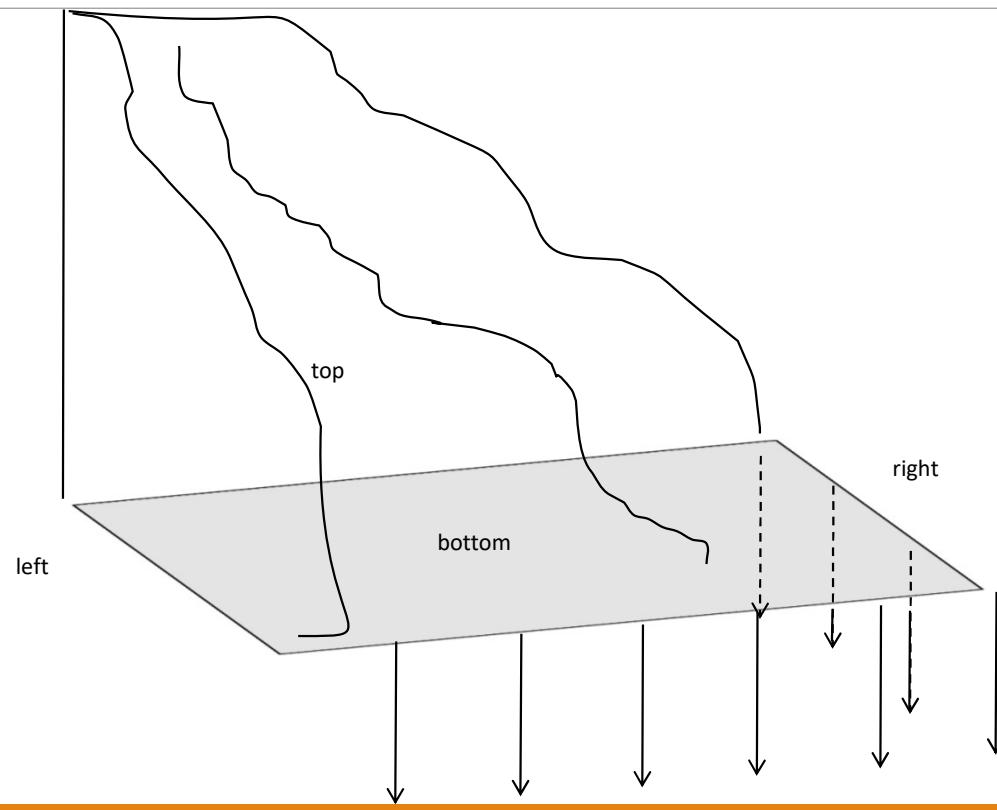


http://en.wikipedia.org/wiki/File:Pink_noise_spectrum.png

Non Linear Dynamics

- In the classic SOC model, sand grains are numerically added to a rectangular lattice building up the slope of a pile.
- When the sand pile is critically poised, the addition of a single sand grain can result in a domino effect where anywhere from a few grains to a large fraction of the sand pile may be shifted.
- SOC systems evolve into a critical state exhibiting avalanches of all sizes without the requirement of finely tuned parameters.

Sandpile Example



Sandpile Example

- The SOC sand pile model consists of a two-dimensional grid with an integer number z_{ij} specified at each grid location (i,j) .
- The ‘idealized sand’ is stacked randomly at each site until the slope at a given site exceeds a threshold value (i.e. $z_{ij} > z_{\text{threshold}}$; e.g. $z_{ij} = 4$ and $z_{\text{threshold}} = 3$)
- Sand is then redistributed to the nearest neighbors as such:
 - 1) first decrease the slope of the site (i,j) by four: $z_{ij} \rightarrow (z_{ij} - 4)$
 - 2) The slope of the four nearest neighbors is then increased by one: $z_{i, j \pm 1} \rightarrow z_{i, j \pm 1} + 1$

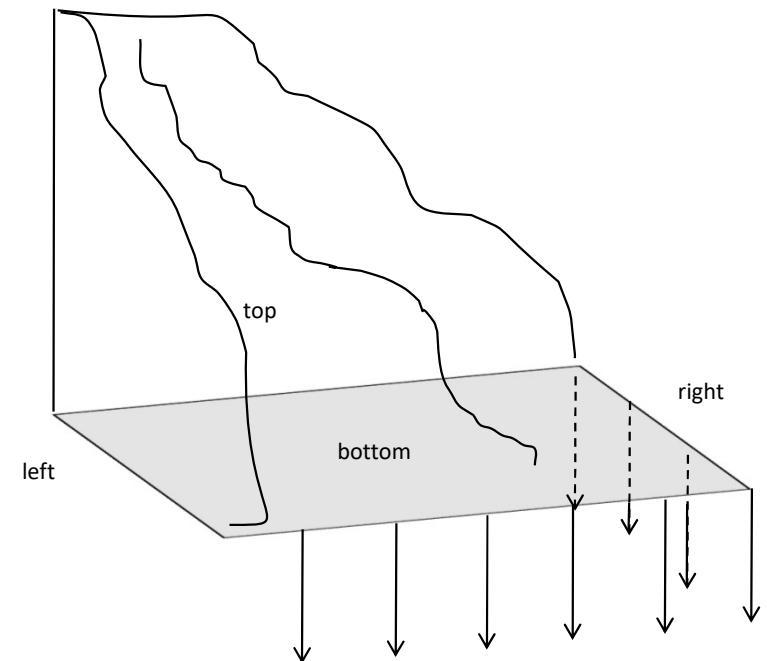
Sandpile Example

- This may cause one or more neighboring sites to exceed the threshold, resulting in further rearrangements until all the sites are less than zthreshold.
- The size of the avalanche is given by the total number N sites toppled
- Avalanche duration is given by the total number of time steps the avalanche propagates.
- Here the resulting power law in avalanche size distribution is:

$$N(s) \propto s^{-\tau} \text{ with } \tau \approx 1.1$$

Boundary conditions of the sand pile model

- 1) Top and left edges are constrained to be zero corresponding to a zero slope of the sand pile at its apex.
- 2) If a site exceeds threshold on the bottom or the right boundary, then 4 is subtracted from every site on these two boundaries.
- 3) This simulates sand falling off the edge as indicated by the arrows in the figure.



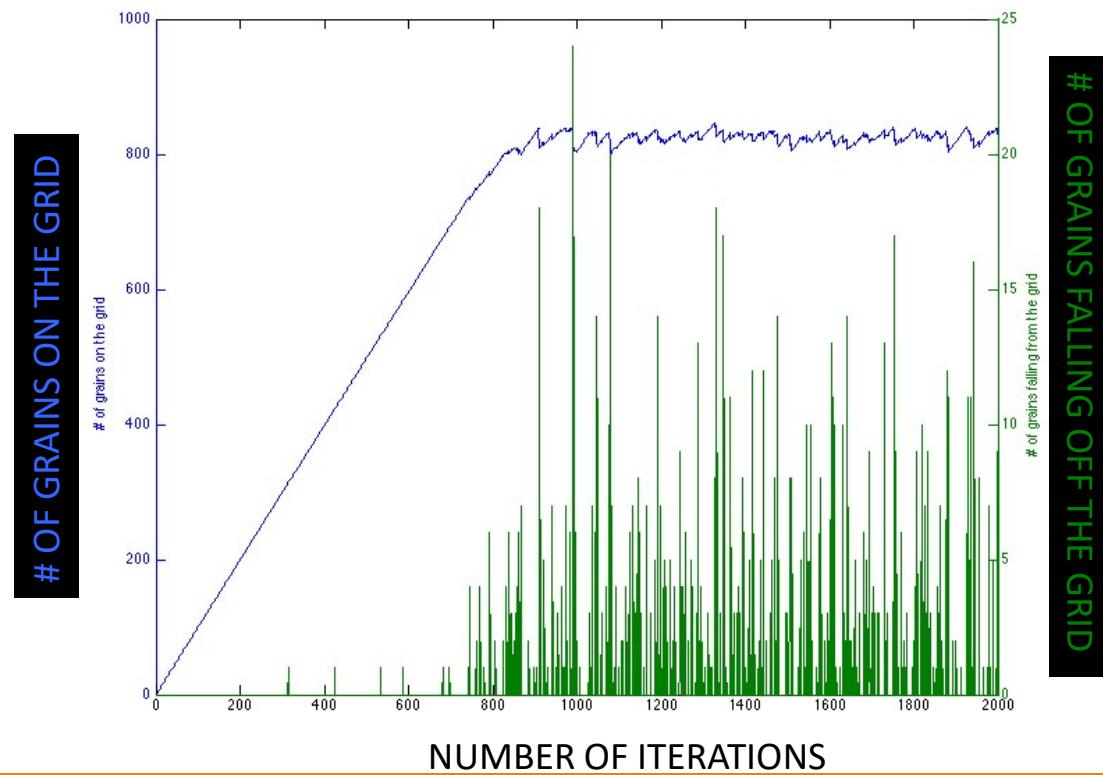
matlab code: sandpile.m

- size of grid matrix is given by siz (e.g. [20 20]) and nrsteps is a scalar positive integer defining the number of time steps of the model run.
- Each time step a sandgrain is added to a random location on the grid.
- When the critical number of grains in a cell exceeds 3 all grains in the cell are turned to the 4 neighbors (von Neumann Neighborhood).

This avalanche may prograde and trigger even more avalanches.

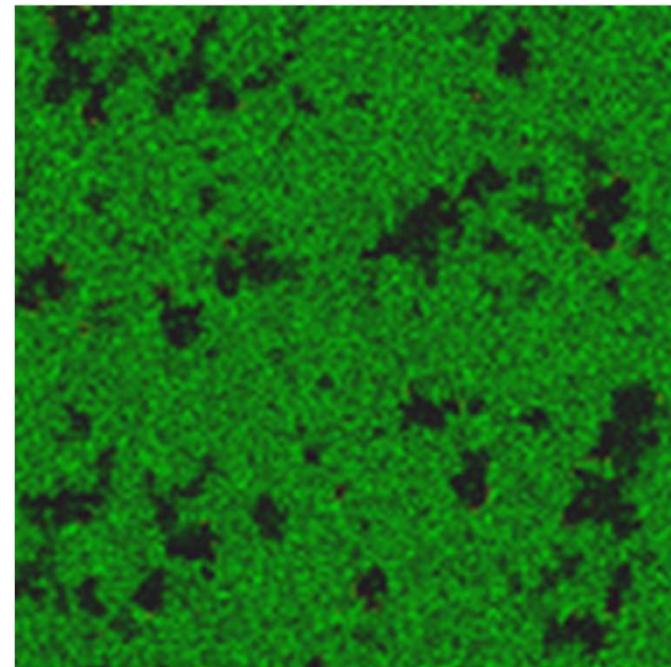
- After a while the sandpile comes to a state of self-organized criticality (SOC).

Matlab code: sandpile.m

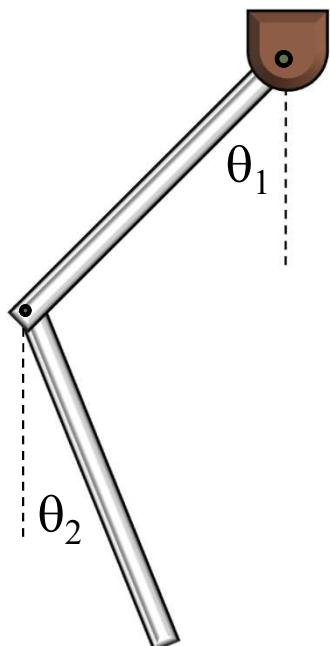


Forest Fire Model

1. A burning cell turns into an empty cell
2. A tree will burn if at least one neighbor is burning
3. A tree ignites with probability f even if no neighbor is burning
4. An empty space fills with a tree with probability p



Chaotic Systems



Simple double pendulum consisting of two connected rods that are free to pivot at their upper ends.

This system will exhibit chaotic oscillations when set in motion with sufficiently large initial values of q_1 and q_2 .

- Chaotic systems are often described by systems of nonlinear differential equations that do not have analytical solutions.
- Physical processes that are purely random (e.g. radioactive decay) are not chaotic.
- In biology systems, complex oscillations can occur in cellular metabolism, population dynamics, heart rhythms, nerve impulses.

Check out matlab code: [double_pendulum_init.m](#)

Chaos final notes

- Chaos occurs in systems whose time evolution is described by nonlinear differential equations.
- BUT: Nonlinearity does not necessarily imply chaotic behavior!
- Chaos never occurs in linear systems or systems with an analytical solution.

Lecture 10

TAYLOR DEVET MASC.

PHD. CANDIDATE BIOLOGICAL AND BIOMEDICAL ENGINEERING

MCGILL UNIVERSITY

SHRINERS HOSPITAL FOR CHILDREN

Todays Aims...



Phase Space



Attractors



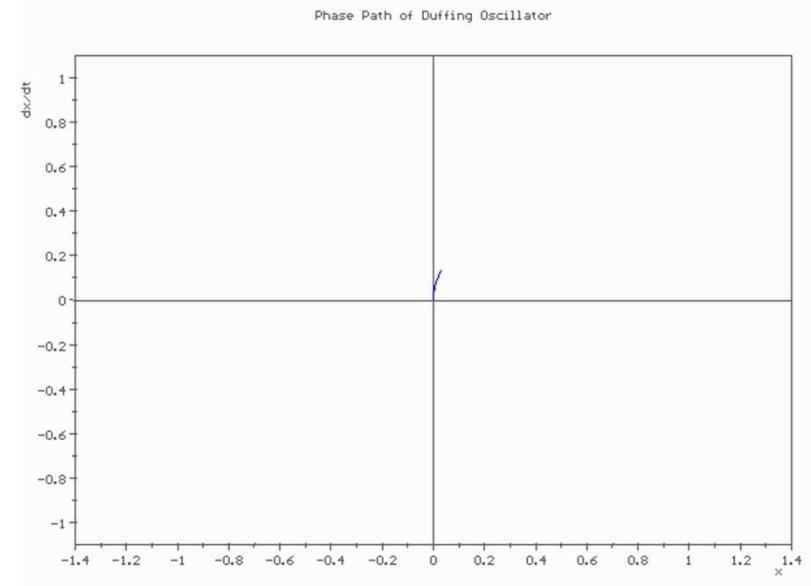
Machine Learning

Other Features of Chaos – Phase Space

- If there are n-variables, then the state of the system can be described as a point in an n-dimensional space whose coordinates are the values of the dynamical variables.

Representations of all possible values a system can take on

- values of variables evolve in time (coordinates of the point in phase space move). Thus can analyze time behavior of the system by analyzing the motion of this point that represents the values of the variables.



Phase Space

- analyze properties of a dynamical system by determining the topological properties of the phase space trajectory

Late 1800s

Henri Poincaré

- Discovered deterministic chaotic systems



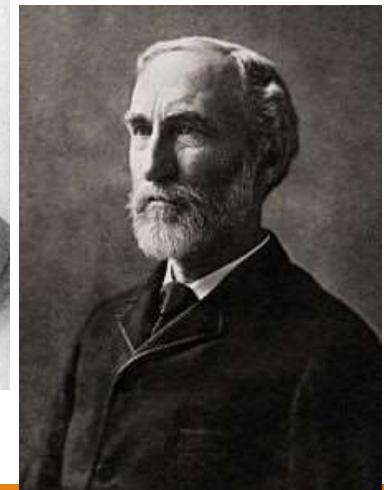
Ludwig Boltzmann

- Studied thermodynamics
- Boltzmann constant



Josiah Willard Gibbs

- Studied thermodynamics
- Gibbs free energy

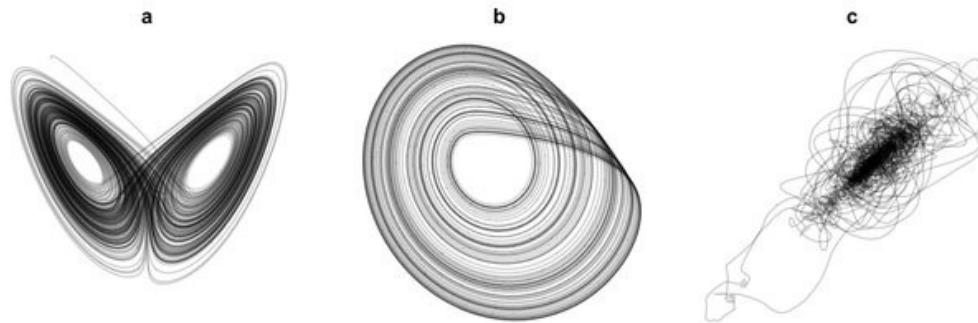


Other Features of Chaos – Attractors

- The point in phase space that represents how the system moves as the values of the variables evolve in time.

If the initial state of the system is not on the attractor, then the phase space point moves exponentially rapidly toward the attractor as time goes by.

Due to sensitivity to initial conditions, two points on the attractor diverge exponentially fast from each other as time goes by, even though they both remain on the attractor.



Other Features of Chaos – Strange attractors

- An attractor is typically finite in phase space.
- Sensitivity to initial conditions
- however, cannot diverge forever

Thus, trajectories from nearby initial points on the attractor diverge and are folded back onto the attractor, diverge and are folded back, etc.

- attractor structure consists of many fine layers.
- Thus, the attractor is fractal.

Other Features of Chaos – Strange and chaotic

"chaotic" = sensitivity to initial conditions

"strange" = fractal attractor

The typical chaotic system is therefore chaotic and strange

- there are also chaotic systems that are not strange - there are non-chaotic systems that are strange

Other Features of Chaos – Dimension of the attractor

The fractal dimension (FD) of the attractor is related to the number of independent variables needed to generate the time series of the values of the variables.

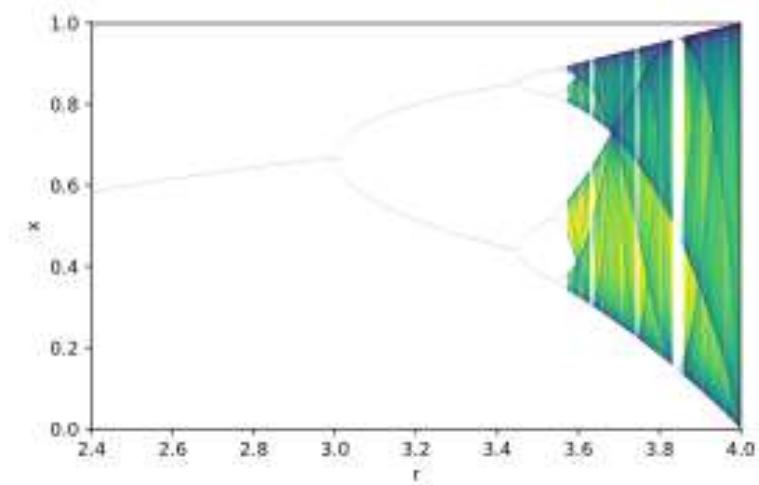
d is the smallest integer greater than the fractal dimension of the attractor

- the time series can be generated by a set of d differential equations with d independent variables.
- e.g. $FD_{\text{attractor}} = 2.03$
- the time series of the values of the variables can be generated by 3 independent variables in 3 coupled nonlinear differential equations.

Other Features of Chaos - Bifurcations

A bifurcation occurs when the form of the trajectory in phase space changes because a parameter passes through a certain threshold value.

- often useful to plot how the form of the dynamics depends on the value of a parameter.



Other Features of Chaos - Control

Linear systems

- small changes in the parameters added to the values of the variables produce small changes in subsequent values of the variables.

Nonlinear systems

- small changes in the parameters added to the values of the variables can produce enormous changes in subsequent values of the variables because of the sensitivity to initial conditions.

Thus, in principle, a chaotic system can be controlled faster and finer and requires smaller amounts of energy for such control than a linear system.

How to analyze real data for Chaotic Dynamics

1. Calculate the FD of the temporal domain (i.e. time data), or frequency data
2. Transform the time series into a geometric object in phase space and then analyze the topological properties of this set
 - i.e. instead of using fractals to analyze the values of the time series itself, use fractals to analyze a representation of the time series in a phase space.

FD of the phase space set is different than the fractal dimension of the values of the time series itself, and it conveys different information about the nature of the process that produced the data.

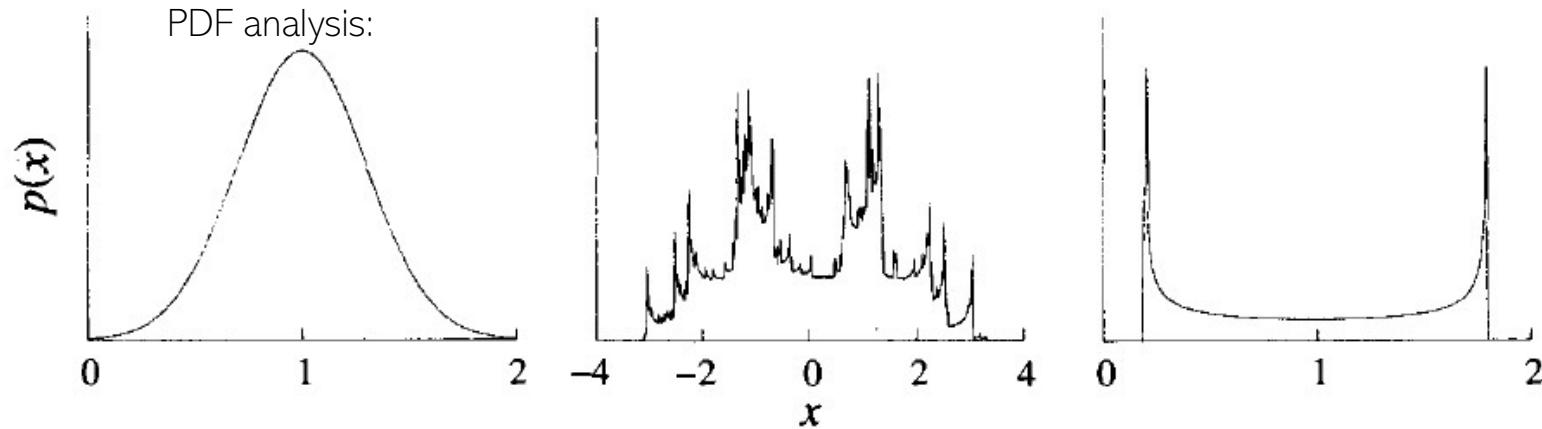
Chaos vs Noise?

A general principle to keep in mind when trying to make the distinction between chaos and noise:

- A very high order system will not be distinguishable from noise unless the system is so dominated by a few variables that it behaves as a low dimensional attractor

Methods for Determining If Signals Are Fractal

1. Power spectral analysis.
2. Autocorrelation function.
3. Dispersion analysis (i.e. RD).
4. Standard statistical measures (e.g. PDF analysis).



Methods suggestive of underlying chaos:

1. Visual inspection for irregularity and bifurcations in periodicities
2. The power spectral density
3. Autocorrelation function

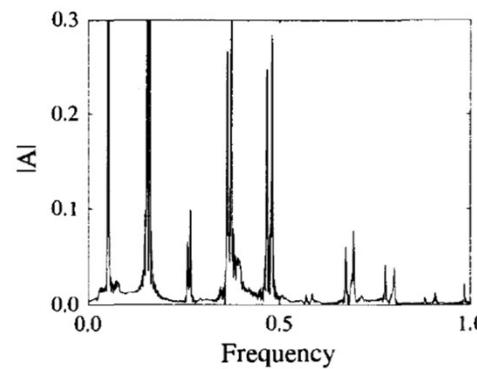
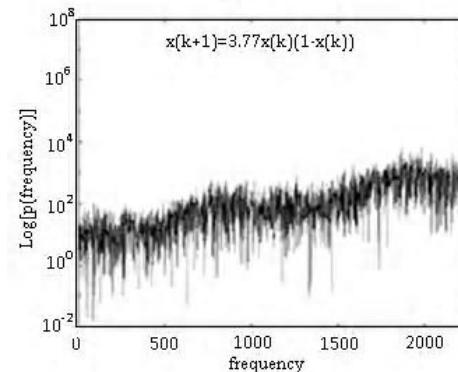
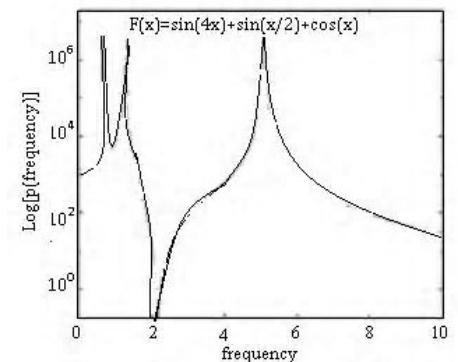
Methods suggestive of underlying chaos: Visual and Power

1. "Visual" methods depend on the appearance of the signal to give:

- an impression of irregularity in rhythm or amplitude,
- the absence of periodic or repeating segments of the same form

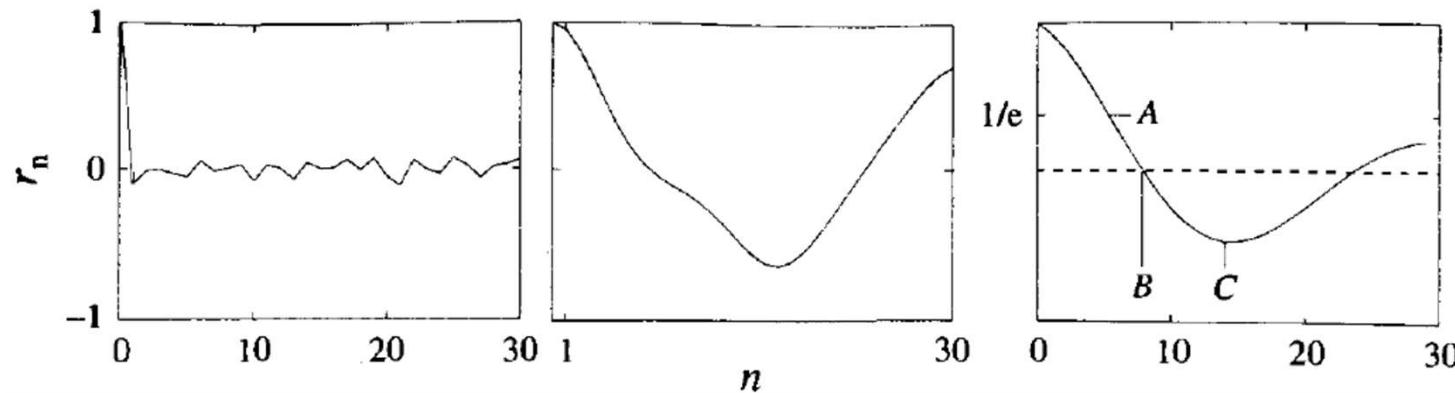
2. Power spectral analysis with FFT

- a chaotic signal commonly shows broadly spread peaks in the power spectrum.



Methods suggestive of underlying chaos: Autocorrelation

- positive correlation over even a short range distinguishes the chaotic signal from random noise even though it will not necessarily distinguish chaos from filtered or smoothed noise.
- noise function shows no correlation even over short times (i.e. $r_n=0$).
- spring function shows only short range correlation, and then some periodicity in the correlation function.
- the fractal correlation also provides an estimate of the time lag to be used in the embedding to create a pseudo-phase space reconstruction of the signal to see if it has the features of a chaotic attractor (right).



Methods suggestive of underlying chaos: Autocorrelation

Typical correlation coefficient:

$$r = \frac{Cov(Y_i, Y_j)}{Var(Y)}$$

Y_i and Y_j are subsets of observations of Y

Autocorrelation Function:

$$r_n = \frac{N}{N-n} \frac{\sum_{i=1}^{i=N-n} Y_i Y_{i+n} - \left(\sum_{i=1}^{i=N} Y_i \right)^2}{\sum_{i=1}^{i=N} Y_i^2 - \frac{\left(\sum_{i=1}^{i=N} Y_i \right)^2}{N}}$$

(i.e. autocovariance / variance)

Special methods for characterizing low dimensional chaotic signals:

1. Visual analysis of phase plane plots
2. Correaltion, and information dimension
3. Estimation of Lyapunov exponents
4. Calculations of entropy

Special situations

1. Observations of changing states exhibiting bifurcations or changes of periodicities
2. Interventions resulting in period doublings or changes in apparent chaotic state

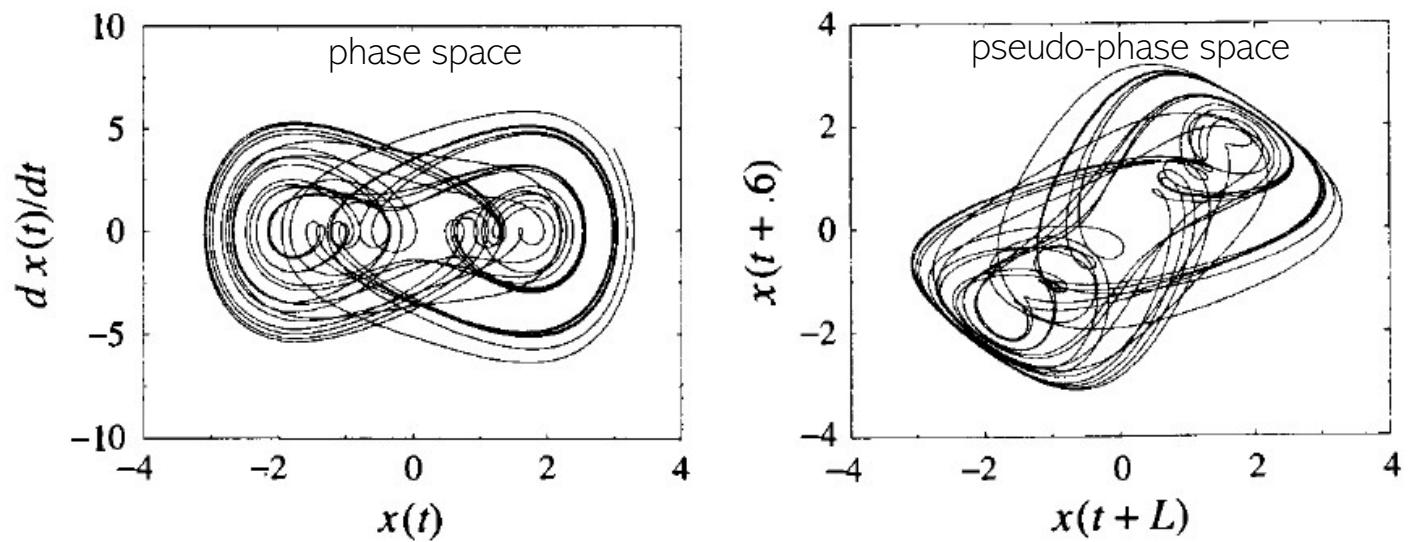
Pseudo-phase space plots

Embedding: Turning the time series into a set in pseudo-phase space.

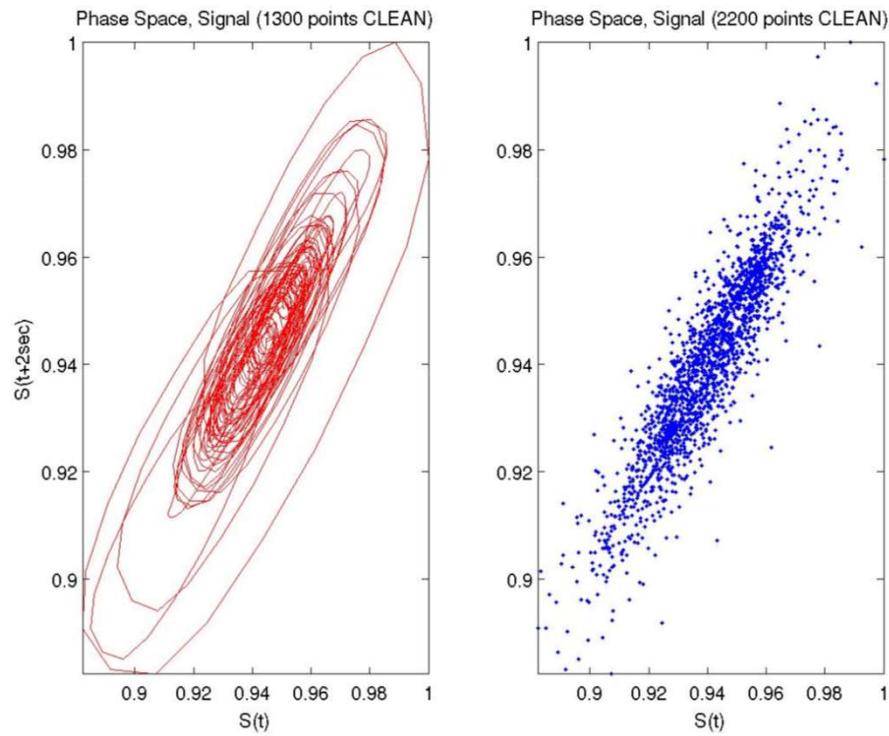
Phase Space: consider 3 independent variables can plot variable X vs. Y OR plot of derivative of a variable vs. variable (left).

Pseudo-phase space:

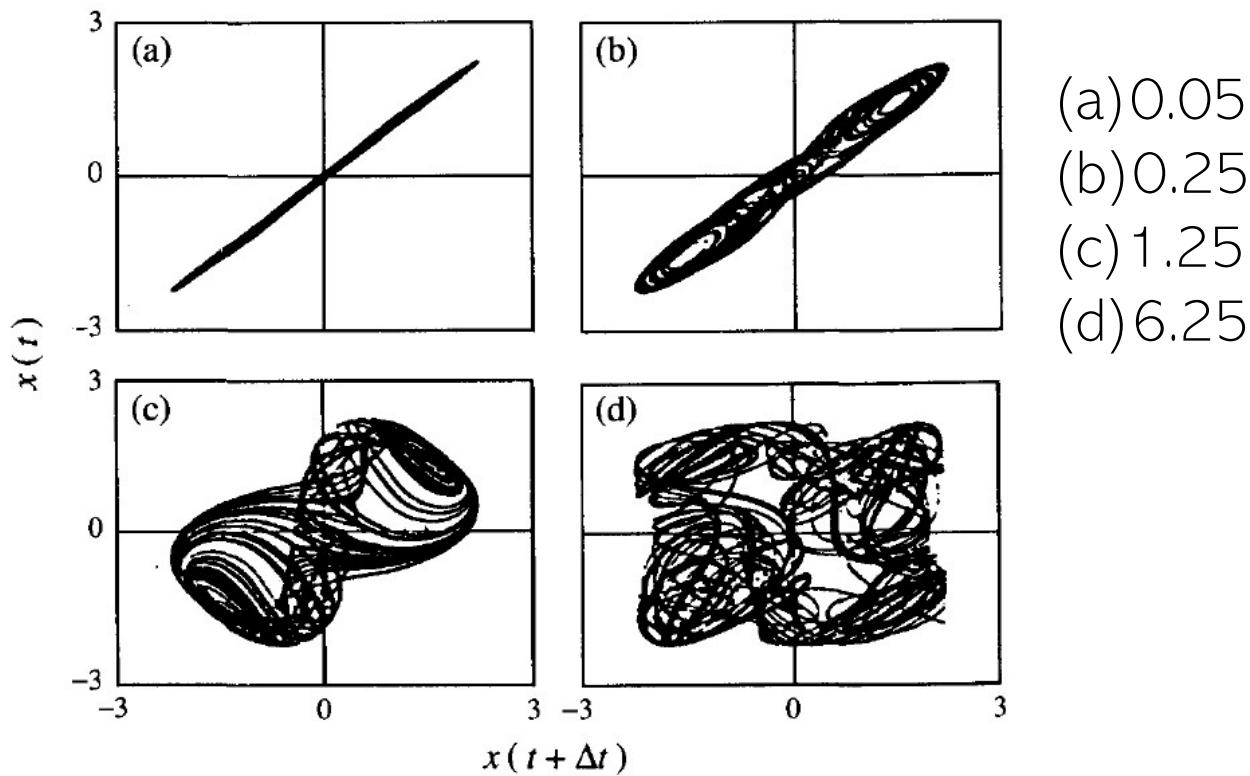
plot variable+lag vs.
variable



Pseudo-phase space plots



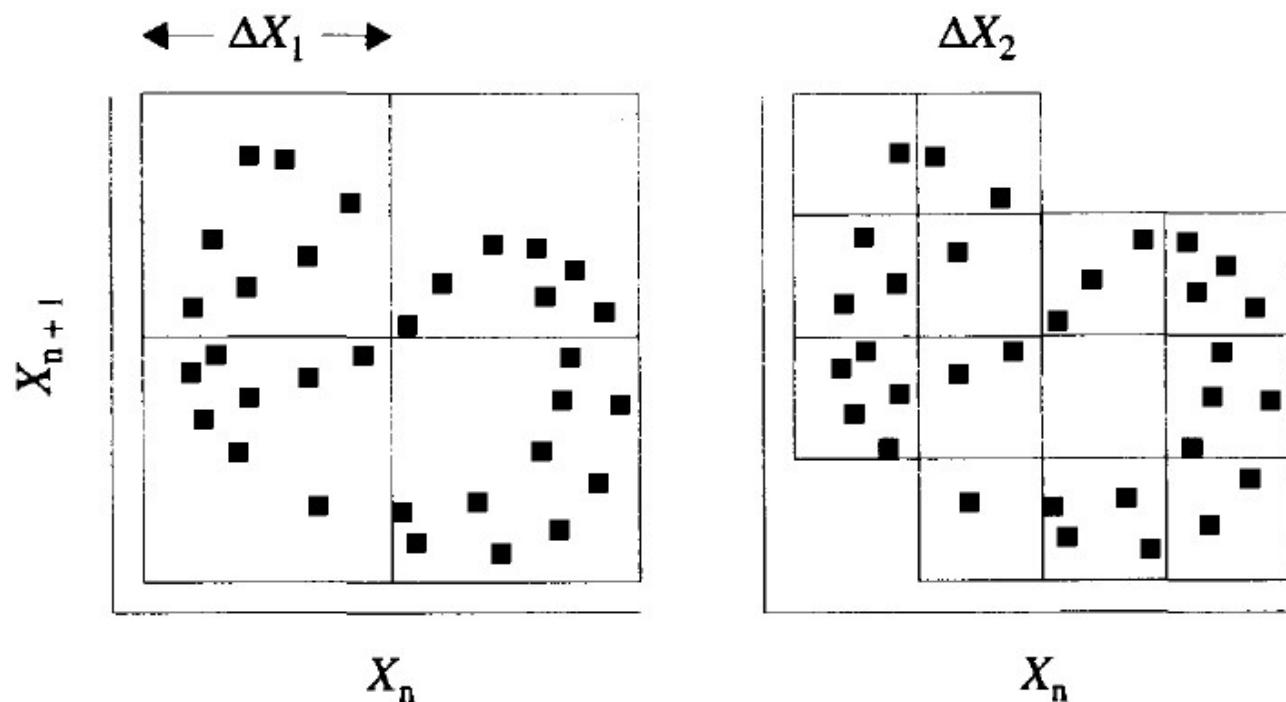
Reconstructions with lags



Box counting:

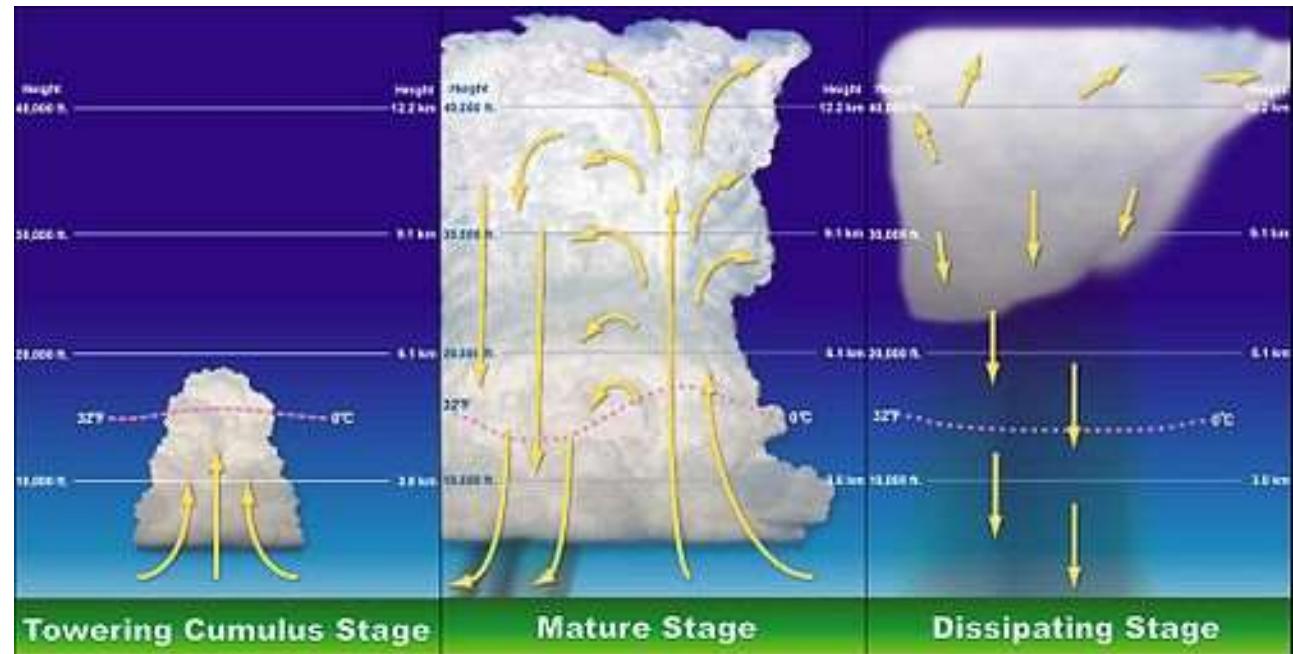
Topology of determining fractal dimension of phase space

Box counting to determine FD of 2D space. The # of boxes $N(r)$ occupied by at least one point of the set is determined for boxes of ever smaller size r . The fractal dimension is the slope of $\log N(r)$ versus $\log(1/r)$.



Lorenz Model

A famous model exhibiting chaotic behavior was first developed in 1963 by Edward Lorenz to describe complex atmospheric convection.



Lorenz Model

The Lorenz equations represent an extreme simplification of the Navier-Stokes equations describing fluid flow between two boundaries held at different temperatures representing the Earth's surface and the upper atmosphere.

The Lorenz model consists of the set of first-order, autonomous, nonlinear differential equations:

σ includes fluid velocity and thermal conductivity (Prandtl number).

r = Rayleigh number

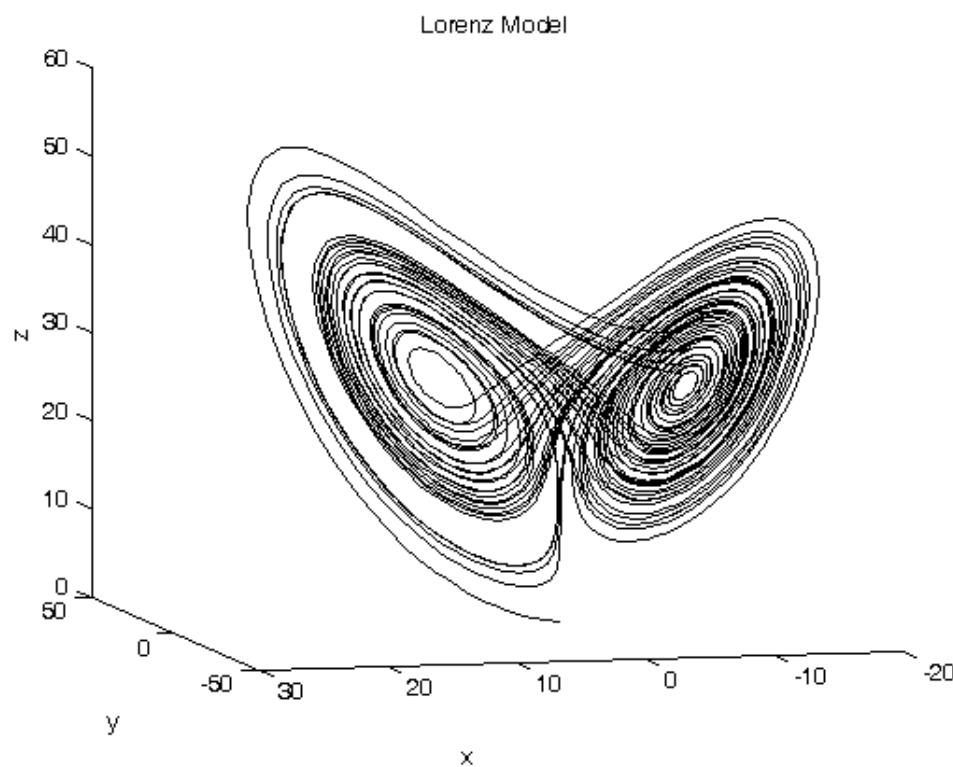
b = geometry dependent factor

$$\frac{dx}{dt} = -\sigma(x + y)$$

$$\frac{dy}{dt} = -xz + rx - y$$

$$\frac{dz}{dt} = xy - bz$$

Lorenz Model



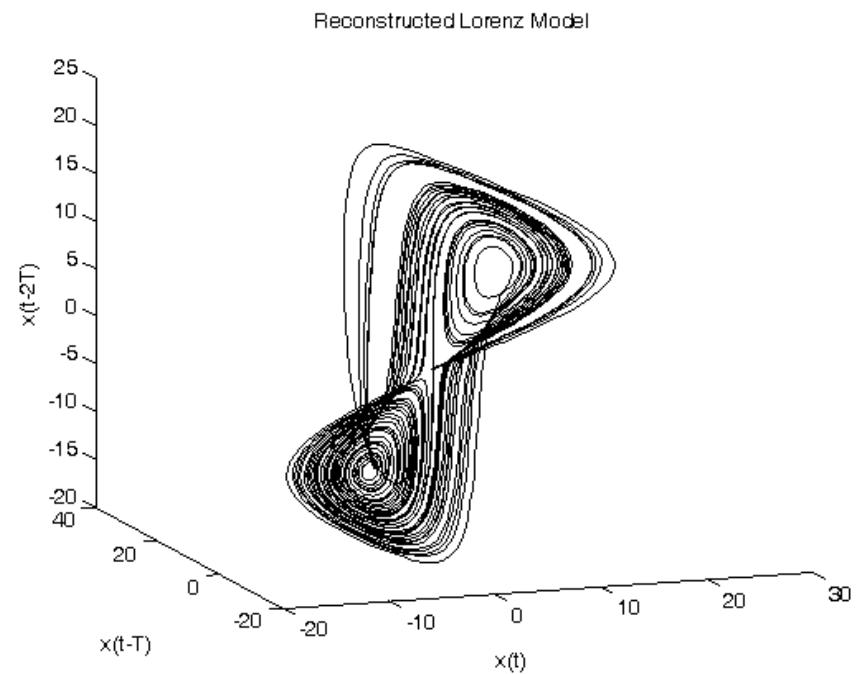
The solution of this system generates a three-dimensional phase-space orbit for
 $\sigma=10$
 $r=28$
 $b=8/3$

Try Matlab code: [Lorenz_reconstruct.m](#)

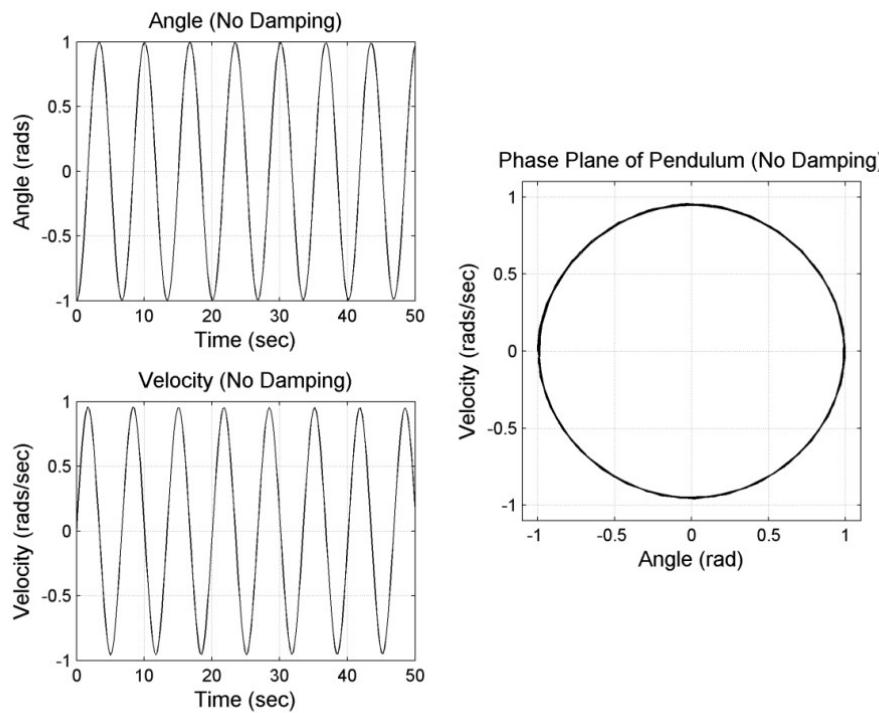
Lorenz Attractor

- The reconstruction is performed in two dimensions by plotting $x(t)$ vs. $x(t-\tau)$ where τ is a suitably chosen time delay.

Lorenz attractor reconstructed from the x time series using time delay coordinates.

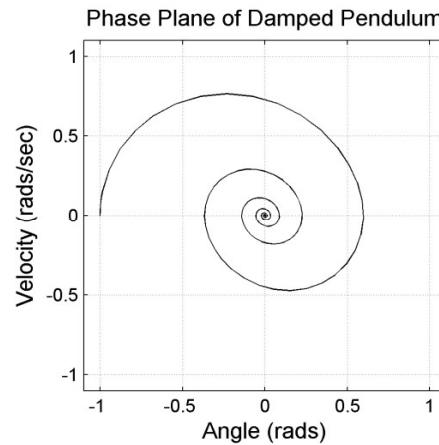
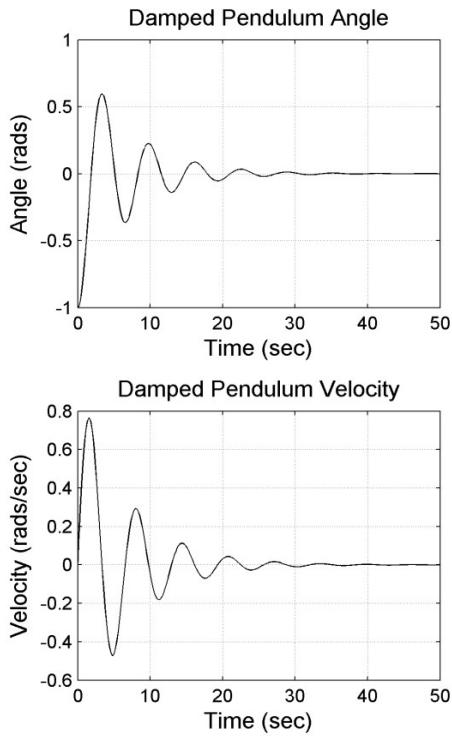


Visualizing the state-space with a phase-plane plot



- A plot of the state variables against each other
- If the phase-space has 2 dimensions, a plot is known as the phase-plane
- Depicts the attractor of the system
- The attractor is the tendency of the system at equilibrium. In this example it is a limit cycle

Visualizing the state-space with a phase-plane plot



- Pendulum with damping
- Here the attractor is a stable node

Phase-space plots of measured data

- The phase-space can give us important information about a system (from its signals)
- Nonlinear systems can have complicated phase-space plots
- However, we often do not have access to all the variables.
 - For example, body temperature is one variable of the inflammation response but others are invasive
- Is there a way to recover, or at least estimate the phase-space of a multidimensional system from a phase-space plot?

Answer: Yes, With Delay Embedding

Floris Takens showed that the method of delayed embedding sufficiently reconstructs a time series

Reconstruction is just an estimate, not perfect

Method

- Use a delayed version of the signal as a “new” dimensional measurement time series. Each additional series is delayed by a multiple of the chosen delay τ

Formally we would say time series $x[n]$ of length N can be reconstructed into multidimensional time series $y[n_d, k]$ of k dimensions from 1 to m , where each delayed vector n_d comes from $x[n]$ delayed by τ

$$y[n_d, k] = x[n + (k - 1)\tau, k] \dots x[N - (m - 1)\tau, m],$$



With MATLAB

- The matlab routine **delay_emb** is provided to perform the operation of delay embedding. The function has a format:

y=delay_emb (x,m,tau)

- Inputs
 - **x**: the original one-dimensional time series
 - **m**: the embedding dimension of new m -dimensional vector **y**
 - **tau**: the delay
- Outputs
 - **y**: a matrix of vectors representing the system in embedded space

Getting a good embedding

Takens showed that

- m should be twice D where D is the “true” dimension of the system
 - True meaning if we had access to each dimension
- Theoretically τ could be any number

However in practice

- m just needs to be “sufficiently” large
- τ can’t be too large or too small
 - Each delayed vector should have some correlation to the previous vector, but not too much

Lorenz Equations

Before continuing, we revisit the Lorenz system

- useful for demonstrating concepts in the rest of the lecture.

Lorenz system

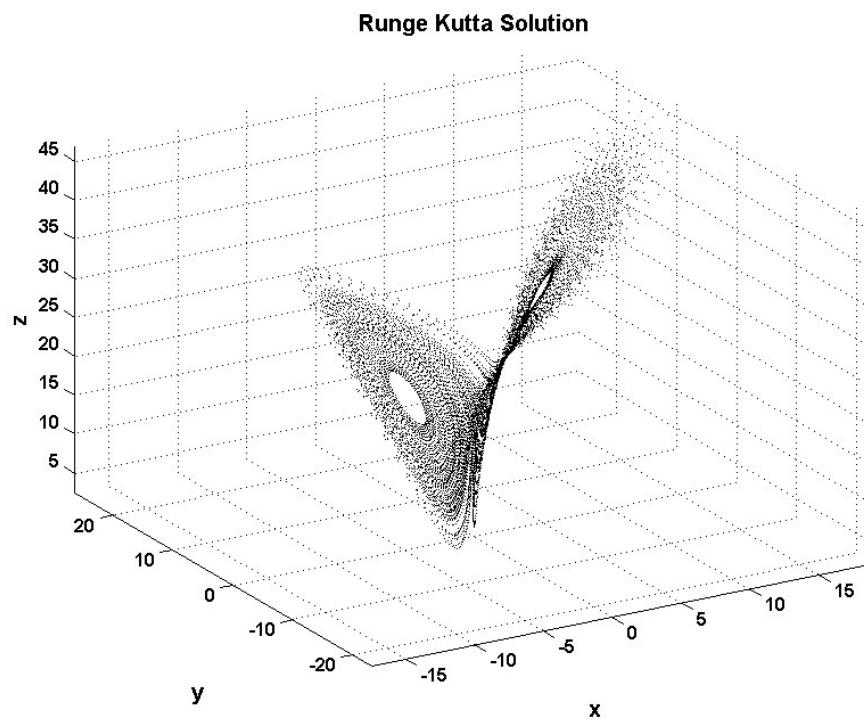
- Simplified fluid undergoing convection (the movement of fluid due to a heat source)
- Here x, y and z are the spatial dimensions
- The constants
 - σ , the Prandtl number, the ratio of the viscosity of the fluid to the thermal conductivity
 - β , the Rayleigh number, a dimensionless number which represents a ratio of buoyant forces to viscous forces within the flow of a fluid
 - ρ , the density of the fluid.

$$\frac{dx}{dt} = \sigma(y - x)$$

$$\frac{dy}{dt} = x(\rho - z) - y$$

$$\frac{dz}{dt} = xy - z\beta$$

Lorenz System Phase Space



Methods of finding the delay

There is more leeway in choosing the best delay compared to choosing the best value for m , a range of delay values may be fine.

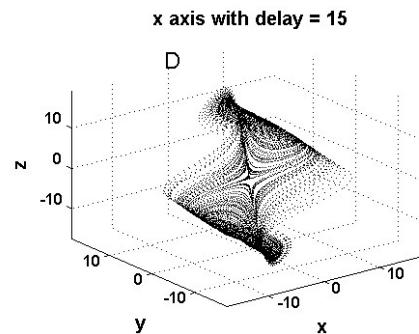
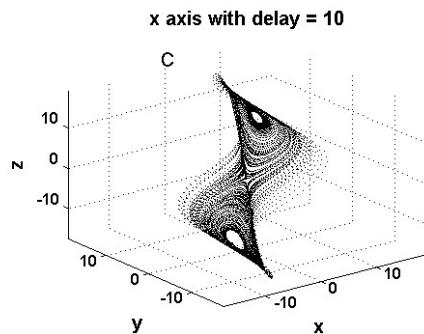
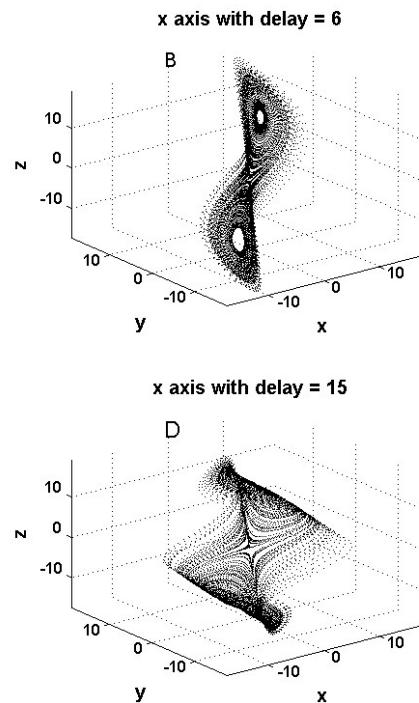
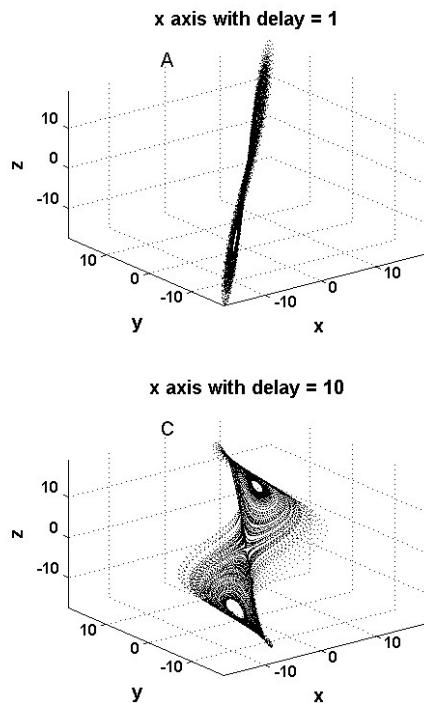
Trial and Error

- Simple but it works

Autocorrelation

- Minimum or 0

Estimating the Delay



Methods of finding embedding dimension

- Trial and error
- False Nearest neighbors
- Principle component analysis

False Nearest Neighbors

- Points along the trajectory of a properly unfolded phase-space have specific neighbors.
- If unfolded in too few dimensions, the neighbors may be false
 - For example points separated in the Z dimension might appear close together projected to the XY plane
- By examining the ratio of the change in false nearest neighbors, we can estimate a reasonable embedding dimension.

Distance of vectors

- Here we define distance of two vectors using the Euclidian no

$$d = \sqrt{(\vec{x}_a - \vec{x}_b)^2} \quad \text{or} \quad d = \|\vec{x}_a - \vec{x}_b\|$$

- Other definitions, such as the max norm, can also be used
 - Max norm is more computationally efficient

With MATLAB

```
numnear = fnumnear(x,tau,em,r)
```

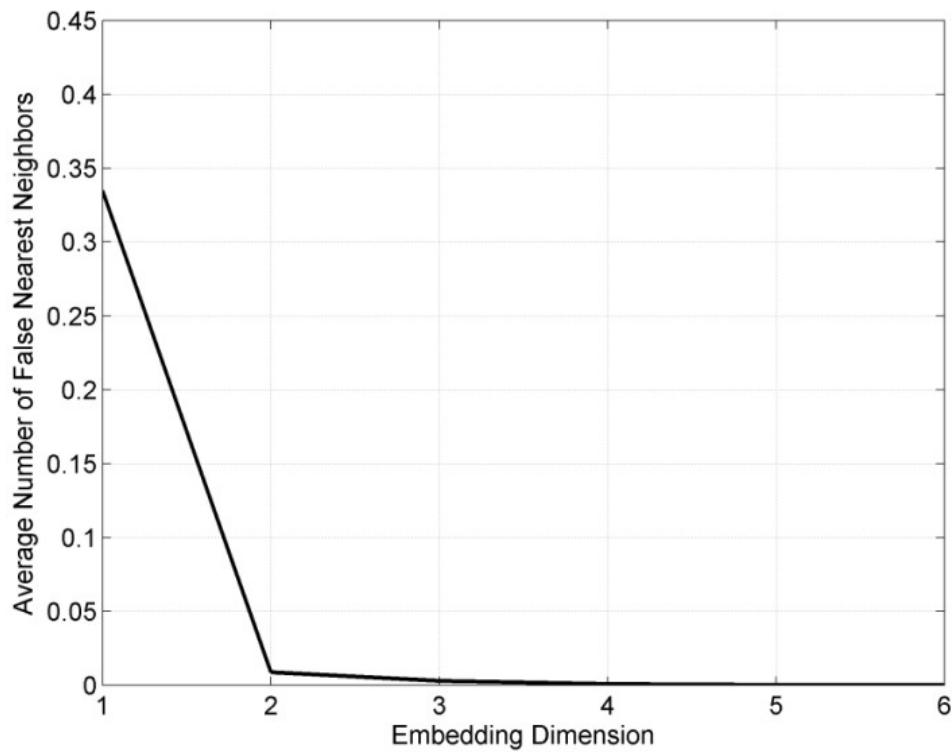
- Inputs

- **x**: input vector (i.e. the signal)
- **tau**: is the delay used for embedding
- **em**: maximum number of dimensions to test
- **r**: the distance that defines nearest neighbors.
 - A range of **r** values should be tested,
 - Good starting point is $0.1 * \text{std}(x)$

- Outputs

- **numnear**: a series of ratio values of the number of neighbors in dimension m to dimension **m+1**

Nearest Neighbors for Lorenz System



We see a large drop off from dimension 1 to 2, and then a small drop off from dimension 2 to 3. After that there is negligible change.

Data Reduction using PCA

The eigenvalues describe how much of the variance is accounted for by the associated principal component and if a component is really necessary.

Eigenvalues are ordered by size; that is:

$$\lambda_1 > \lambda_2 > \lambda_3 \dots > \lambda_M.$$

If an eigenvalue is zero or 'close to' zero, then its associated principal component contributes little to the data and can be eliminated. This component accounts for only a small amount of the variance in the data.

This tells us the effective dimension of the data set.

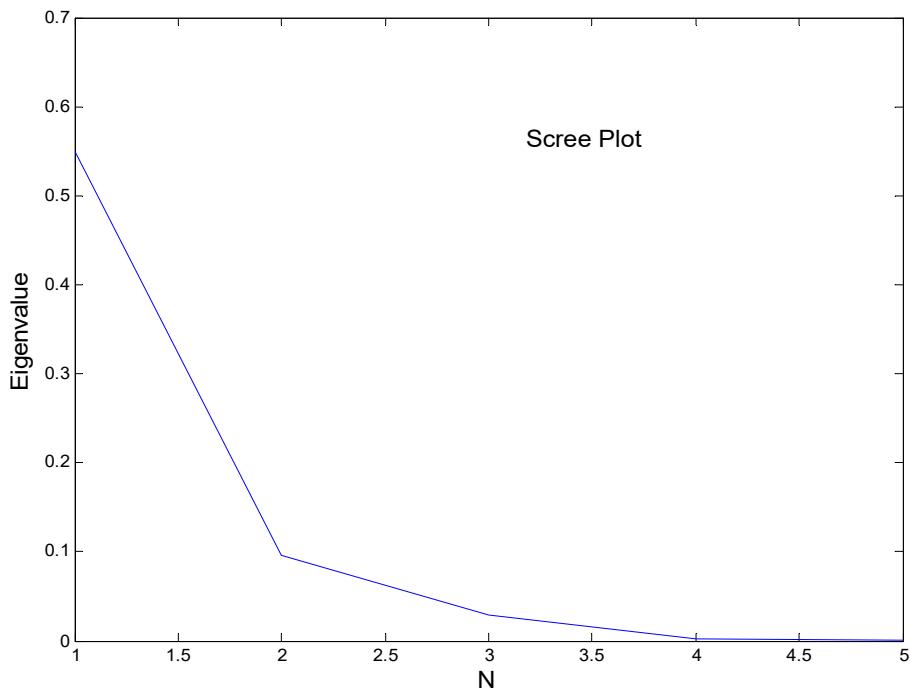
How do you decide if an eigenvalue is small enough so that its associated component can be removed from the data set?

The Scree Plot

The eigenvalues are in order of large to small. Plotting them in order will show their relative value.

Such a plot is called the Scree plot

The actual dimension of the data set is taken where the Scree plot becomes more-or-less flat.

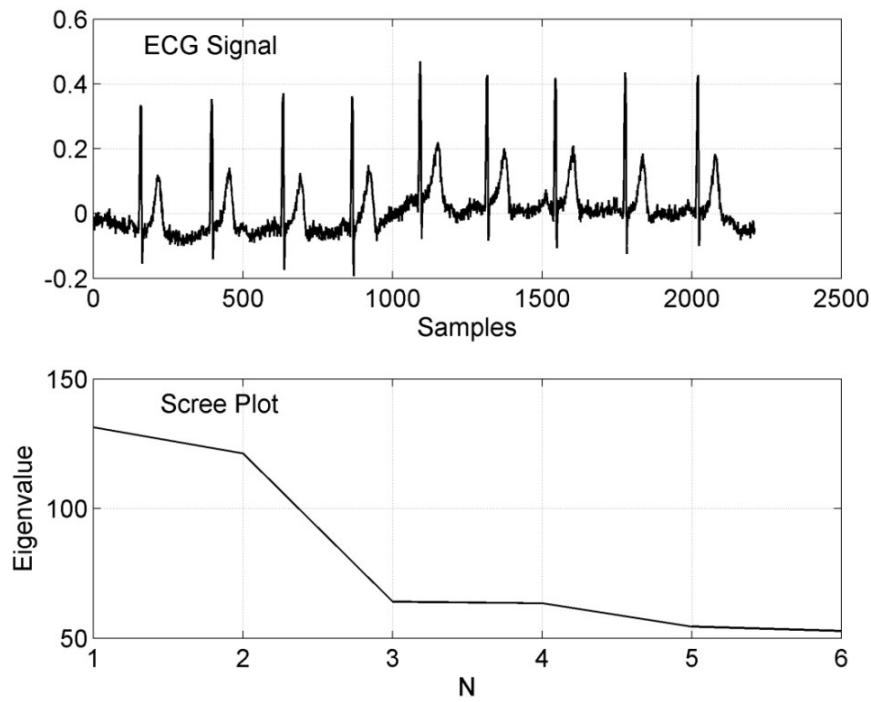


Finding m : PCA

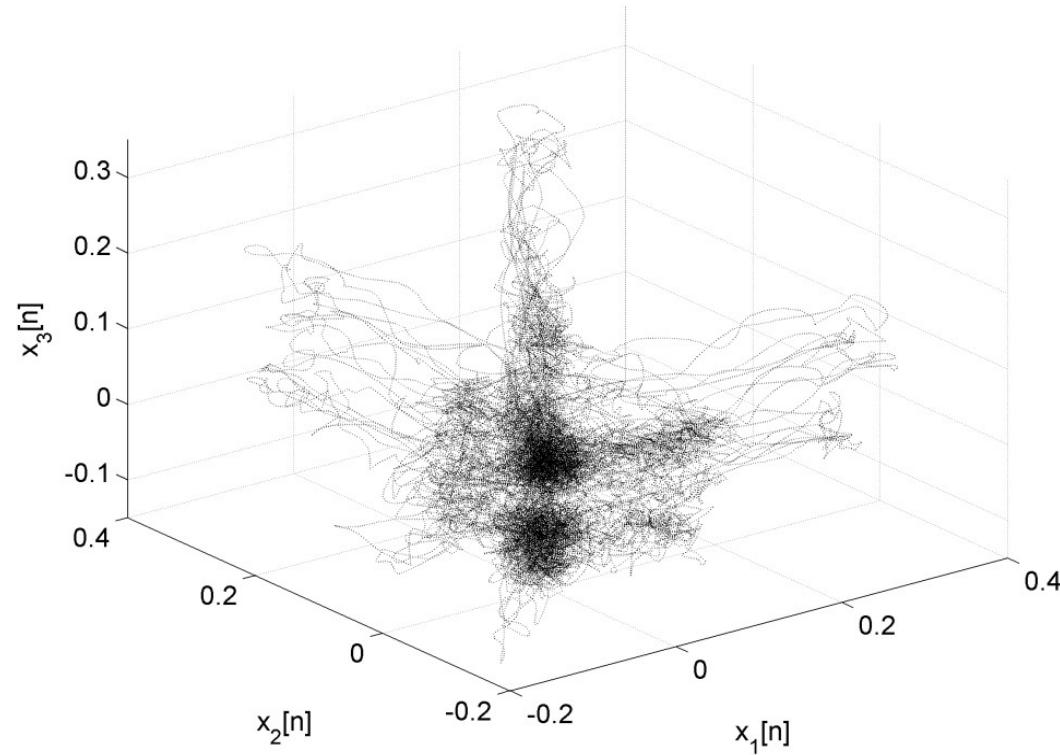
Use principle component analysis to determine the number of principle components and use that as your estimate for m

```
% Determination of embedding dimension using PCA
%
x=load('ECGtest.csv');           % Load the ECG file.
fs = 2100;                      % Sampling frequency
m = 6;                          % Embedding dimension
%
X = delay_emb(x,m,tau);         % Generate delayed signals
[U,S,pc]= svd(X' , 'econ');    % Perform the decomposition
eigen = diag(S).^2;             % Get the eigenvalues
```

Scree plot showing 3 Dimensions



ECG embedded in 3 dimensions



The delay was found using the autocorrelation method and taken as 381 samples

Another nonlinear system: Logistic Map

- A simple nonlinear system is known as the Logistic mapping function, a simplification of a Predator Prey model
- $x[n+1] = r \cdot x[n] \cdot (1 - x[n])$
- Exhibits chaotic behavior if $r > 3.58$
- If $0 < x_0 < 1$, the output is always bounded by 0 and 1
- Phase-plane can be visualized by plotting $x[n+1]$ against $x[n]$

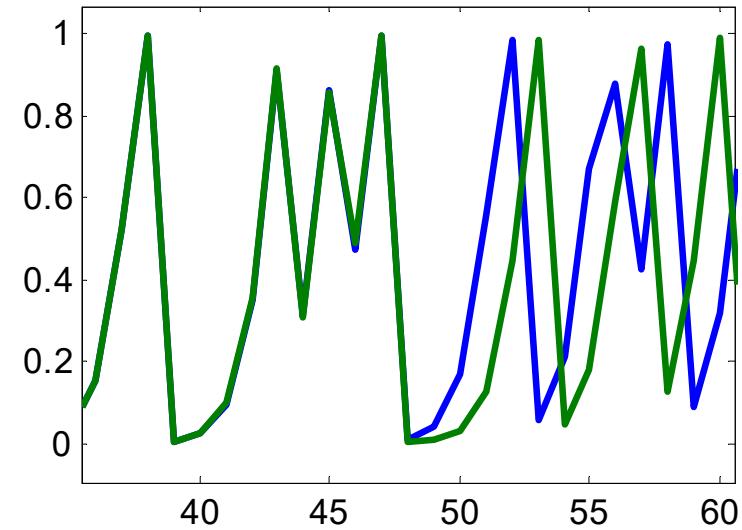
Properties of Chaos

- Exponential Divergence

- Divergence is the phenomenon of trajectories of a system that begin with similar initial conditions ending up with very different trajectories.
 - This is the opposite of convergence, in which systems tend towards the same value over long periods of time
 - While nonchaotic systems may show divergence, only chaotic systems have trajectories that diverge exponentially.

Sensitivity to Initial Conditions

- Since the divergence is exponential, a plot of the log divergence should give a straight line
- The slope of the line gives an estimate of the Lyapunov exponent, a measurement of how quickly the divergence happens



Properties of the Lyapunov Exponent

For a multidimensional system, we measure divergence along the phase-space (or the estimated phase space)

The Lyapunov exponent (λ) is a measure of the divergence

- $\lambda > 0$ signifies chaotic behavior
- $\lambda = 0$ signifies a limit cycle
- $\lambda < 0$ converging trajectories.

A chaotic system does not have a single Lyapunov exponent

- There is a λ for each direction in phase-space: The *Lyapunov spectrum*
- Typically we are interested in the largest exponent or the sum of all exponents, as these give a general description of the system

When we refer to the Lyapunov exponent (symbol λ) we are referring not to the Lyapunov spectrum, but the single largest Lyapunov exponent.

Finding λ with MATLAB

Not always possible to take many measurements at similar initial conditions

Instead use nearest neighbors as substitutes for initial conditions of different trajectories

- provided they are sufficiently separated in time, why?

Since λ is not constant across the phase-space, we need to sample many areas of the phase-space

- The average of all the areas should approach the largest λ

After averaging the log divergence curves we examine the mean curve and look for a linear region

MATLAB Implementation

```
[lambda, s_mean,linear_end] = max_lyp(x,m,tau,fs,radius);
```

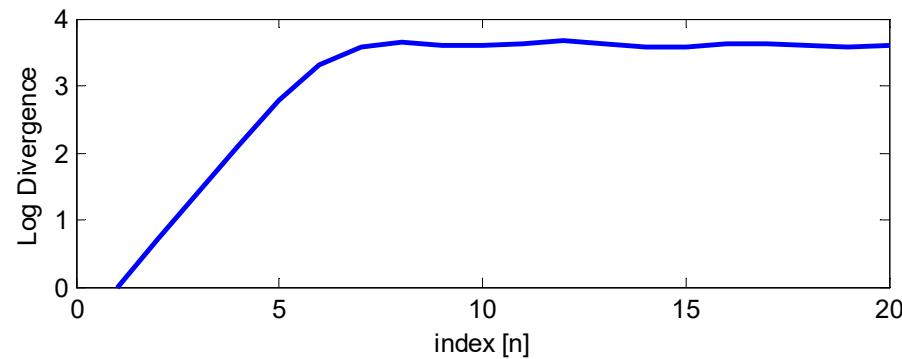
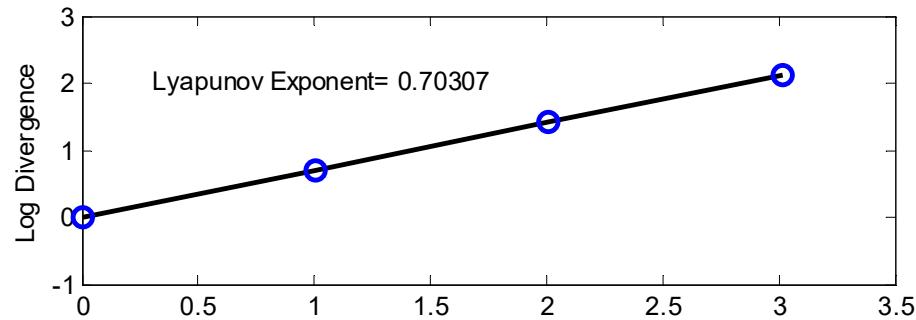
Inputs

- **x**: the time series to be analyzed the embedding
- **m**: and **tau**: the embedding parameters
- **fs**: the sampling frequency
- **radius** : the nearest neighbors cutoff.

Outputs

- **lambda** : the estimate of λ
- **linear_end**: the index of the end of the linear region

max_lyp output for Logistic map

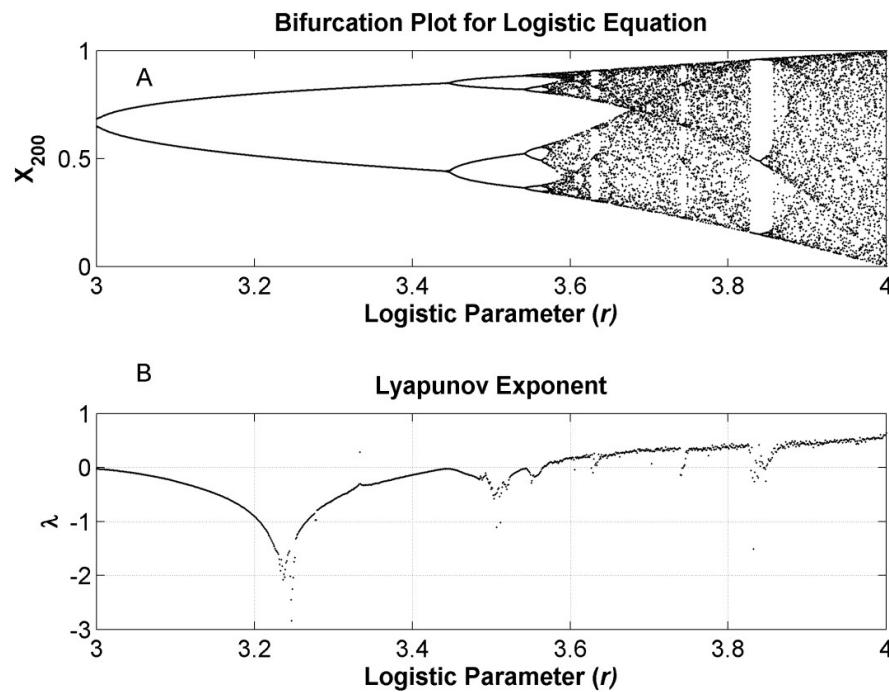


Important

- `max_lyp` will give you an estimate for λ whether it is appropriate or not

You must double check that the divergence is indeed linear

Bifurcation plot of the Logistic Map

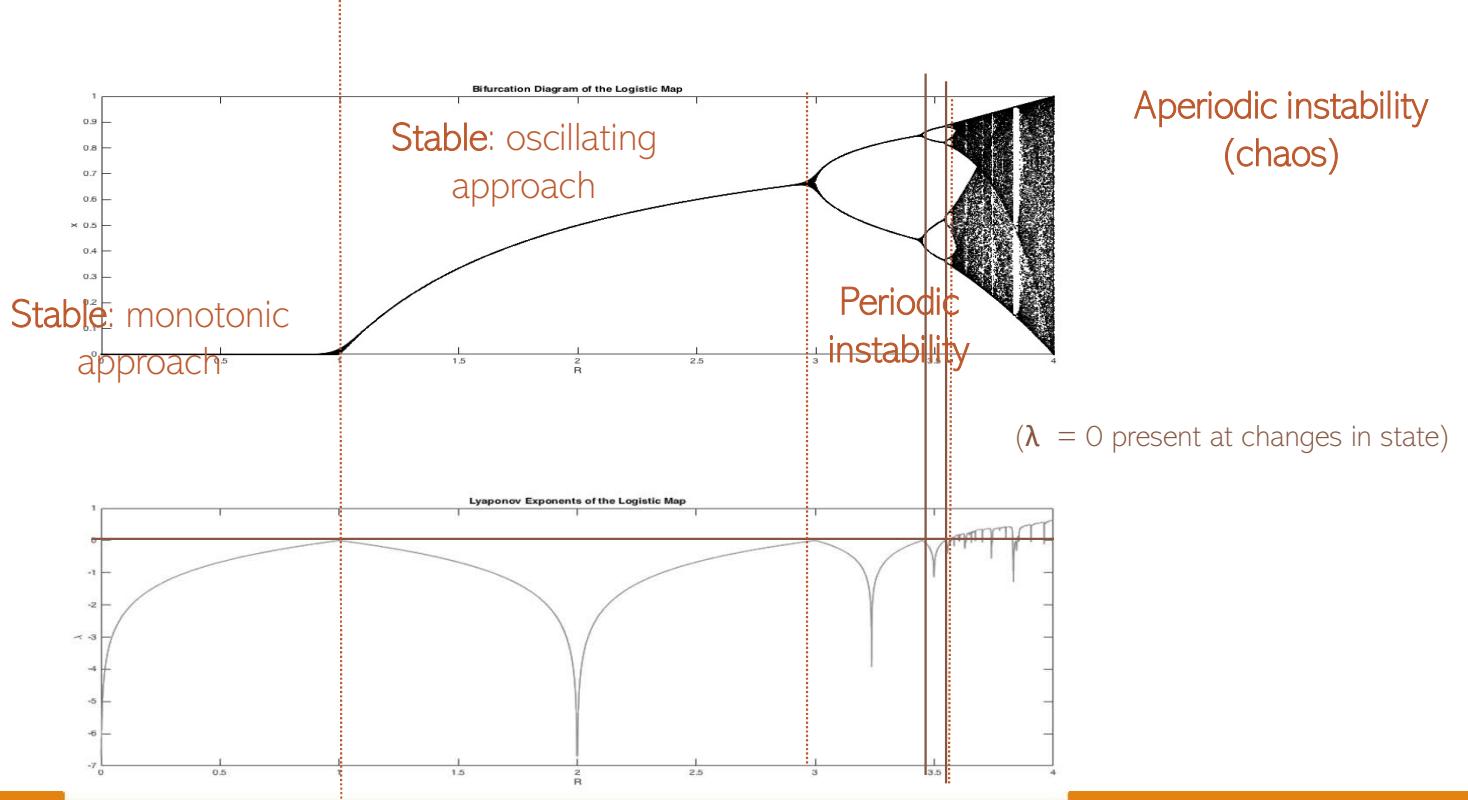


Lyapunov exponent

$\lambda > 0$ signifies chaotic behavior

$\lambda = 0$ signifies a limit cycle

$\lambda < 0$ converging trajectories



Node, Focus, or Saddle point

If the pair of ODEs are NOT LINEAR... you can approximate them using linear equations in the neighbourhood of the fixed point (x^*,y^*)

$$A = \left. \frac{\partial f}{\partial x} \right|_{x^*,y^*} \quad B = \left. \frac{\partial f}{\partial y} \right|_{x^*,y^*}$$
$$C = \left. \frac{\partial g}{\partial x} \right|_{x^*,y^*} \quad D = \left. \frac{\partial g}{\partial y} \right|_{x^*,y^*}.$$

$$\begin{aligned}\frac{dX}{dt} &= AX + BY, \\ \frac{dY}{dt} &= CX + DY.\end{aligned}$$

$f(x,y)$ and $g(x,y)$ are
the non-linear ODEs

Approximated
linear equations

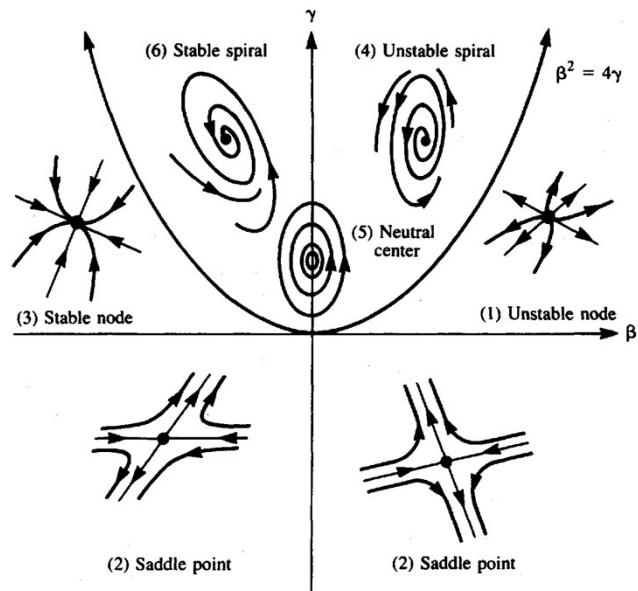


Node, Focus, or Saddle point

Table 5.1 Linear Systems of two ODEs

	Full algebraic notation	Equivalent Vector-Matrix Notation
Equations	$\frac{dx}{dt} = a_{11}x + a_{12}y$ $\frac{dy}{dt} = a_{21}x + a_{22}y$	$\frac{dx}{dt} = \mathbf{Ax}, \quad \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$
Significant quantities	$\beta = a_{11} + a_{22}$, $\gamma = a_{11}a_{22} - a_{12}a_{21}$, $\delta = \beta^2 - 4\gamma$	$\text{Tr } \mathbf{A}$, $\det \mathbf{A}$, $\text{disc } \mathbf{A}$
Characteristic equation	$\lambda^2 - \beta\lambda + \gamma = 0$	$\det(\mathbf{A} - \lambda\mathbf{I}) = 0$
Eigenvalues	$\lambda_{1,2} = \frac{\beta \pm \sqrt{\delta}}{2}$	$\lambda_{1,2} = \frac{\text{Tr } \mathbf{A} \pm \sqrt{\text{disc } \mathbf{A}}}{2}$
Identities	$\lambda_1 + \lambda_2 = \beta$,	$\lambda_1 + \lambda_2 = \text{Tr } \mathbf{A}, \quad \lambda_1\lambda_2 = \det \mathbf{A}$
Eigenvectors	$\begin{pmatrix} a_{12} \\ \lambda_1 - a_{11} \end{pmatrix}, \begin{pmatrix} a_{12} \\ \lambda_2 - a_{11} \end{pmatrix}$	$\mathbf{v}_1, \mathbf{v}_2$ such that $(\mathbf{A} - \lambda_i\mathbf{I})\mathbf{v}_i = 0$
Solutions	$x = c_1 a_{12} e^{\lambda_1 t} + c_2 a_{12} e^{\lambda_2 t}$, $y = d_1 e^{\lambda_1 t} + d_2 e^{\lambda_2 t}$, where $d_1 = c_1(\lambda_1 - a_{11})$, $d_2 = c_2(\lambda_2 - a_{11})$.	$\mathbf{x} = c_1 \mathbf{v}_1 e^{\lambda_1 t} + c_2 \mathbf{v}_2 e^{\lambda_2 t}$.

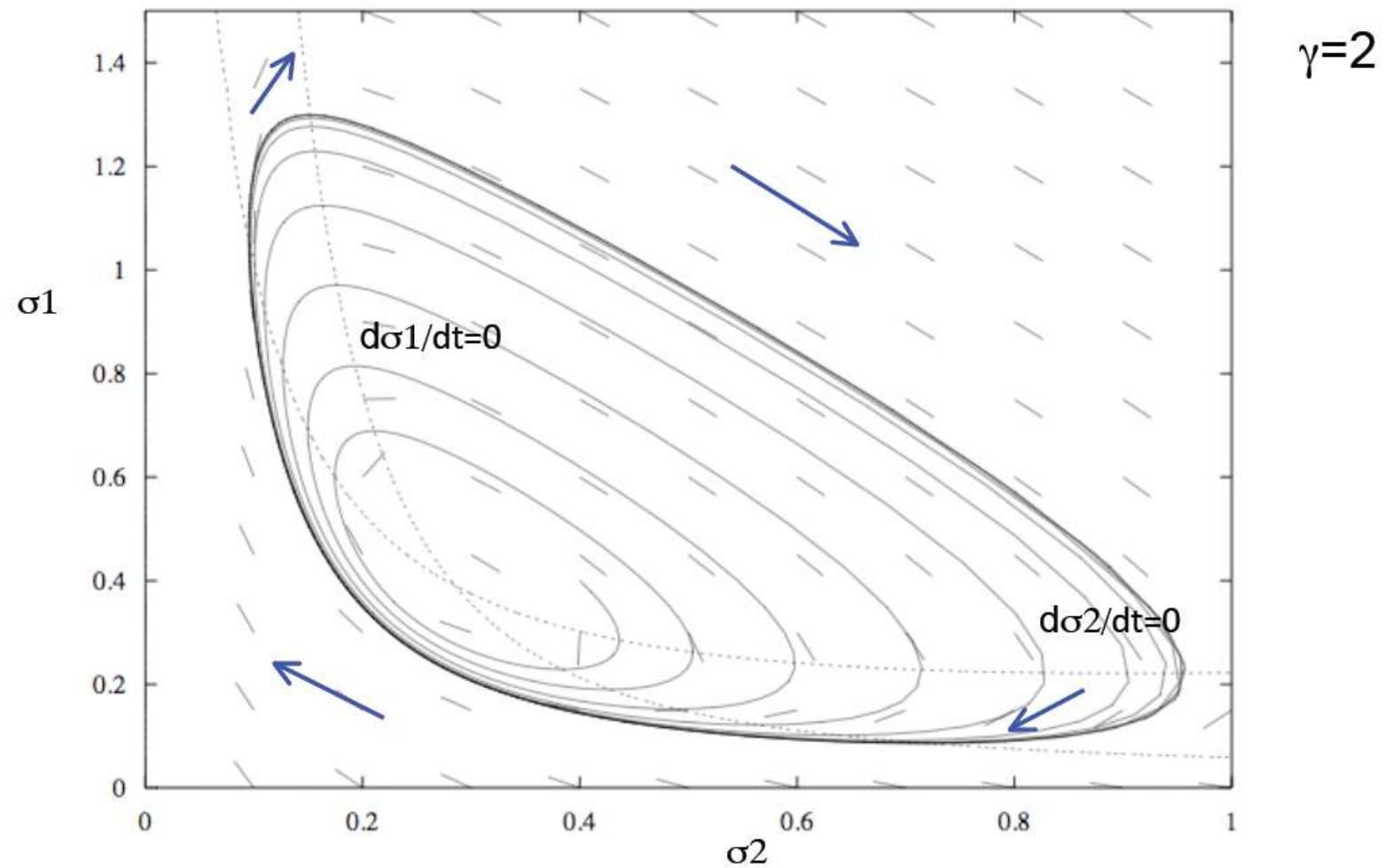
Node, Focus, or Saddle point



*focus = spiral

To summarize, the steady state can be classified into six cases as follows:

1. Unstable node: $\beta > 0$ and $\gamma > 0$.
2. Saddle point: $\gamma < 0$.
3. Stable node: $\beta < 0$ and $\gamma > 0$.
4. Unstable spiral: $\beta^2 < 4\gamma$ and $\beta > 0$.
5. Neutral center: $\beta^2 < 4\gamma$ and $\beta = 0$.
6. Stable spiral: $\beta^2 < 4\gamma$ and $\beta < 0$.



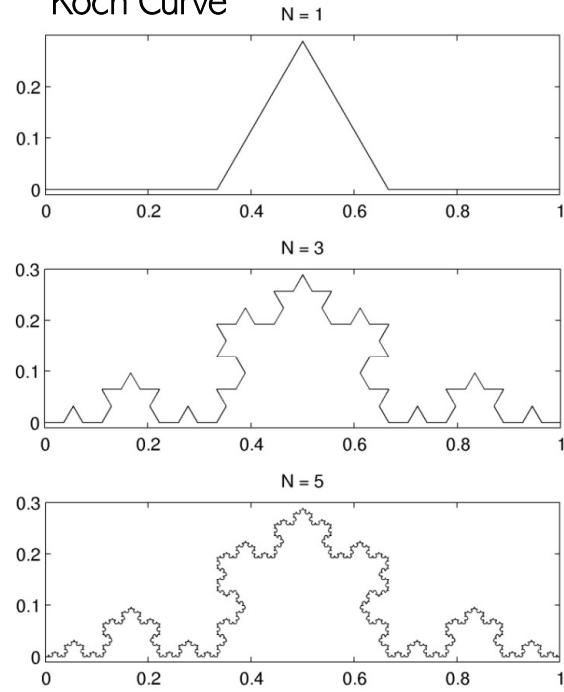
Slide courtesy Dr. Didier Gonze. Université Libre de Bruxelles, Belgium

More phase-space parameters: Dimensional Analysis

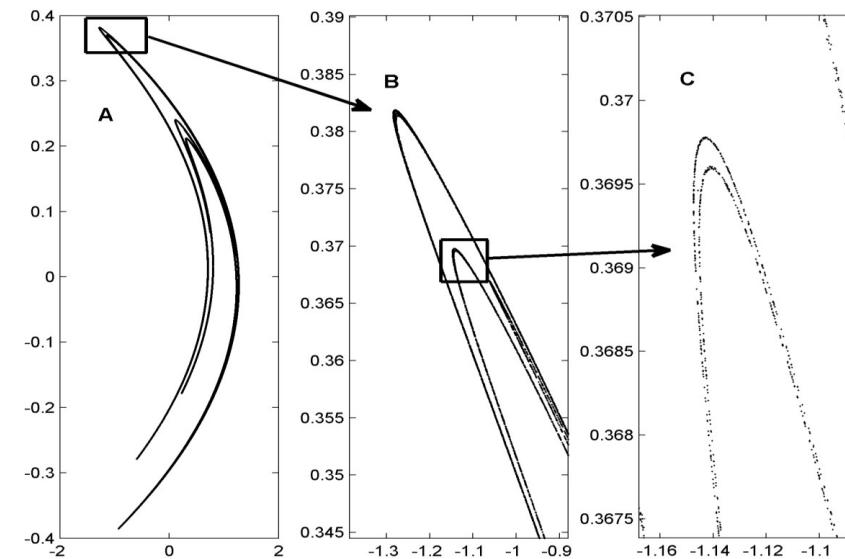
- Embedding dimension can only be integer values
 - For example, from PCA or False Nearest Neighbor analysis
- Chaotic attractors however can take on dimensions of fractional values
 - For example 2.05, 3.5
- Such objects are called fractals
 - Fractal objects have the property of self affinity
 - Their shapes are made of repeated patterns

Some Fractal Objects

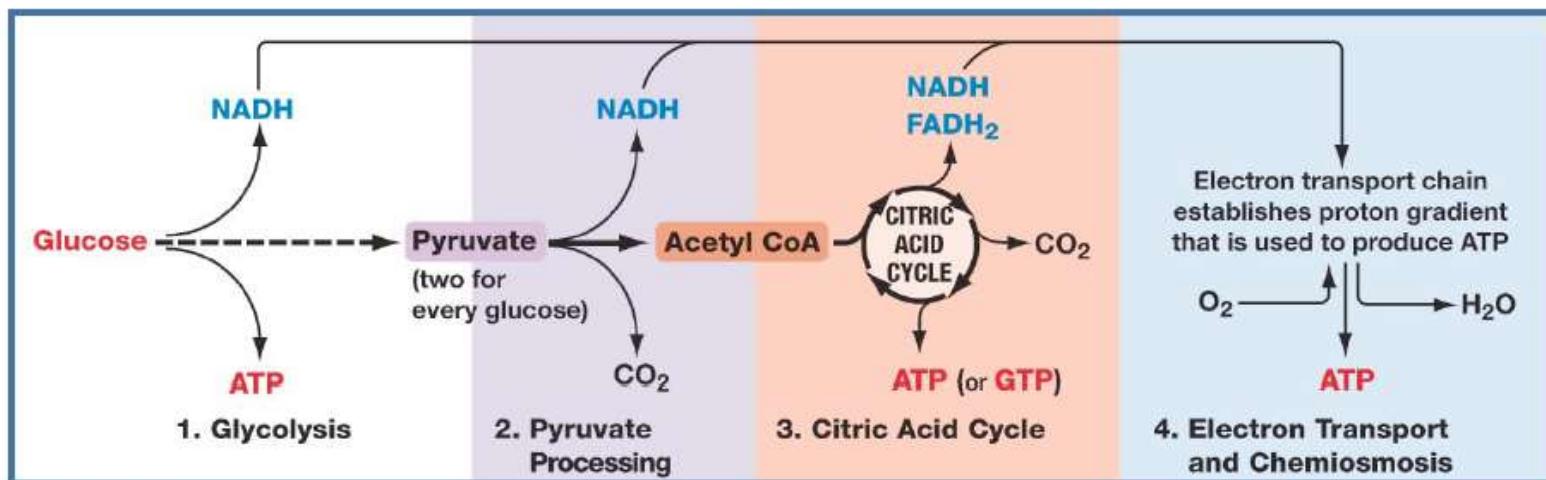
Koch Curve



Hénon Map

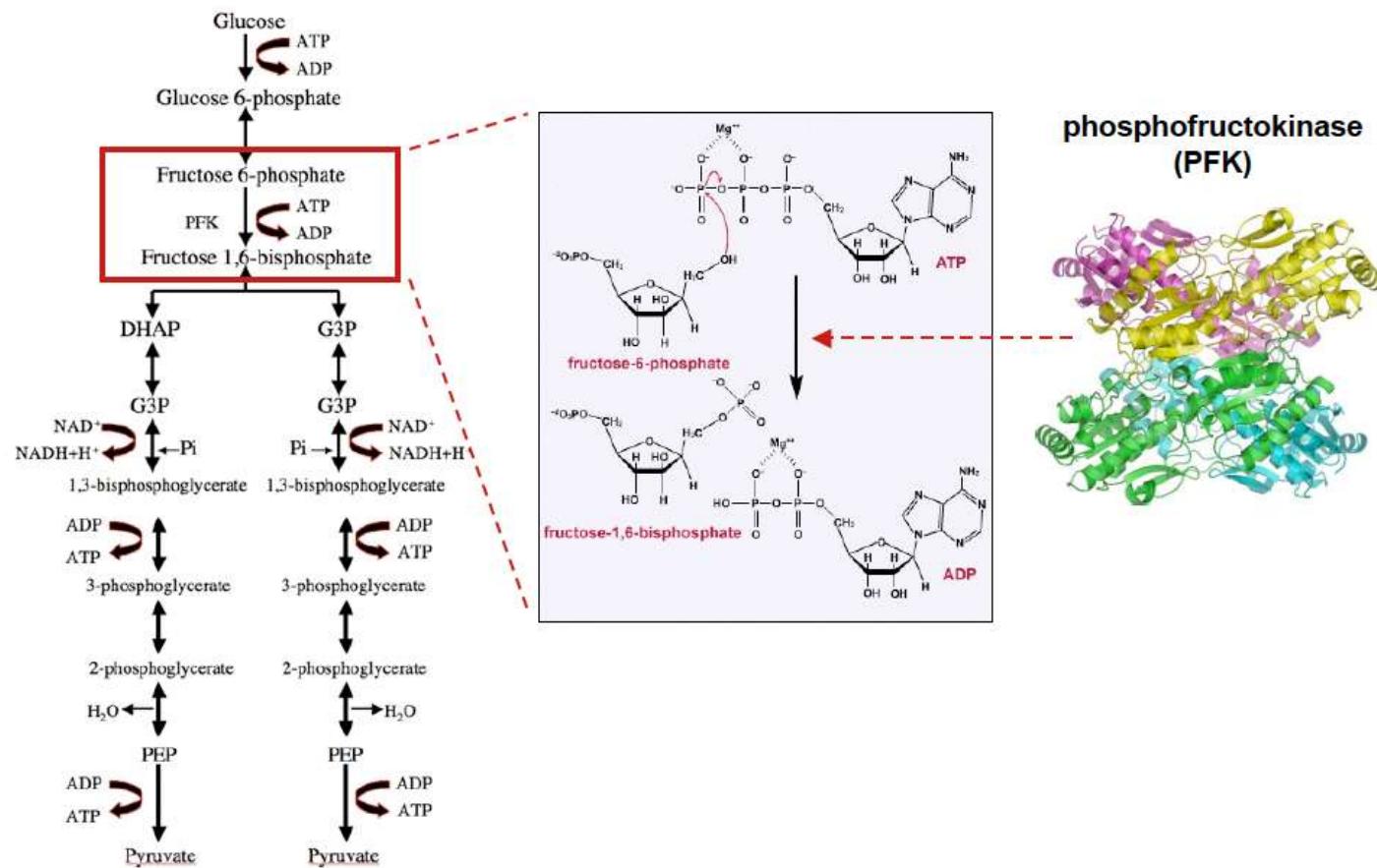


Glycolysis

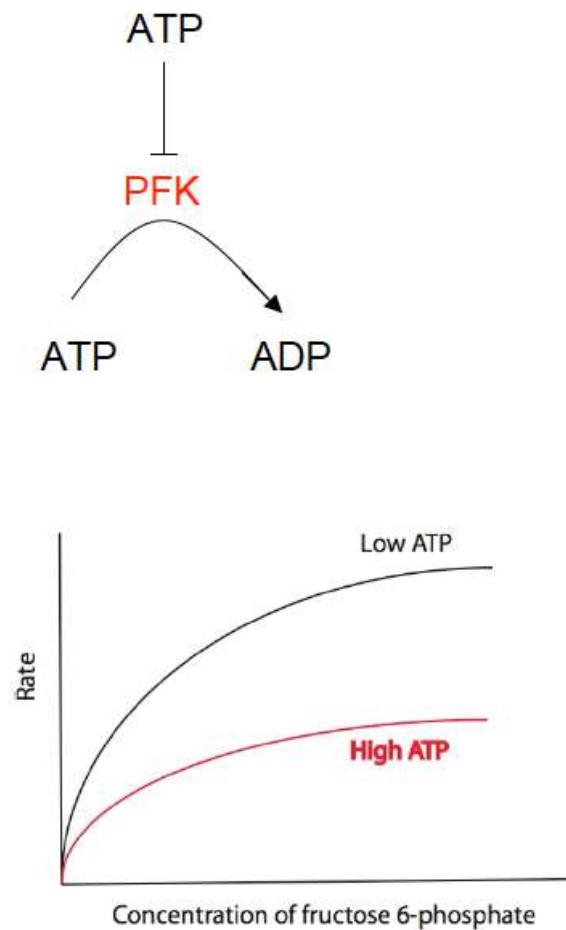
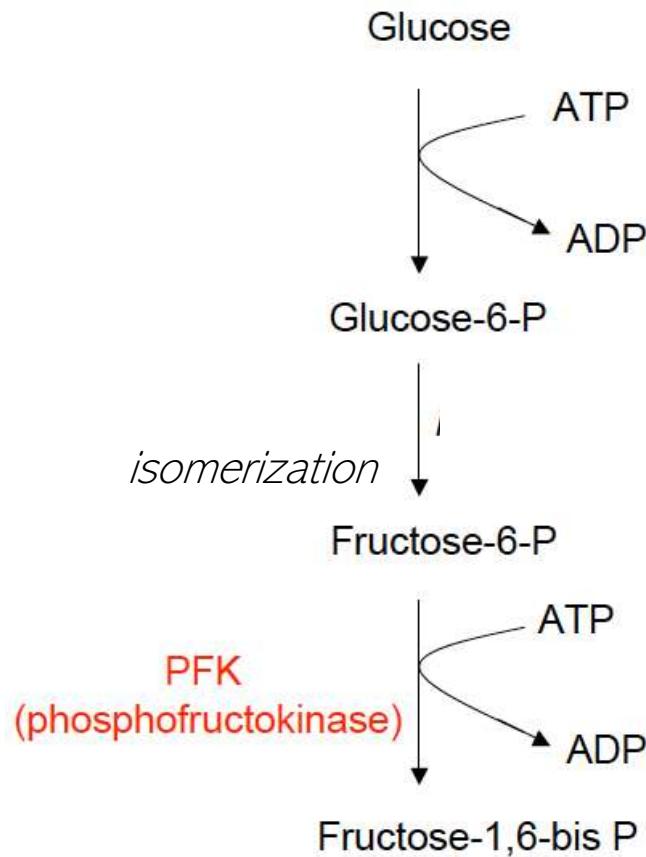


- Glycolysis literally means "splitting sugars."
- Glycolysis is the metabolic pathway that converts glucose into pyruvate.
- During glycolysis, two molecules of pyruvate are formed for every molecule of glucose.
- Pyruvate is then used in the Kreb cycle.
- Glycolysis also yields 2 molecules of ATP and 2 molecules of NADH.
- Glycolysis takes place in the cytoplasm.

Glycolysis & PFK

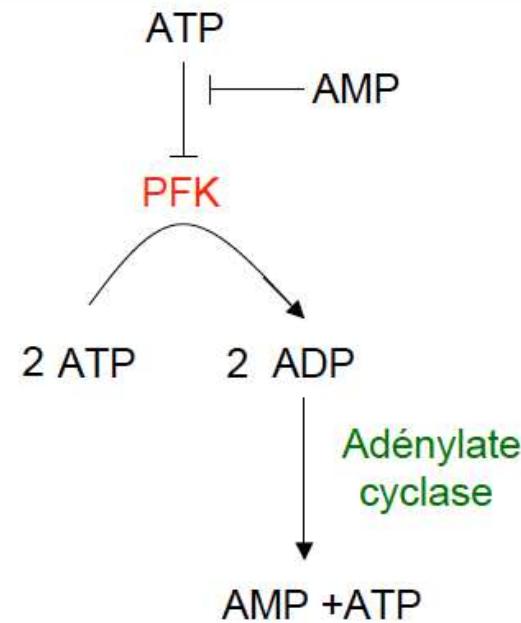
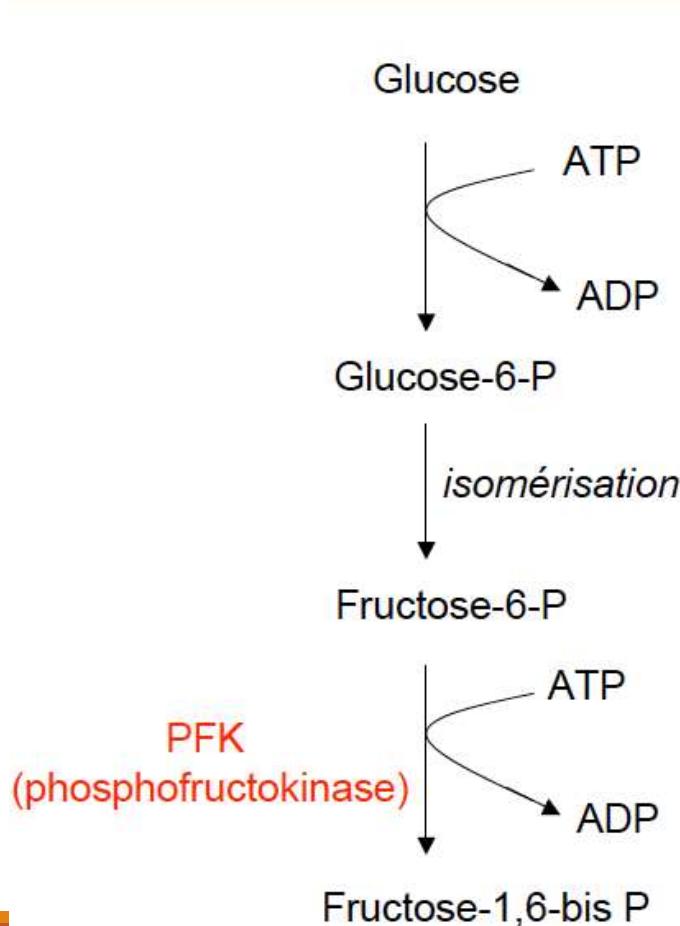


Glycolysis & PFK



Slide courtesy Dr. Didier Gonze. Université Libre de Bruxelles, Belgium

Glycolysis & PFK



Allosterically inhibited by ATP and allosterically activated by AMP (indicating cell's energetic needs). So, if PFK is active the ratio of [ATP]/[AMP] decreases.

Glycolysis Example

Consider: the nonlinear, coupled system proposed by Sel'kov describing glycolysis at the rate limiting phosphofructokinase (PFK) step:

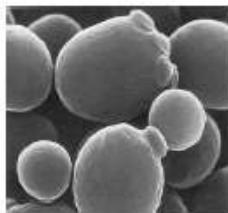
$$\frac{\partial x}{\partial t} = -x + ay + x^2y$$

$$\frac{\partial y}{\partial t} = b - ay - x^2y$$

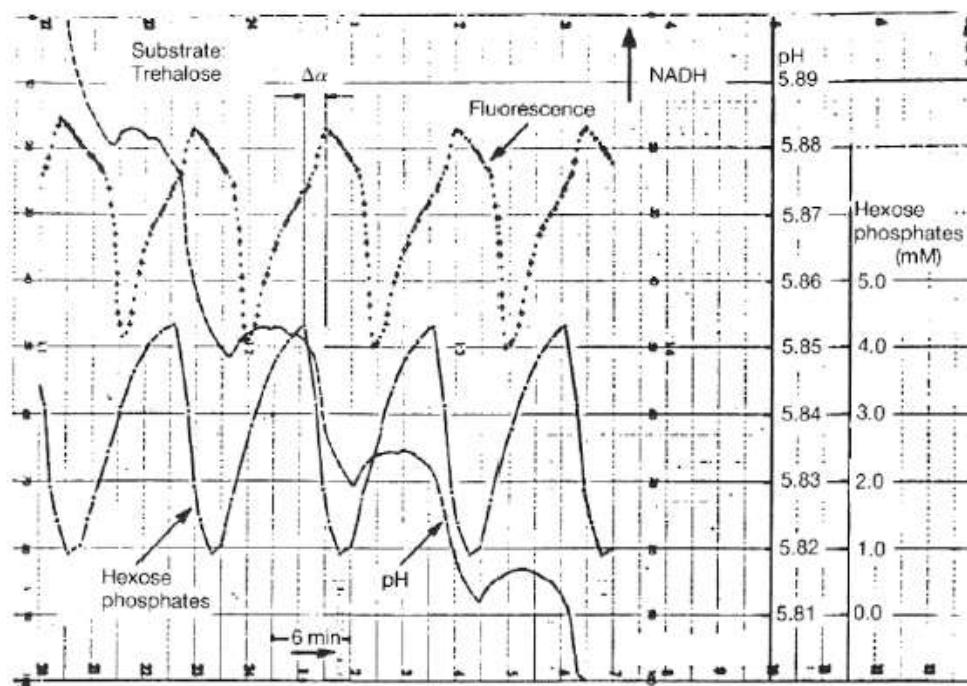
- a and b are rate constants, x=[ADP] and y=[F6P]
- there is a non-linear term (x^2y) but this is not chaotic (2 degrees of freedom- ADP and F6P). Also time is implicit in these differential equations.

If explicit time dependence was included in this system (e.g. change b such that $b=b\sin(\omega t)$) then there would be 3 degrees of freedom including time, and chaos would be possible

Glycolytic oscillations in *Saccharomyces cerevisiae*



S. cerevisiae
(yeast)



Glycolytic oscillations in a yeast extract subjected to constant injection of the substrate (trehalose). Chemical analyses show that the various hexoses oscillate with the same frequency as NADH.

Hess & Boiteux (1968) In Regulatory Functions of Biological Membranes. Ed. J. Jarnefelt, Elsevier.

Hess B, Boiteux A, Krüger J (1969) Cooperation of glycolytic enzymes. Adv Enzyme Regul. 7:149-67

Slide courtesy Dr. Didier Gonze. Université Libre de Bruxelles, Belgium

Ranges of Glycolytic Oscillation in Yeast Extract

<u>Input rate*</u> mm/hr	<u>Period</u> min	<u>Amplitude</u> in mm NADH	Damping	Waveform
< 20	—	steady high level of NADH	—	—
20	8.6	0.2–0.4	—	double periodicities, nonsinusoidal
40	6.5	0.6	—	nonsinus-sinus
60–80	5.0	0.3	—	stable sinus
120	3.5	0.2	—	stable sinus
> 160	—	steady low level of NADH	+++	

* Fructose or glucose serve as substrates. Cell-free extract of ~60 mg/ml.

Hess B, Boiteux A, Krüger J (1969) Cooperation of glycolytic enzymes. Adv Enzyme Regul. 7:149-67

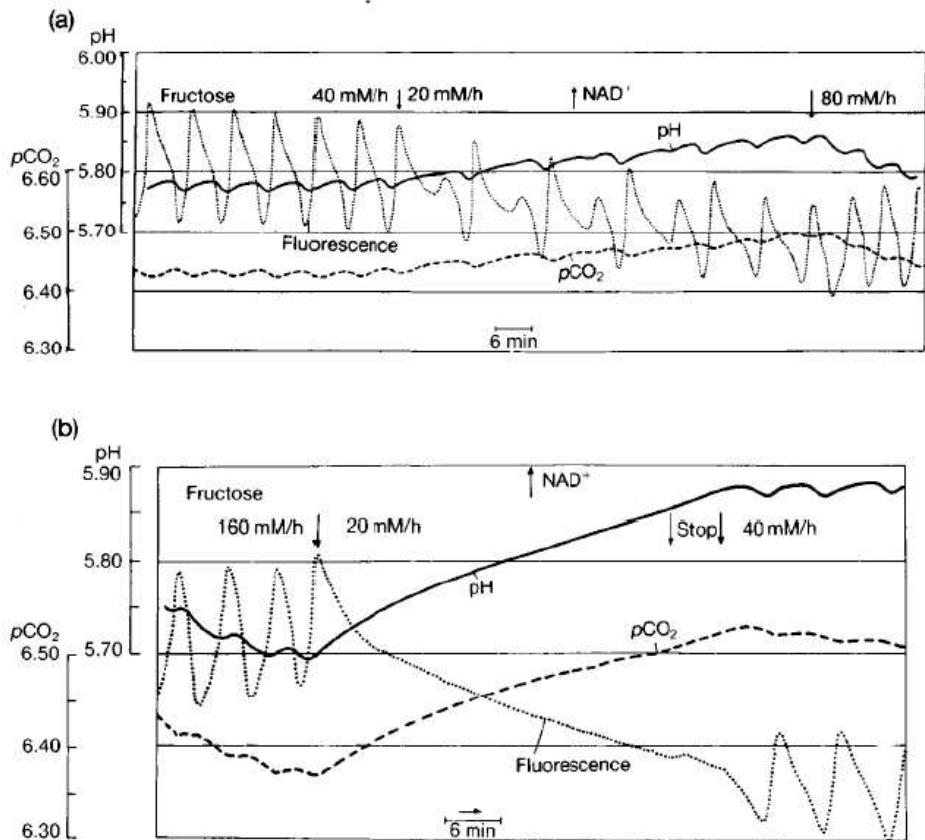


Fig. 2.4. Control of glycolytic oscillations in yeast extracts by the substrate injection rate. (a) The diminution of the rate of injection of fructose from 40 to 20 mM/h causes a lengthening of the period as well as a change in the waveform of oscillations; this change is reversible. (b) Decreasing the injection rate below 20 mM/h causes the reversible suppression of the oscillations (Hess & Boiteux, 1968b).

Control of glycolytic oscillations by the substrate injection rate:

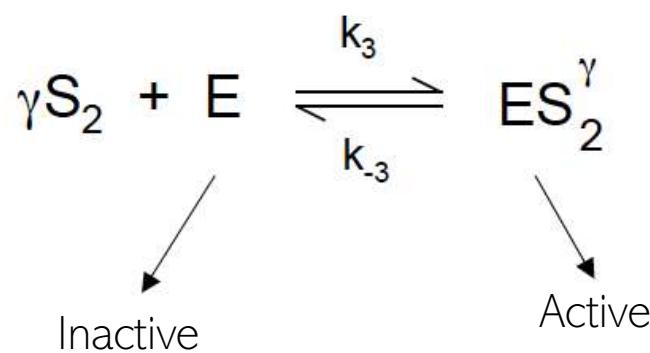
- (a) The diminution of the injection rate causes a lengthening of the period
- (b) Decreasing the injection rate below a certain threshold causes the reversible suppression of the oscillations

Hess B, Boiteux A (1968)
Hoppe Seylers Z Physiol
Chem. 349:1567-74.

PFK is responsible for the glycolytic oscillations

- 1) If glucose-6-phosphate or fructose-6-phosphate is taken as substrate, oscillations are still observed.
- 2) If fructose 1,6-bis-phosphate is used as substrate, there are no oscillations.
- 3) NH_4^+ (activates PFK) inhibits oscillations
- 4) Citrate (inhibits PFK) inhibits oscillations
- 5) Amplitude and frequency of oscillations can be varied by adding purified PFK to the cultured cells.

Hypothesis: activity of PFK is stimulated by 1 or several ADP molecules



Sel'kov (1968) Self-oscillations in Glycolysis, *Eur J Biochem* 4: 79-86.

Slide courtesy Dr. Didier Gonze. Université Libre de Bruxelles, Belgium

Evolution equations ($s_1 = [\text{ATP}]$; $s_2 = [\text{ADP}]$; $e = [\text{E}]$; $x_1 = [\text{ES}_2^\gamma]$; $x_2 = [\text{S}_1\text{ES}_2^\gamma]$)

$$\left\{ \begin{array}{lcl} \frac{ds_1}{dt} & = & v_1 - k_1 s_1 x_1 + k_{-1} x_2 \\ \frac{ds_2}{dt} & = & k_2 x_2 - k_3 s_2^\gamma e + k_{-3} x_1 - v_2 s_2 \\ \frac{dx_1}{dt} & = & -k_1 s_1 x_1 + (k_{-1} + k_2) x_2 + k_3 s_2^\gamma e - k_{-3} x_1 \\ \frac{dx_2}{dt} & = & k_1 s_1 x_1 - (k_{-1} + k_2) x_2 \end{array} \right. \quad \begin{array}{l} \text{with:} \\ e_0 = e + x_1 + x_2 \\ (\text{total concentration in PFK}) \end{array}$$

$$u_1 = \frac{x_1}{e_0} \quad u_2 = \frac{x_2}{e_0} \quad \sigma_1 = \frac{k_1 s_1}{k_{-1} + k_2} \quad \sigma_2 = \left(\frac{k_3}{k_{-3}} \right)^{1/\gamma} s_2$$

$$\tau = \frac{e_0 k_1 k_2}{k_{-1} + k_2} t \quad u_1 + u_2 + \frac{e}{e_0} = 1 \text{ (enzyme conservation)}$$

$$\begin{cases} \frac{d\sigma_1}{dt} = v - \frac{k_{-1} + k_2}{k_2} u_1 \sigma_1 + \frac{k_{-1}}{k_2} u_2 \\ \frac{d\sigma_2}{dt} = \alpha \left(u_2 - \frac{k_{-3}}{k_2} \sigma_2^\gamma (1 - u_1 - u_2) + \frac{k_{-3}}{k_2} u_1 \right) - \eta \sigma_2 \\ \epsilon \frac{du_1}{dt} = u_2 - \sigma_1 u_1 + \frac{k_{-3}}{k_{-1} + k_2} (\sigma_2^\gamma (1 - u_1 - u_2) - u_1) \\ \epsilon \frac{du_2}{dt} = \sigma_1 u_1 - u_2 \end{cases}$$

where $\epsilon = \frac{e_0 k_1 k_2}{(k_2 + k_{-1})^2}$ $v = \frac{v_1}{k_2 e_0}$ $\eta = \frac{v_2 (k_{-1} + k_2)}{e_0 k_1 k_2}$ $\alpha = \frac{k_{-1} + k_2}{k_1} \left(\frac{k_3}{k_{-3}} \right)^{1/\gamma}$

Analysis of the 2-equation system (for σ_1 et σ_2) in the phase space

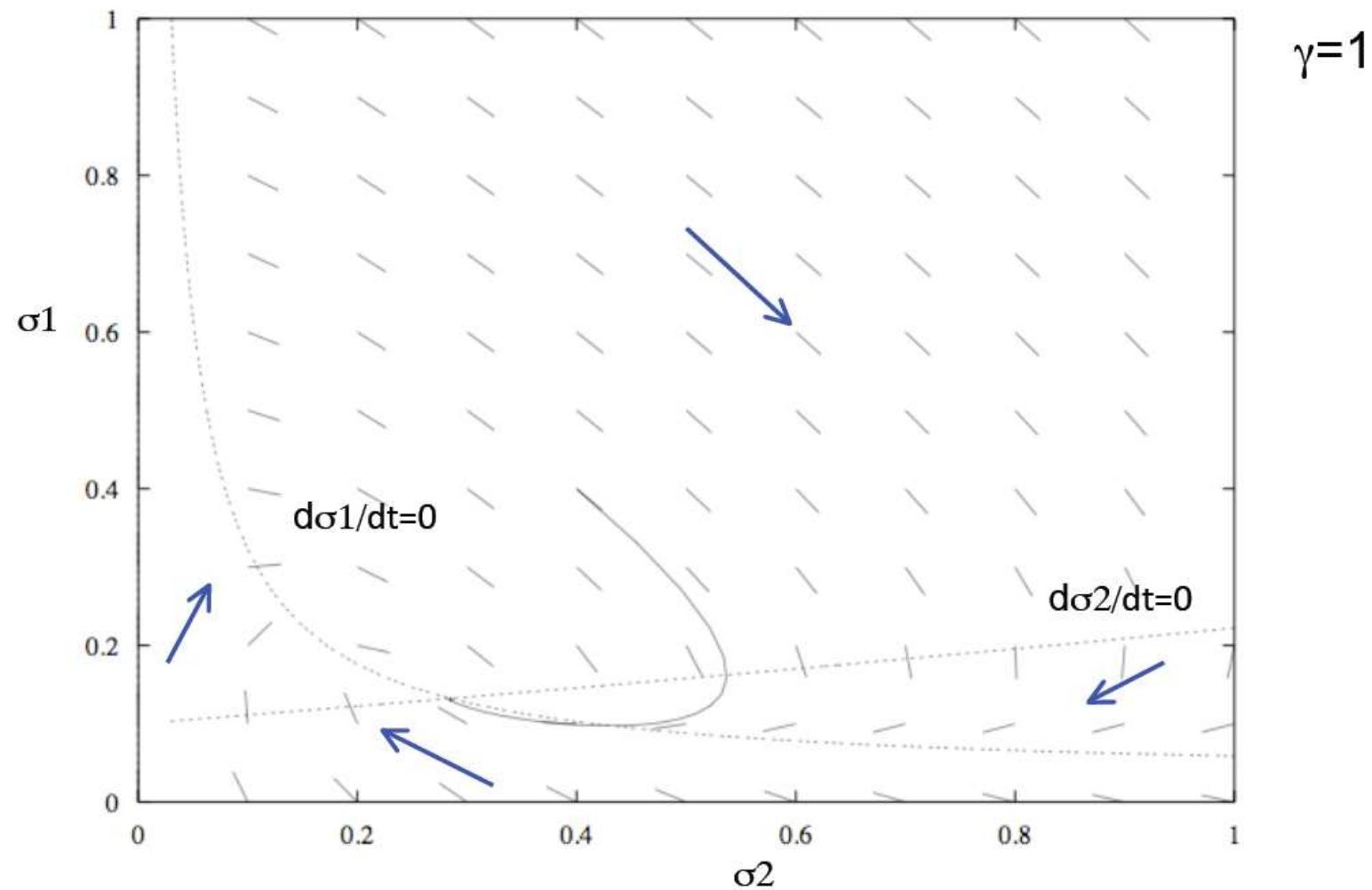
Nullclines:

$$(1) \quad V = f(\sigma_1, \sigma_2) \quad \text{Hence:} \quad \sigma_1 = \frac{V}{1-V} \frac{1+\sigma_2^\gamma}{\sigma_2^\gamma}$$

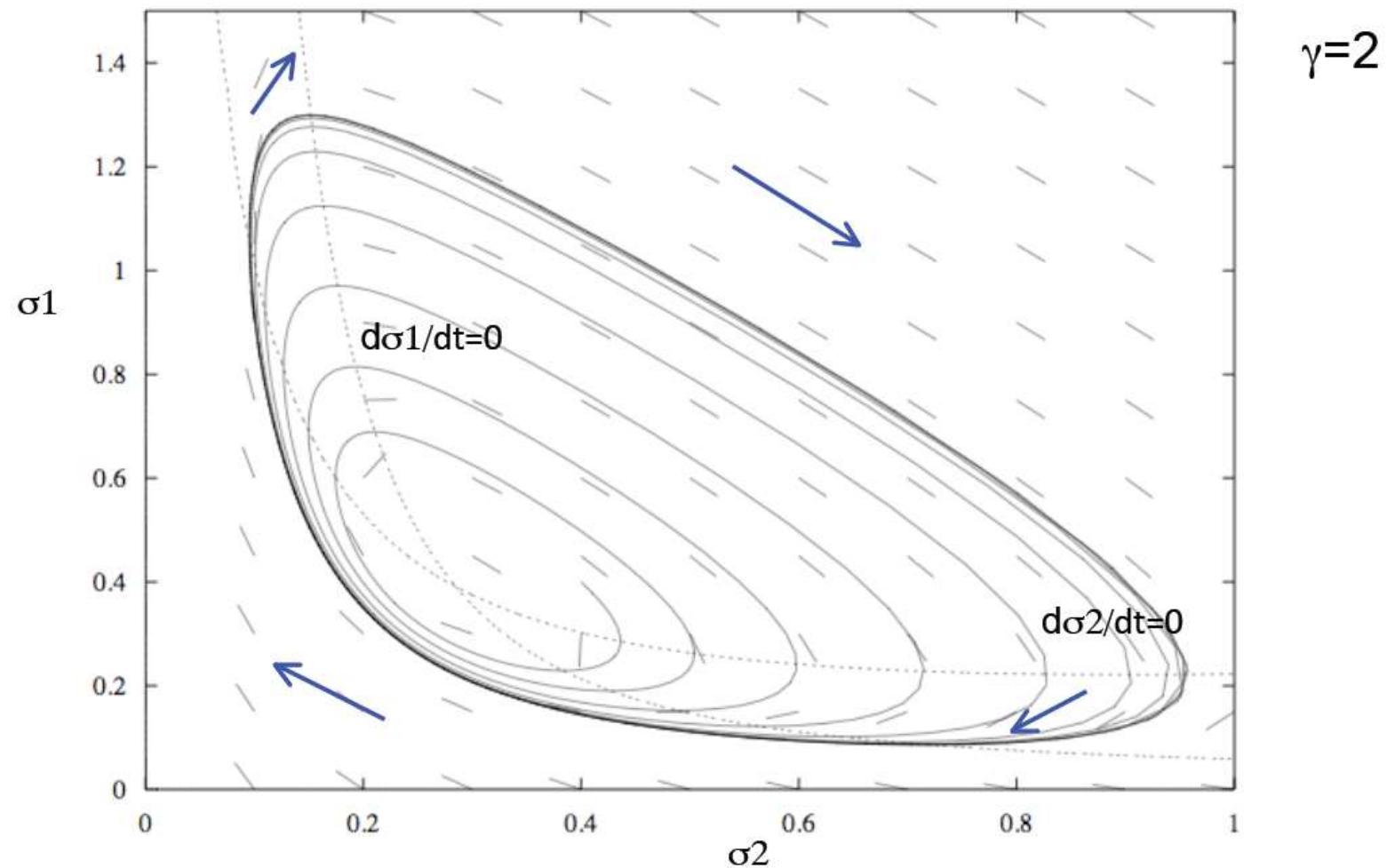
$$(2) \quad \alpha f(\sigma_1, \sigma_2) = \eta \sigma_2 \quad \text{Hence:} \quad \sigma_1 = \frac{1+\sigma_2^\gamma}{\sigma_2^{\gamma-1}(p-\sigma_2)} \quad \text{where} \quad p = \frac{\alpha}{\eta}$$

The intersection of the nullclines defines the steady state: $\sigma_1^{SS}, \sigma_2^{SS}$

recall γ is number of ADP molecules over the reaction

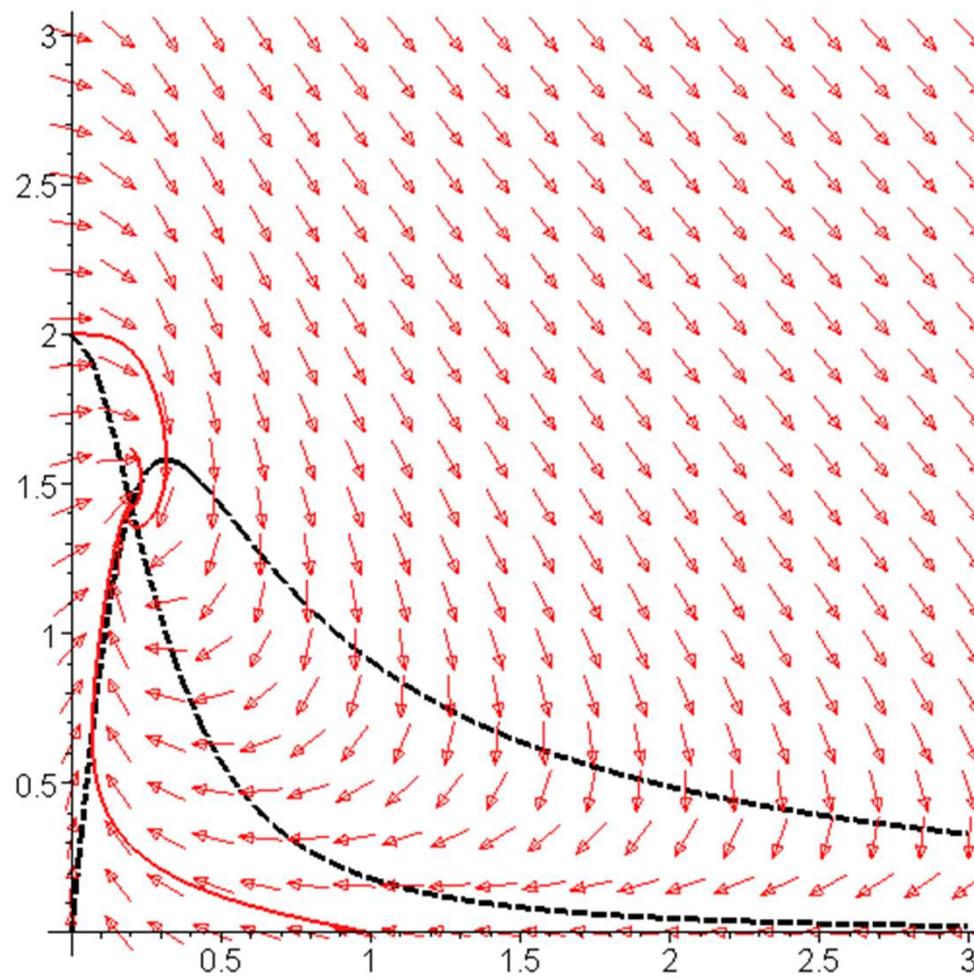


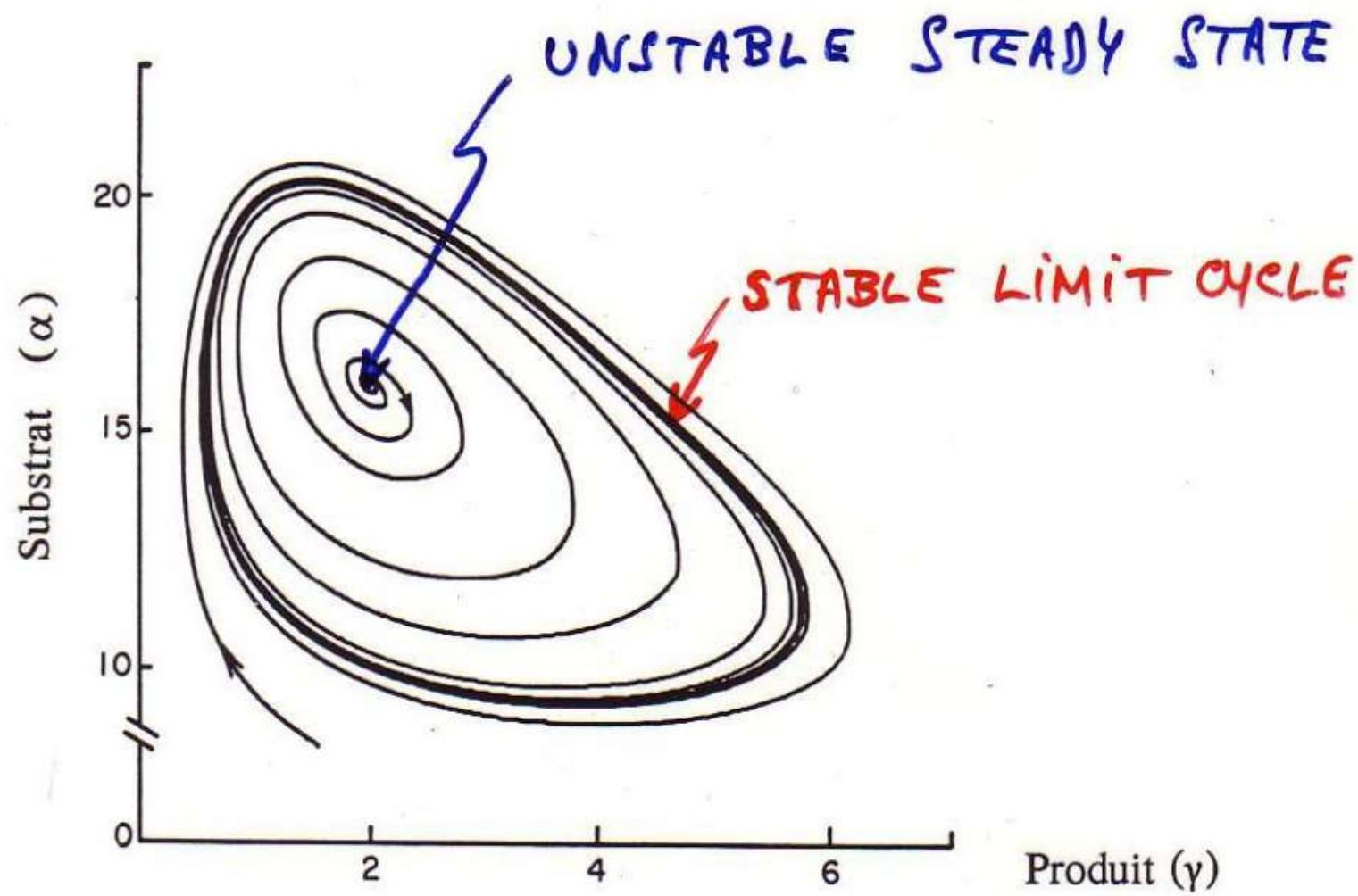
Slide courtesy Dr. Didier Gonze. Université Libre de Bruxelles, Belgium



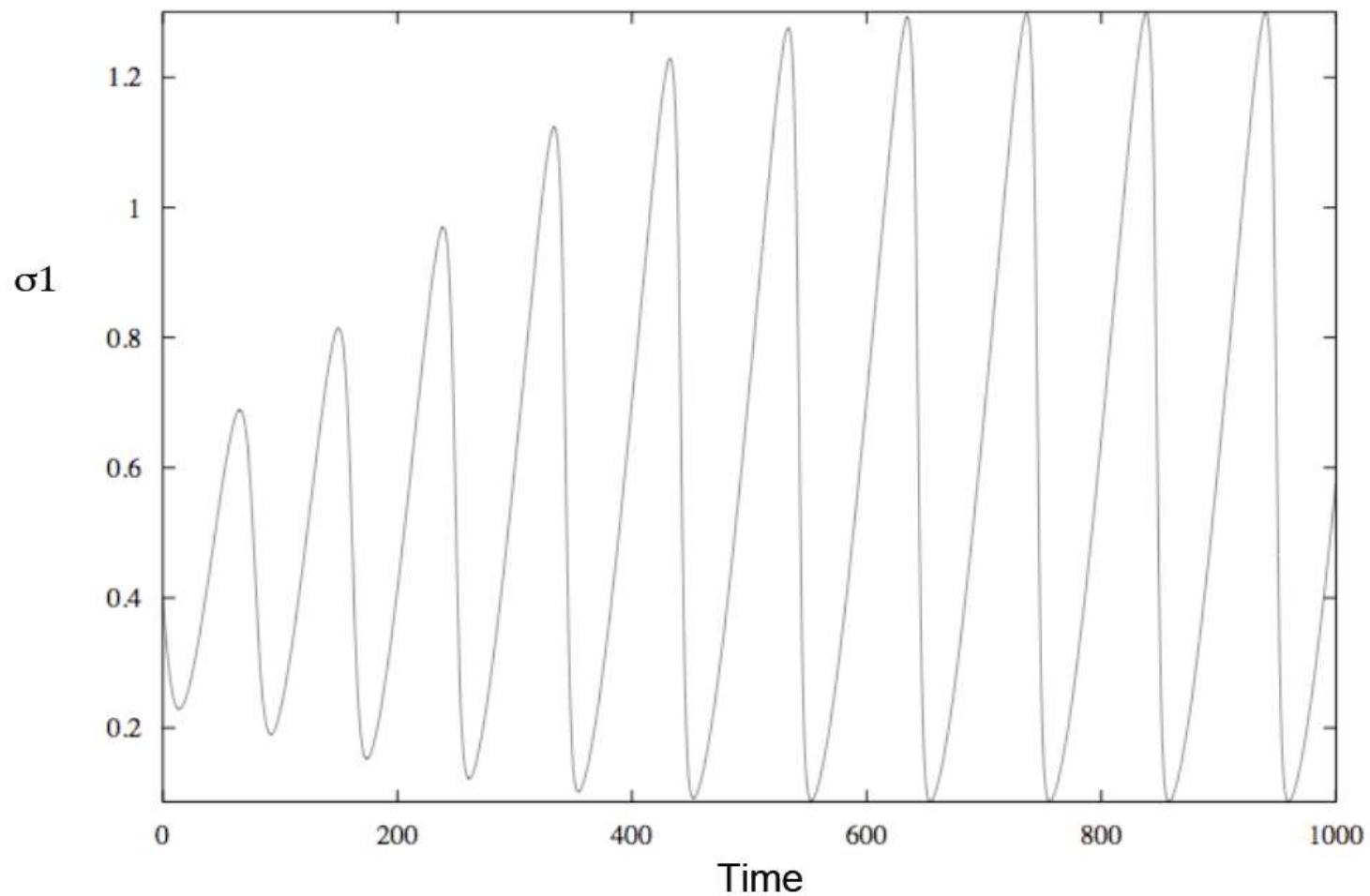
Slide courtesy Dr. Didier Gonze. Université Libre de Bruxelles, Belgium

$$a = .1; b = .2$$

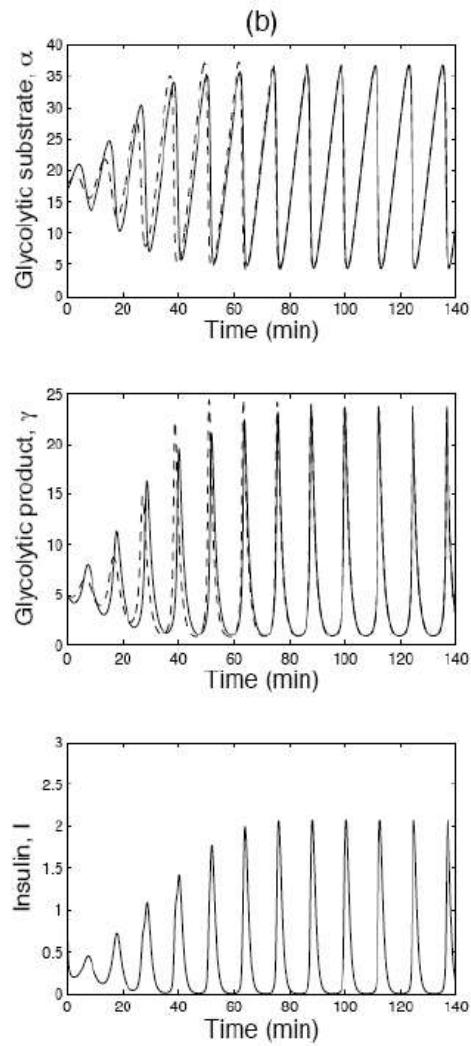
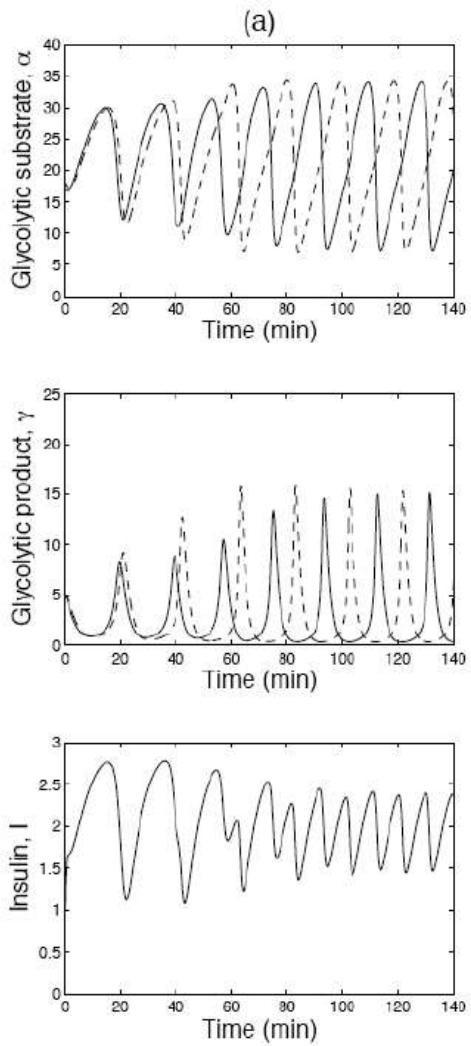




Slide courtesy Dr. Didier Gonze. Université Libre de Bruxelles, Belgium



Slide courtesy Dr. Didier Gonze. Université Libre de Bruxelles, Belgium

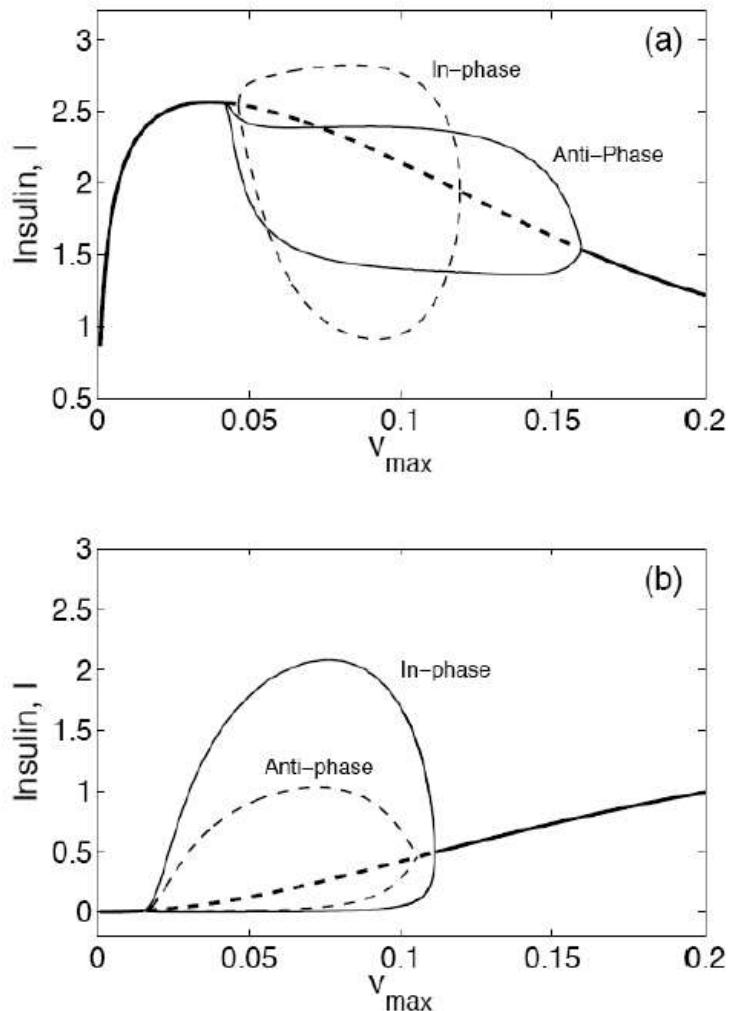


The mode of synchronization depends on the way the two oscillators are coupled:

(a) When insulin release is controlled by the glycolytic **substrate** (α), the oscillations are in anti-phase

(b) When insulin release is controlled by the glycolytic **product** (γ), the oscillations are in phase.

Slide courtesy Dr. Didier Gonze. Université Libre de Bruxelles, Belgium



Bifurcation diagram as a function of the maximum rate of glucose input (V_{max}) into the cell.

(a) When insulin release is controlled by the glycolytic **substrate**, the stable limit cycle regime corresponds to anti-phase synchronization, while the unstable limit cycle regime corresponds to in-phase oscillations.

(b) When insulin release is controlled by the glycolytic **product**, the stable limit cycle regime corresponds to in-phase synchronization and the unstable limit cycle regime corresponds to antiphase oscillations.

Other Information:

Check out the Glycolysis Simulator by Dr. Dr. Bernhard Palsson

http://gcrg.ucsd.edu/sites/default/files/Attachments/Images/publications/books/systemsBiolog_y1/MetlabNotebooks/glycolysis.zip

https://www.youtube.com/watch?v=QgPvBLFQ_xU

Conclusions

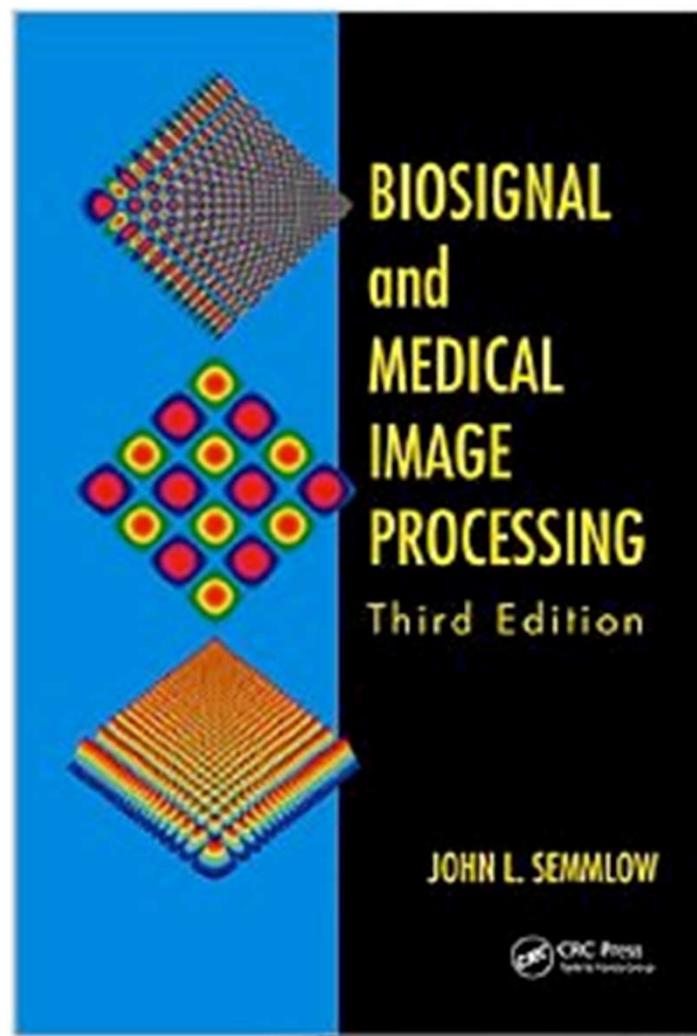
Nonlinear analysis

- Gives more information about nonlinear signals than linear analysis alone
- Has to be used carefully because it can take a long time and give misleading results

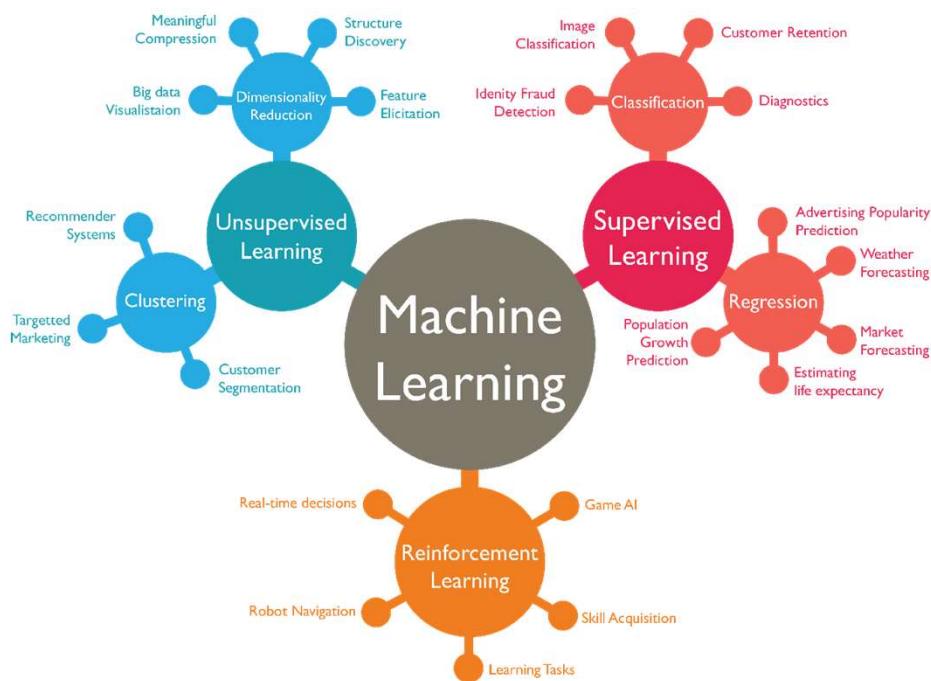
Surrogate data testing is needed to establish nonlinearity

- Otherwise nonlinear testing could be useful for distinguishing between signal types, but won't be advantageous

Note: All materials, code and data for the following materials are from Biosignal and Medical Image processing 3rd Edition



Machine Learning



<https://www.wordstream.com/blog/ws/2017/07/28/machine-learning-applications>

Statistical Learning

- Machine learning is the general term but if a model uses statistical methods to achieve its goal it is statistical learning

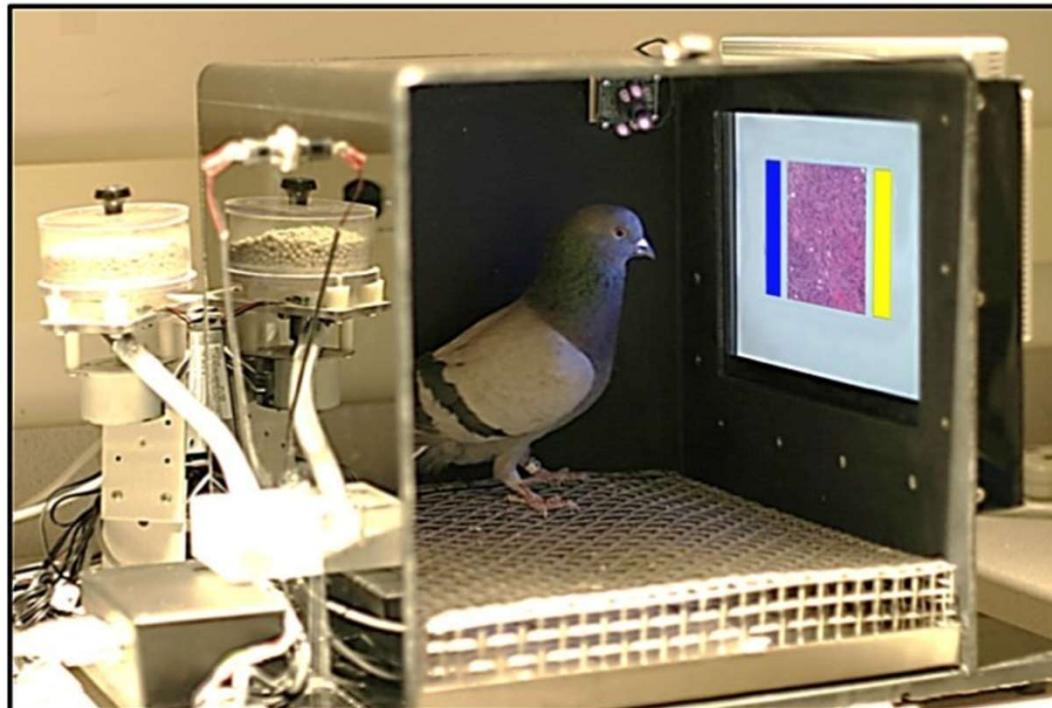
Goal

- build a model that makes predictions based on evidence in the presence of uncertainty.

Paging Dr. Pigeon; You're Needed in Radiology

By NICHOLAS BAKALAR NOV. 24, 2015

The New York Times



The pigeons' training environment at the University of Iowa included a food pellet dispenser, a touch-sensitive screen that projected medical images, and blue and yellow choice buttons on either side.

University of Iowa/Wassermann Lab

Examples

Consider machine learning for a complex task or problem involving a large amount of data and/or lots of variables, but no existing formula or equation.

For example:

- 1) face recognition and speech recognition.
- 2) When rules of a task are constantly changing—as in fraud detection from transaction records.
- 3) If the nature of the data keeps changing, and the program needs to adapt such as for predicting shopping trends.

Real World Applications

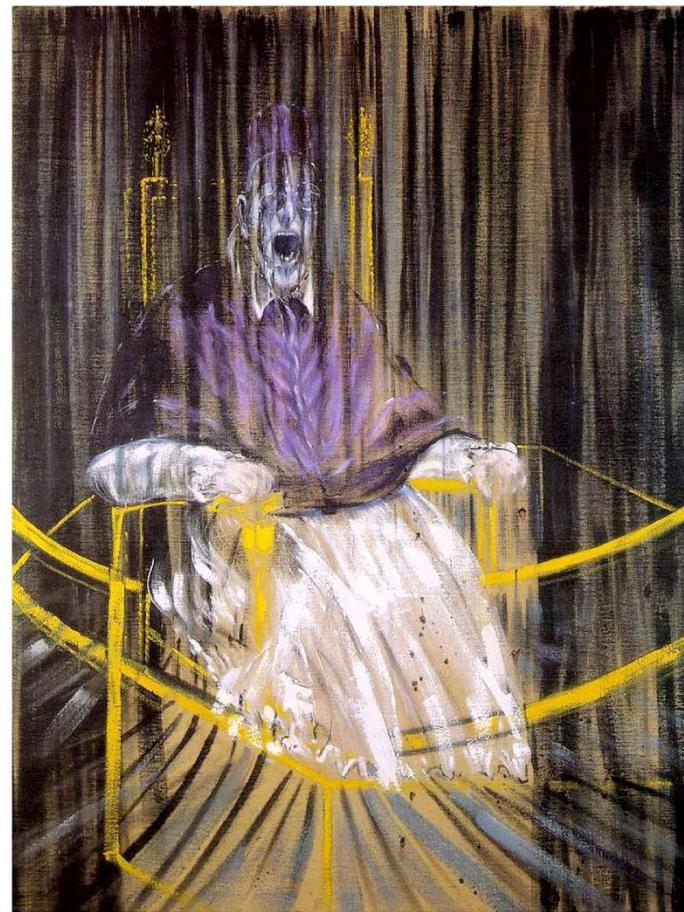
- Finance
 - credit scoring
 - Predicting stock trends
 - Targeted ads & suggestions
 - Fraud Detection
- Image processing and computer vision
 - facial recognition
 - motion detection
 - object detection
- Computational biology
 - tumor detection
 - drug discovery
 - DNA sequencing
- Energy production
 - price
 - load forecasting
- Automotive, aerospace, and manufacturing
 - predictive maintenance
 - Autonomy
- Natural language processing
 - Google Home/Amazon Alexa
 - Automated phone systems

Creating Algorithms that Can Analyze Works of Art

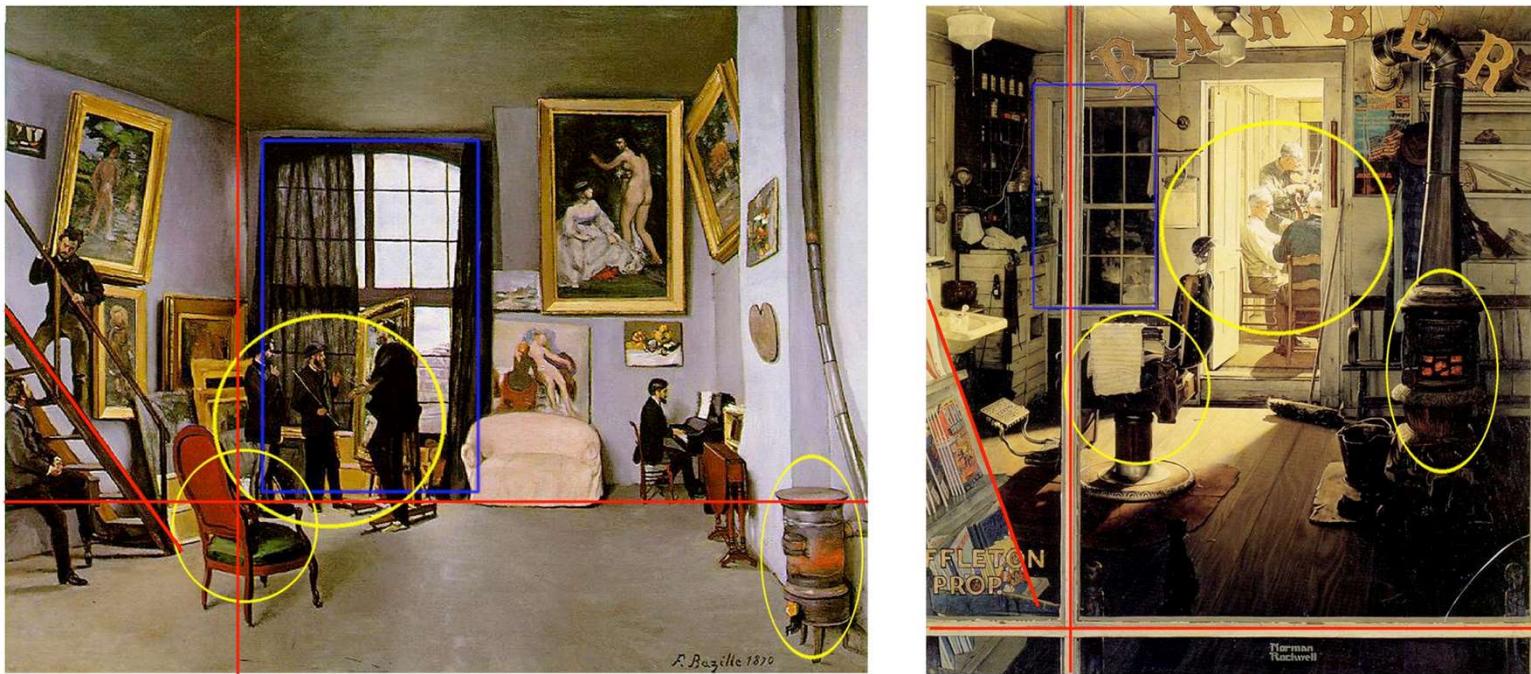
- Art and Artificial Intelligence Laboratory (Rutgers University)
 - used machine learning to classify paintings by style, genre, and artist.
 - Algorithms they developed classified the styles of paintings in large database with 60% accuracy, outperforming typical non-expert humans.
-
- tested >1,700 paintings from 66 different artists working over a span of 550 years. - machine learning readily identified connected works



Diego Velazquez's
Portrait of Pope Innocent X"

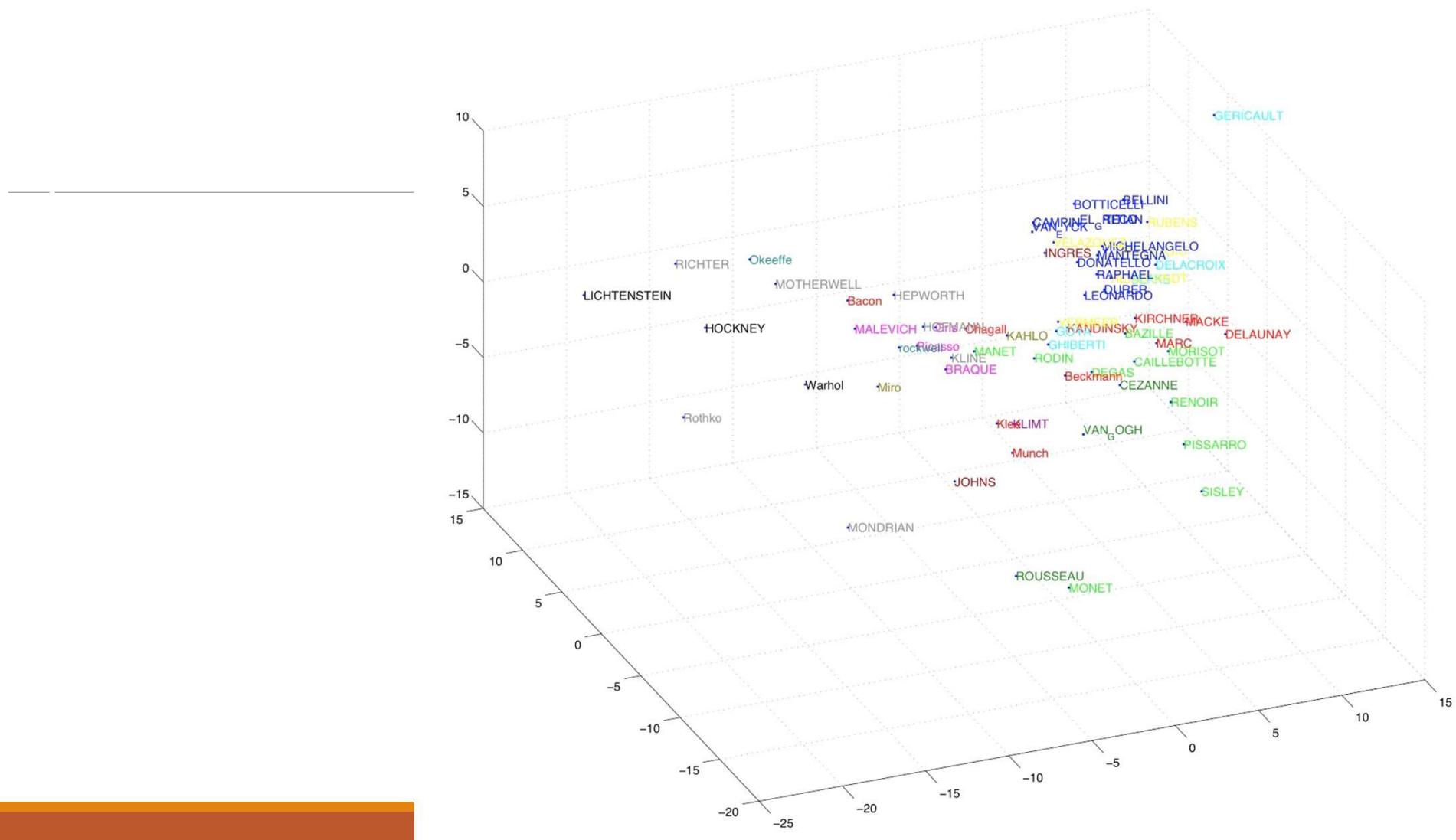


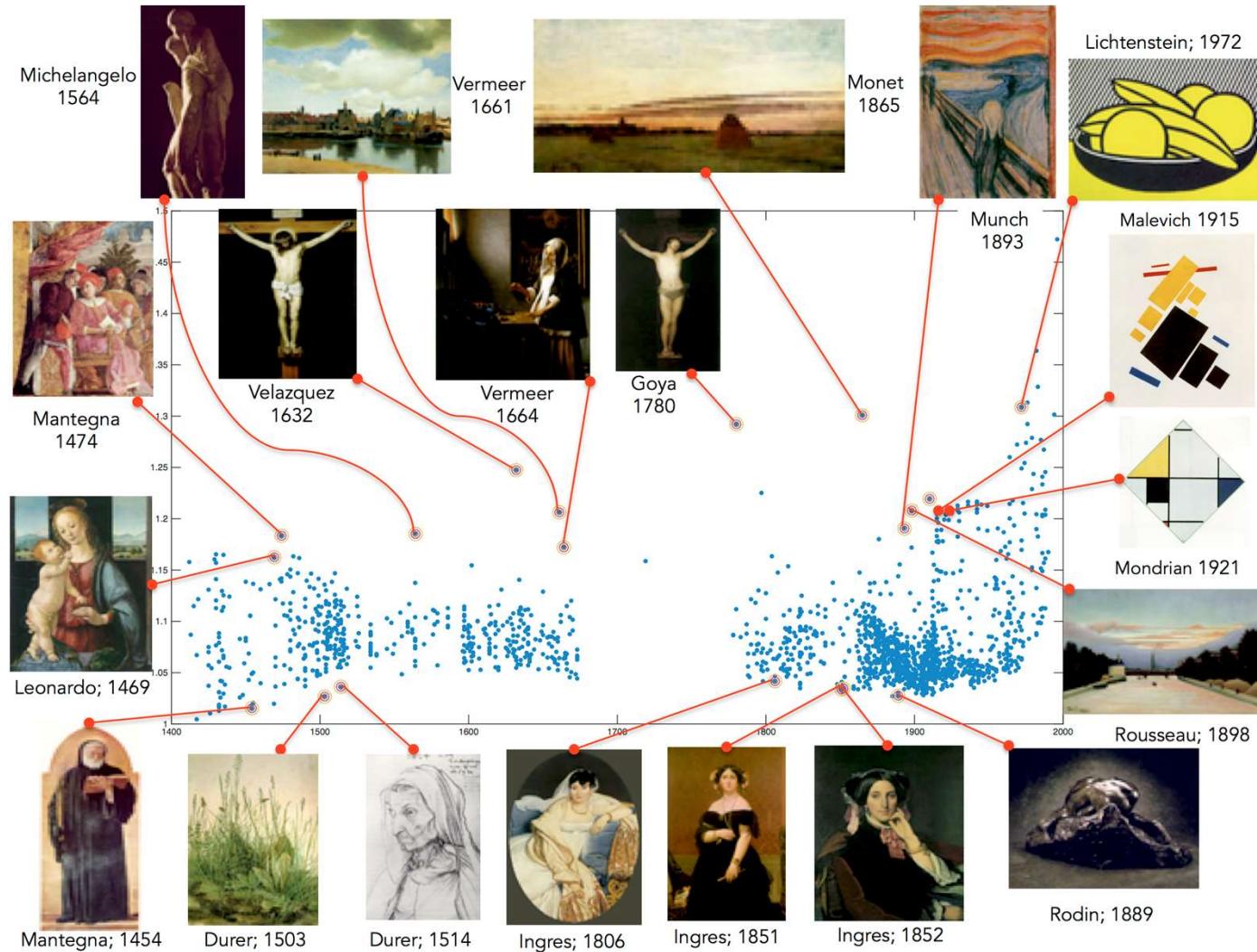
Francis Bacon's
"Study After
Velazquez's Portrait of Pope Innocent X."



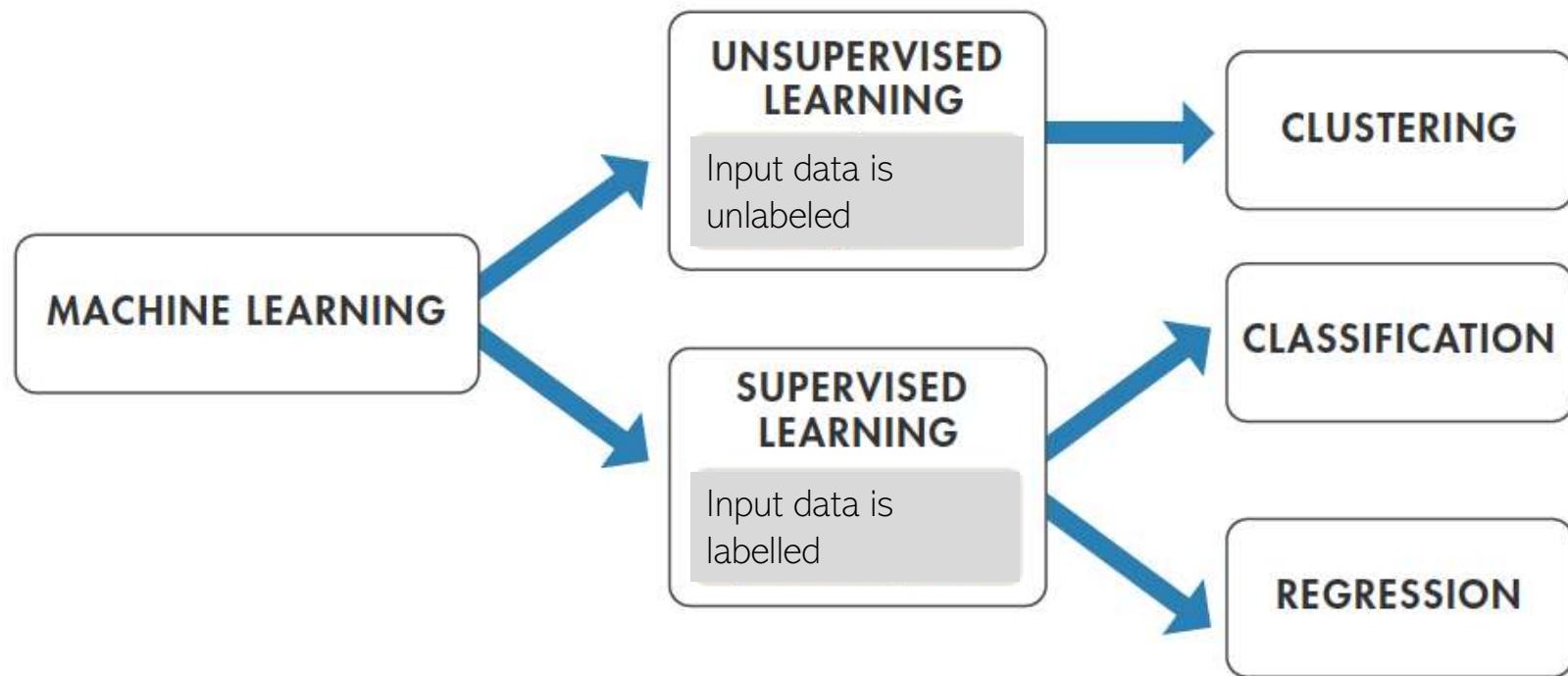
Frederic Bazille's Studio 9 Rue de la Condamine (left) and Norman Rockwell's Shuffleton's Barber Shop (right).

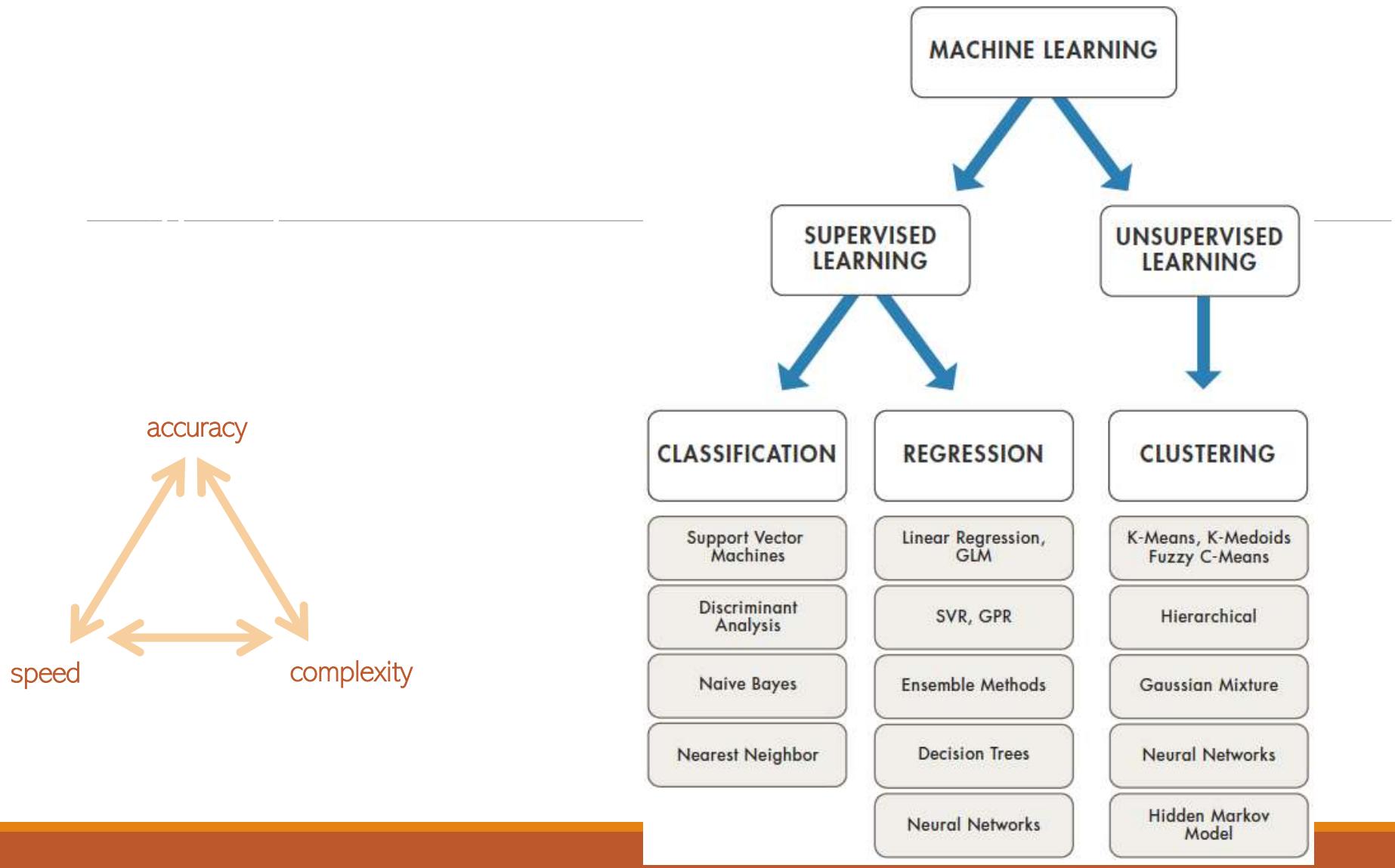
- The composition of both paintings is divided in a similar way.
- Yellow circles indicate similar objects, red lines indicate composition, and the blue square represents similar structural element.





Types of Learning





Supervised Learning

Supervised learning is the most common approach in Biomedical Engineering applications.

- uses data that is labelled into classes to train an algorithm on how to sort data
- Has a data set for training and one for validation

2 Steps:

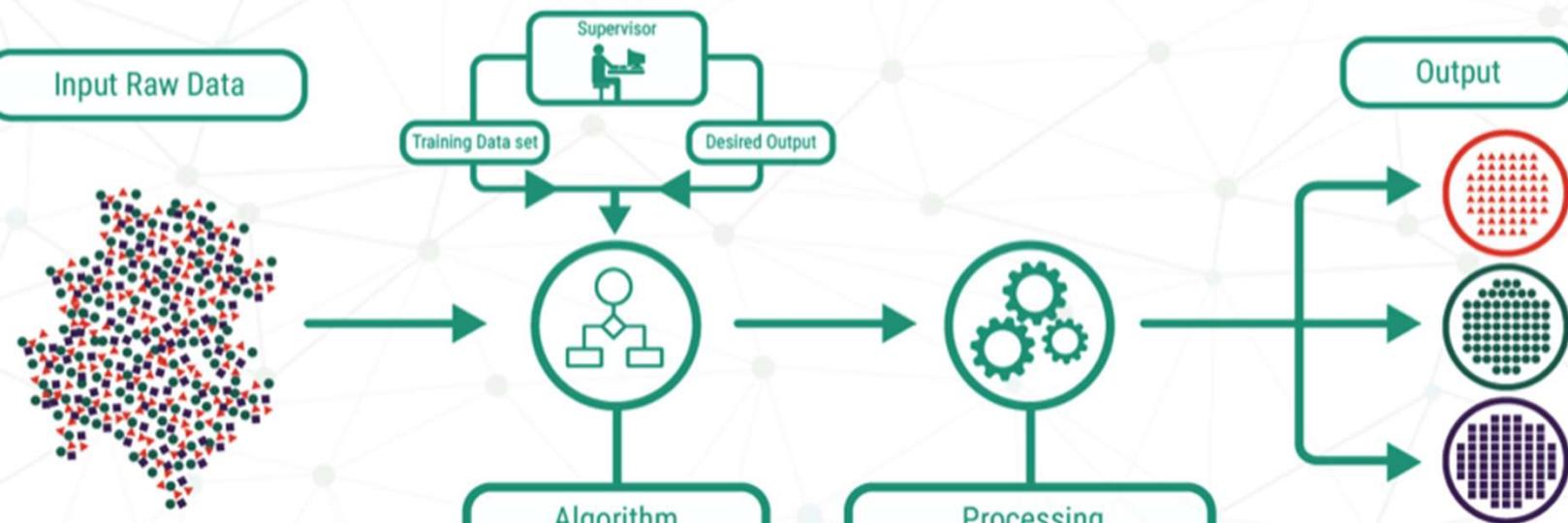
Training

- Parameters start out classifier free and are adjusted to minimize classification errors
- Use a subset of the data called the training set to let algorithm know which group the data belongs in

Validation

- A validation subset is used that the algorithm has never seen
- correct classification is also known but is not used to modify classifier parameters, just check accuracy
- Classifier should perform with minimum error on data that it has never seen.

SUPERVISED LEARNING



<https://medium.com/@himanshuit3036/supervised-learning-methods-using-python-bb85b8c4e0b7>

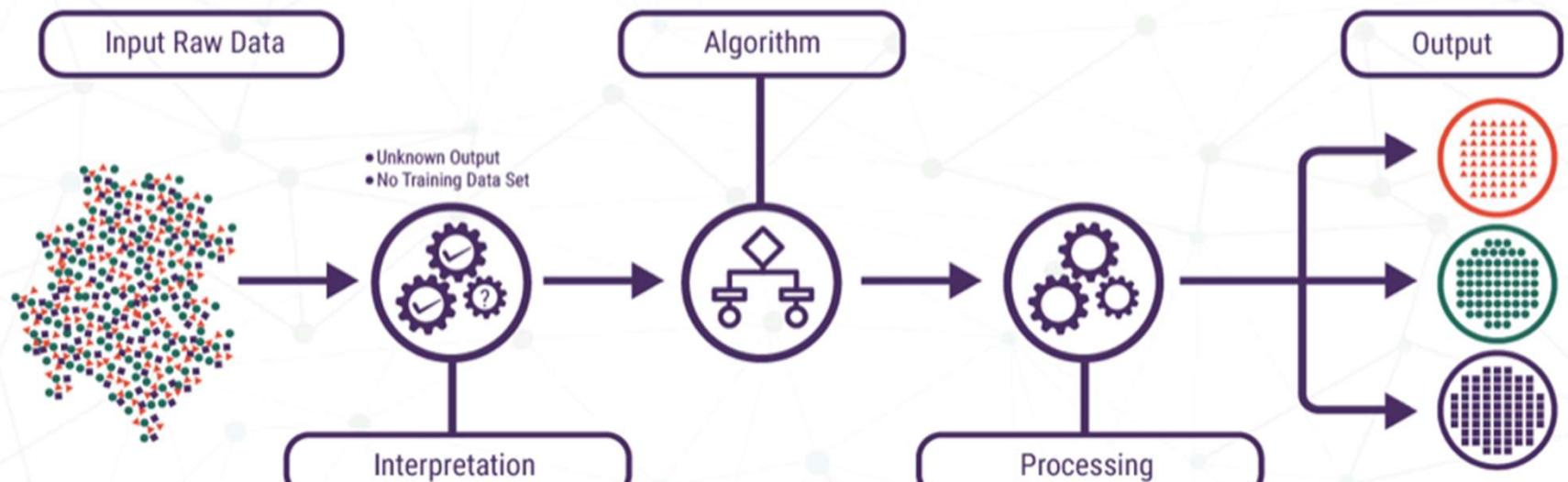
Unsupervised Learning

- Data is not labelled
- the classifier attempts to find patterns within the data with no a priori knowledge of the data patterns and, perhaps, not even the number of classes that exist.
- used to draw inferences from datasets consisting of input data without labeled responses.

Clustering

- most common unsupervised learning technique.
- used for exploratory data analysis to find hidden patterns or groupings in data.
- Applications include gene sequence analysis, market research, and object recognition

UNSUPERVISED LEARNING



<https://technative.io/why-unsupervised-machine-learning-is-the-future-of-cybersecurity/>

Supervised Learning

Pros

Easier to understand the sorting

Know how many classes there are before sorting

Can be very picky with class definition to tailor the decision boundary accuracy

Once it is trained, you can stop changing the algorithm

Cons

Expensive to label

Don't always have knowledge to create label

Fairly simple systems, can't handle complex sorting as well

Doesn't give you unknown information

- You have to understand which features and groups you are giving it

Possible to over-train

Unsupervised Learning

Pros

Useful for very large sets

Useful for unlabeled data

- Labelling is time consuming and expensive
- Labels can be confirmed after they have been sorted

Finds patterns that may be difficult to find otherwise

Good at identifying outliers

Cons

Time consuming to actually complete the algorithm

- Analyzes and calculates all possibilities for all data

The more features added, the more complicated and time consuming

Algorithm is always changing with the addition of new data

Semi Supervised

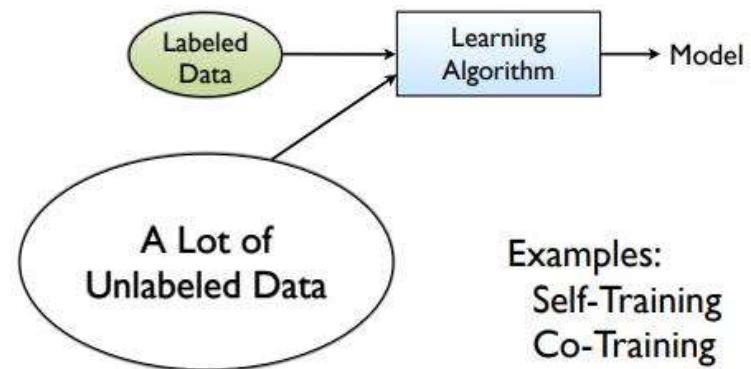
Unsupervised algorithms generate labels

These labels are used for a supervised algorithm

Humans may label some data

Can help with high cost of labelling

Semi-Supervised Learning (SSL)



Examples:
Self-Training
Co-Training

<https://www.programmersought.com/article/19703919020/>

Loss Function

Maps variable vs its “cost”

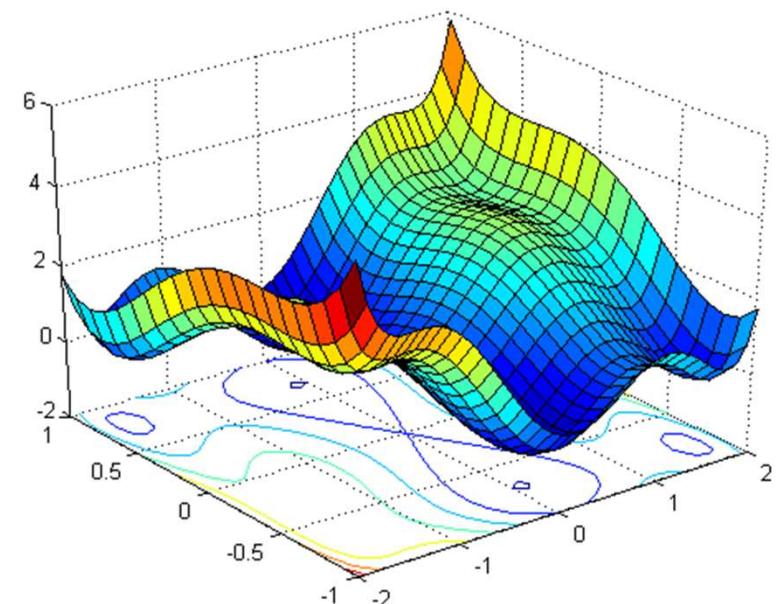
Plots the difference between the estimated and true value of a variable

Also known as error / cost function

In general the goal of any optimization algorithm is to minimize the loss function

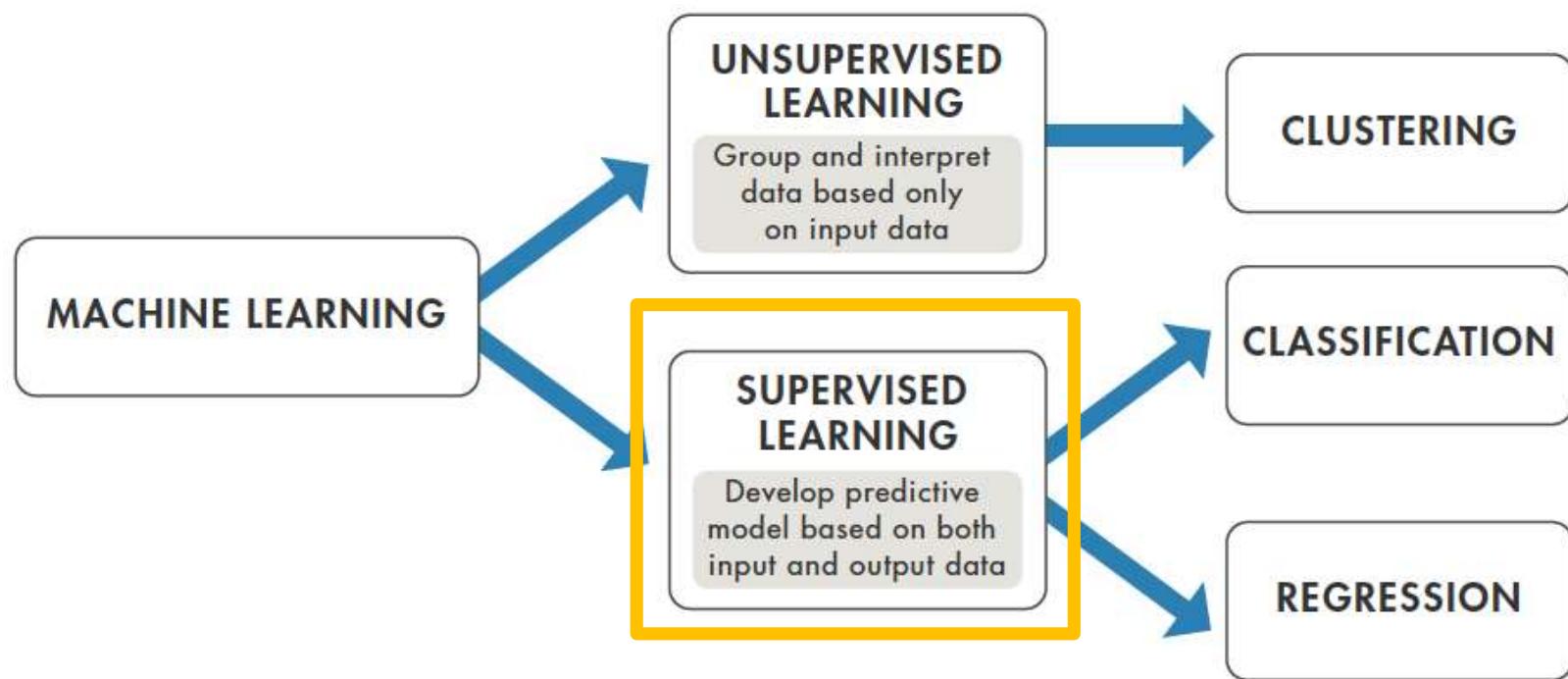
Different ways to calculate Loss:

- Mean squared error
- Log-loss
- Likelihood loss



<https://algorithmia.com/blog/introduction-to-loss-functions>

Types of Learning



Supervised Learning

Trains a model on known input and output data so that it can predict future outputs

Classification techniques

- Predict discrete responses (e.g. is a tumor is cancerous or benign).
- classify input data into categories.
- typical applications include medical imaging and speech recognition.

Regression techniques

predict continuous responses (e.g. changes in metabolic demand with exercise)

- engineering application example is electricity load forecasting.

Supervised Learning (cont.)

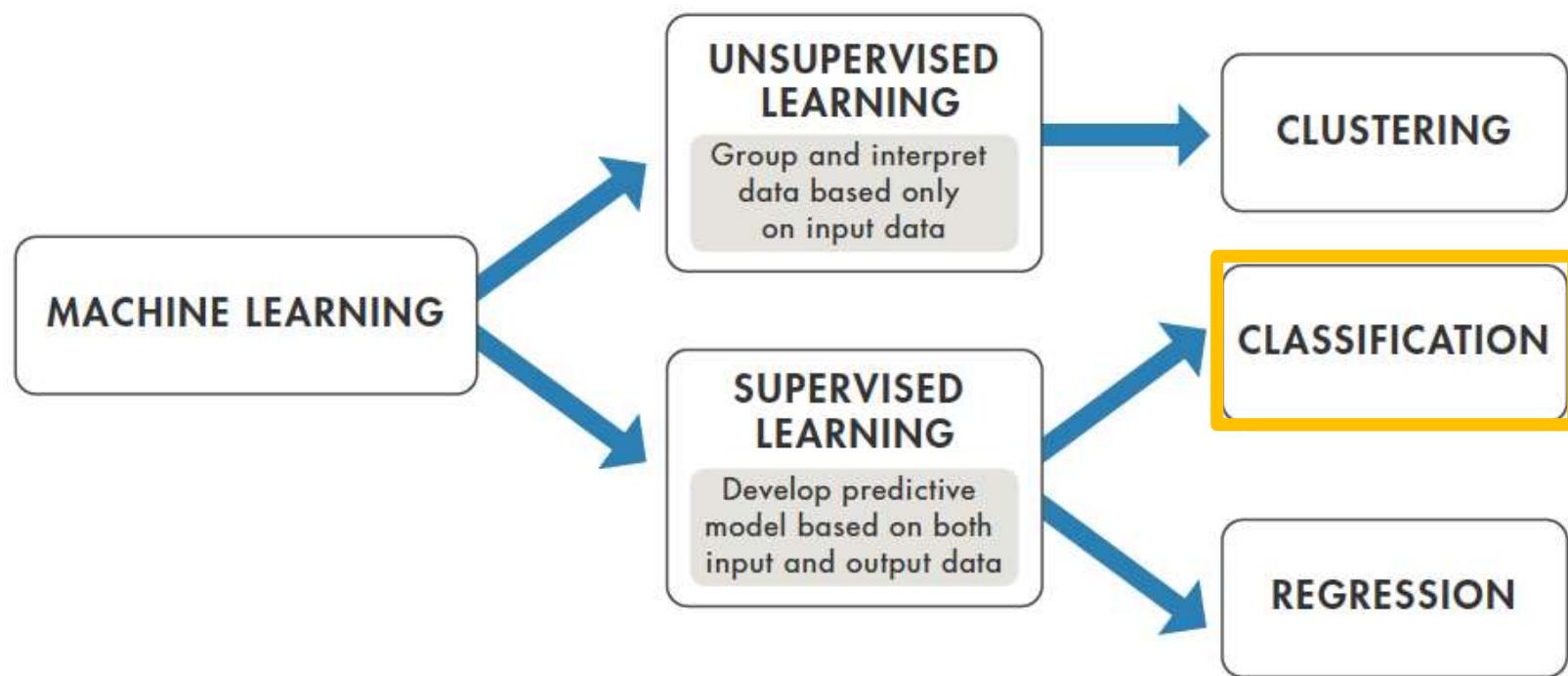
When training is complete, the classifier is applied to a test set where it performs its designed function to determine the most likely condition based on a given data pattern.

It is only the classification error that occurs during the testing phase that really matters.

A common failure of classifiers is that they perform well on the training set (after training) ... but then do poorly on the test data aka their real world application.

Such classifiers are said to generalize poorly and/or be over fit

Types of Learning



Classification

- Determining a disease or condition from a range of measurements or other diagnostic data is an application of classification.
- Classification is applied to a pattern of descriptors. It finds a class which best fits each pattern.
- Classification attempts to associate a pattern of input variables with either a specific class or another variable.
 - If the output is a variable, the analysis is referred to as “regression”
 - If the output is a discrete number identifying a specific class, the analysis is referred to as “classification.”

Classifiers

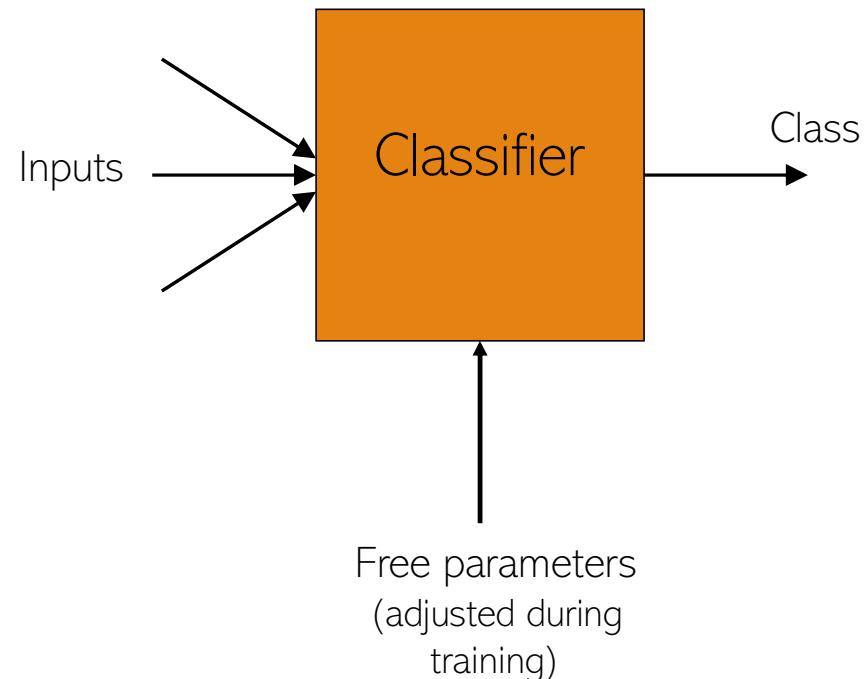
Classifiers establish a relationship between an input pattern and a discrete output

they can be viewed as mathematical functions and classifier development can thought of as function approximation

For the pattern of a number of inputs, it determines the mostly likely condition associated with that pattern.

The inputs can have any value, but typical output values are ± 1 or 0

Classification is done using two basic strategies:
supervised and unsupervised learning.



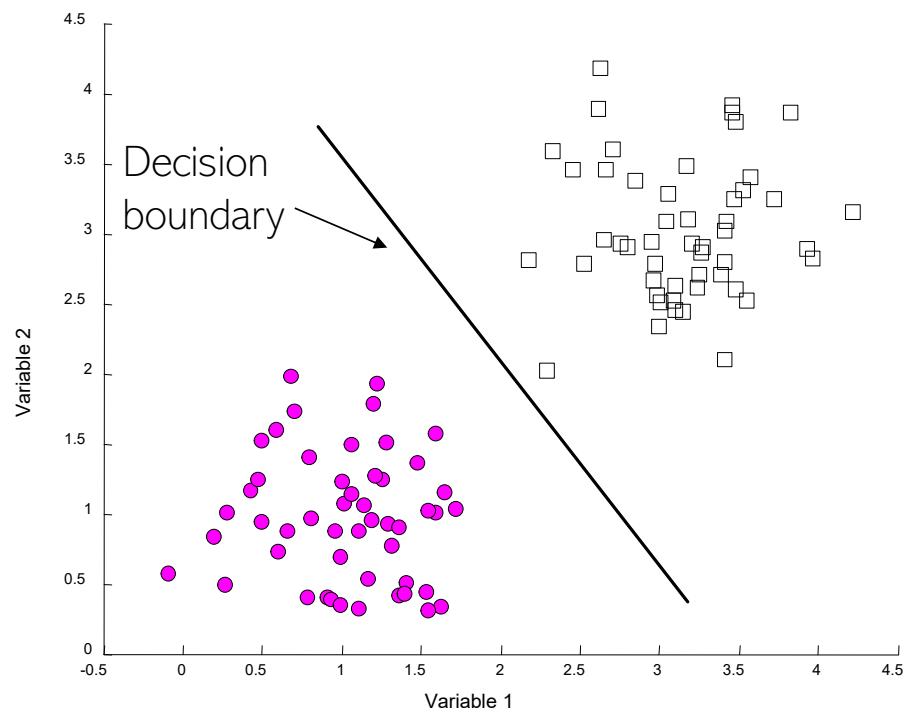
The Classification Problem

This figure shows a graphical representation of a typical classification problem.

There are 2 classes and 2 descriptive variables.

The 2 classes can be easily identified from the 2-variable scattergram.

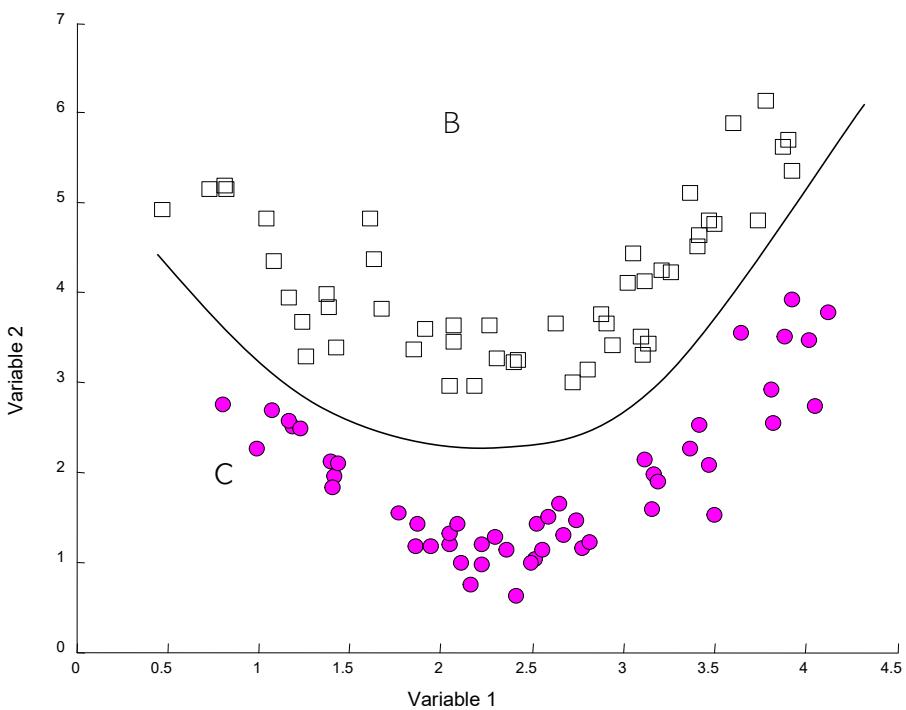
A decision boundary separates measurement patterns associated with the two classes.



Nonlinearly Separable

Data is still separable but requires a more complex shape to do so

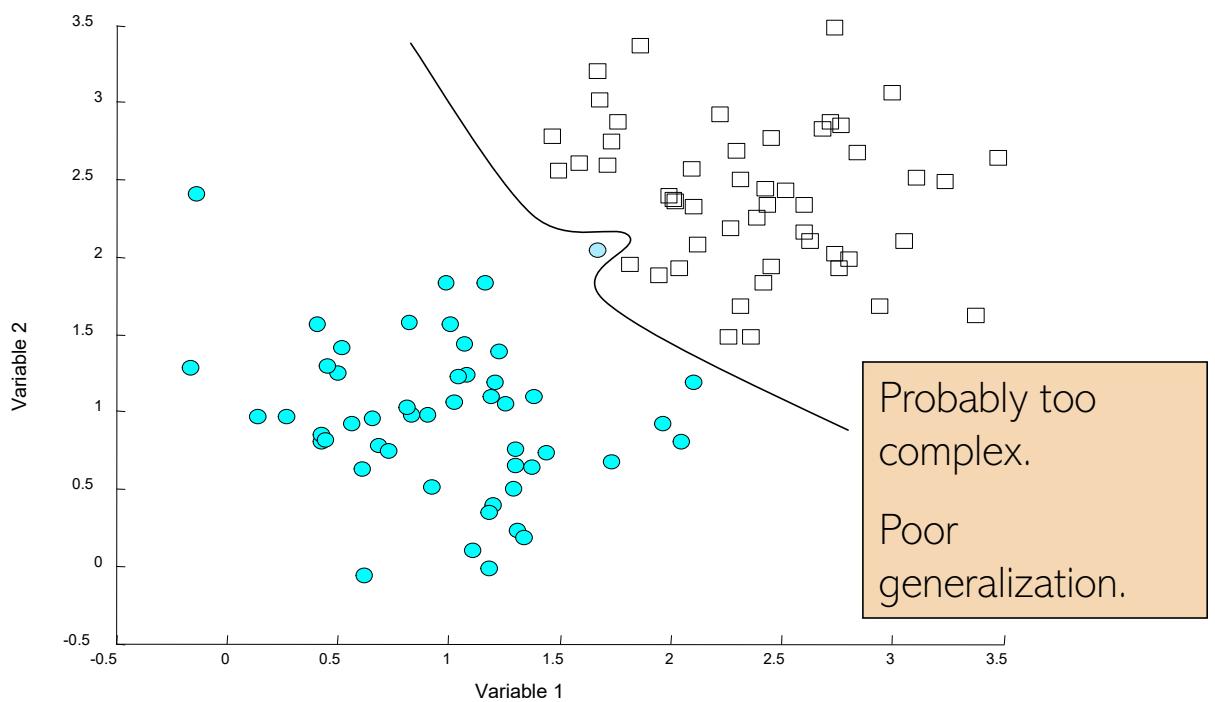
These two classes can still be correctly identified, but a curve is needed to separate them.



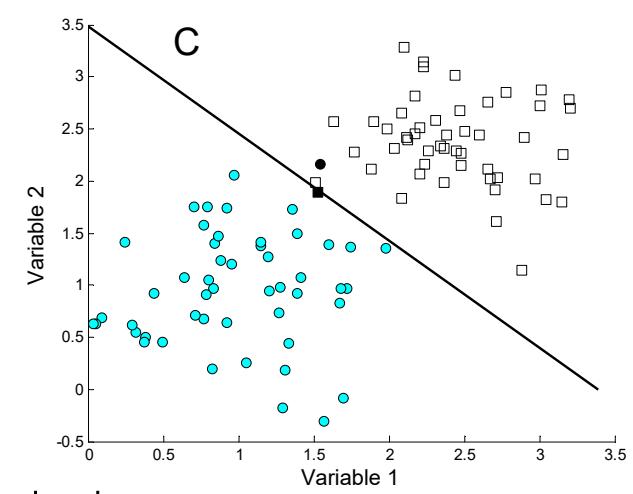
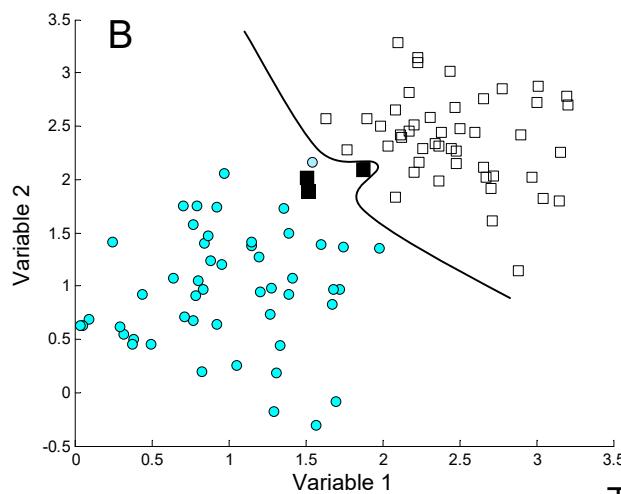
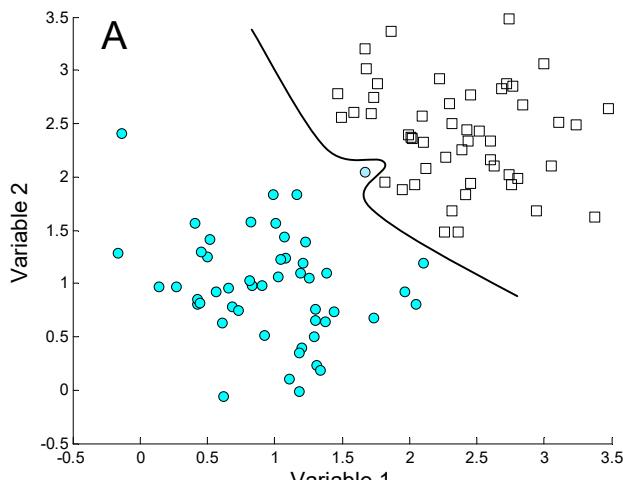
More complicated...

The two classes overlap somewhat, and a very complicated boundary is required to separate them.

If this is a training set, the boundary shown is unlikely to generalize well.



Overtraining



Three errors occur when this complicated boundary is applied to a test set.

A simple straight line produces only two errors.

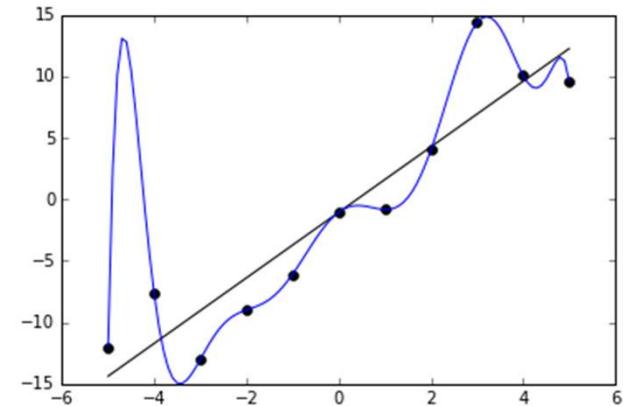
Over training

Algorithm works very well with training set
but doesn't work on other data

Occurs when there is too little data or the
data is inhomogeneous

Over fitting

- Model contains more variables than can be explained by the data
- Model follows data too closely and cannot be generalized and used for other data sets



https://en.wikipedia.org/wiki/Overfitting#/media/File:Overfitted_Data.png

Machine Capacity

The complexity of the boundary is determined by the classification algorithm.

Since classification is a form of machine learning, classification algorithm complexity is termed machine capacity.

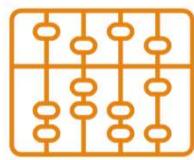
If capacity is too large for the data, the classifier will overtrain. It will perform well on the training set, but will not generalize well and perform poorly on the test set.

A machine with too little capacity will show excessive errors in training and sub-par performance in classifying the test set.

Lecture 12

TAYLOR DEVET MASC.
PHD. CANDIDATE BIOLOGICAL AND BIOMEDICAL ENGINEERING
MCGILL UNIVERSITY
SHRINERS HOSPITAL FOR CHILDREN

Todays Aims...



Linear Classifiers



Unsupervised Learning



Neural Networks

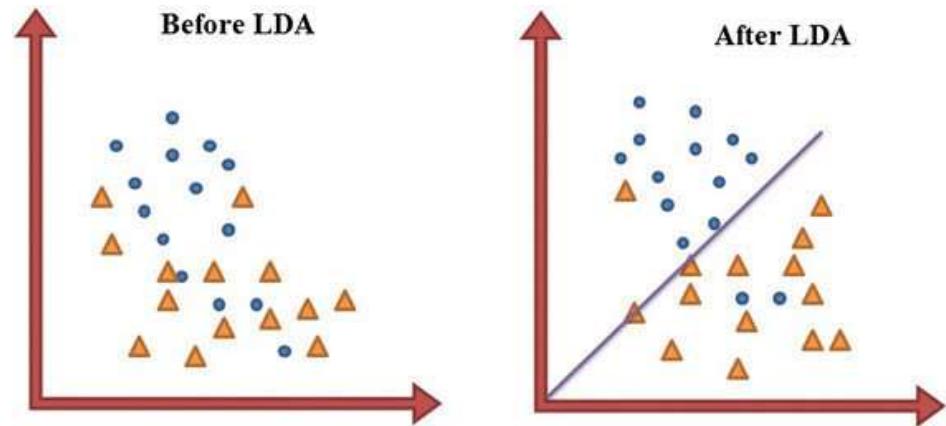
Linear Classifiers

Linear Discriminators

Linear classifiers use decision boundaries that are linear

- straight lines for two variables
- planes for three variables
- hyperplanes for four or more variables

These classifiers produce a single boundary so they can separate only two classes at a time.



https://www.researchgate.net/publication/288002528_Data_mining_EEG_signals_in_depression_for_their_diagnostic_value/figures?lo=1&utm_source=google&utm_medium=organic

Linear Discriminators

The class predicted by a linear discriminator is given by the output of a linear equation:

$$y = \sum_{i=1}^M x_i w_i + b$$

where M is the number of input variables

x_i are the input variables

w_i are the weights → orientation of separation plane/line

b is the bias or offset. → determines location of separation place/line

Linear Classifiers (cont)

The output of the linear classifier to any set of input variables is wholly determined by the weights w_i and bias, b .

$$y = \sum_{i=1}^N x_i w_i + b$$

The bias, b , can be included as one of the weights, by adding an input $x(M+1)$ that is always 1.0
(i.e., the inputs become: $x_i = [x_1, x_2, x_3, \dots, x_M, 1]$)

$$y = \sum_{i=1}^{M+1} x_i w_i = \mathbf{X}\mathbf{w} \quad \text{where } x_{(M+1)} = 1$$

$$\text{Essentially: } w_{(M+1)} = b$$

Linear Classifiers (cont)

If $y > 0.5$, then the classifier is predicting that the input data belong in class 1. Class 1 = 1

If $y \leq 0.5$ the classifier is predicting that the input data belong to class 0. Class 0 = 0.

“Class 0” and “class 1” are arbitrary names for the two classes. They stand for two categories such as “normal” and “diseased” or “malignant” and “benign.”

Linear Classifiers (cont)

The weights (and bias) constitute the free parameters of this classification method.

Sometimes linear classifiers are optimal

Linear classifiers can be easily implemented and quickly trained.

Training Set Class Identification

The known class is specified by variable d where:

$$d = \begin{cases} 0 \text{ or } -1 & \text{Class 0} \\ 1 & \text{Class 1} \end{cases}$$

In the case where there is more than two classes, d is usually a vector of 0's and 1's

$$d = \begin{cases} 1000 & \text{Class 0} \\ 0100 & \text{Class 1} \\ 0010 & \text{Class 3} \\ 0001 & \text{Class 4} \end{cases}$$

This unusual structure makes programming easier as will become apparent later.

- These identifiers are an essential component of any training set.

Training Methods

Least Squares Algorithm

The sum of squares error between the output of the linear classifier and the correct class is given by:

$$\varepsilon^2(w) = \sum_{i=1}^{N+1} (d_i - x_i^T w)^2$$

In matrix notation:

$$\varepsilon^2(w) = (\mathbf{d} - \mathbf{X}\mathbf{w})^T (\mathbf{d} - \mathbf{X}\mathbf{w})$$

Minimizing $sse(w)$ is done by differentiating $sse(w)$ with respect to w , setting to zero, and solving for w . If $\mathbf{X}^T\mathbf{X}$ is nonsingular, a unique solution is:

$$w = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{d}$$

Example 1

Generate a test set consisting of two Gaussian distributions with centers 3.0 standard deviations apart.

The test set should include the correct classification vector d .

Apply the least squares method to classify these two data sets and then plot the results.

Plot the two classes as circles and squares and plot any misclassified data points filled in black.

Finally plot the decision boundary produced by the linear classifier.

Plotting the Results

The equation for the boundary can be determined from the basic equation and weights, w_i . The boundary occurs at $y = 0.5$:

$$y = \sum_{i=1}^{N+1} x_i w_i = \mathbf{X}\mathbf{w} = 0.5$$

For two inputs, this equation becomes:

$$w_1 x_1 + w_2 x_2 + w_3 = 0.5$$

Solving for x_2 in terms of x_1 :

$$(y = mx + b) \quad x_2 = \left(\frac{-w_1}{w_2} \right) x_1 - \frac{0.5 - w_3}{w_2}$$

This equation is used in `linear_eval` to plot the boundaries

linear_eval: Routine to Evaluate Performance of a Linear Classifier

```
function [sensitivity, specificity] = linear_eval(X,d,w)
% Function to evaluate performance.

% X      data set
%
% d      Correct classification
%
% w      Classifier weights (totally defines classifier)

threshold = .5;                      % The decision threshold

% Initialize counters, etc.

[r,c] = size(X);                     % Determine data set size

y = X*w;                             % Evaluate the output

% Plot the results
```

Plotting and Counting the Four Possibilities

```
% Assumes Class 0 is 0 and Class 1 is 1.           % 2) False positive
Threshold = 0.5

% Evaluates each point for all four
possibilities

for i = 1:r
    if d(i) > threshold & y(i) > threshold
        plot(X(i,1),X(i,2), 'sqk')
        tn = tn + 1;
    %1) True negative
    elseif d(i) > threshold & y(i) <=
threshold
        plot(X(i,1),X(i,2), 'sqk')
        fp = fp + 1;
    % 2) False positive
    elseif d(i) <= threshold y(i) <= threshold
        plot(X(i,1),X(i,2), 'ok');
        tp = tp + 1;
    % 3) True positive
    elseif d(i) <= threshold & y(i) >
threshold
        plot(X(i,1),X(i,2), 'ok');
        fn = fn + 1;
    % 4) False negative
    end
end
```

Plotting the Linear Decision Boundary

```
clf; hold on;

y_lim = get(gca, 'YLim');                                % Used to reset axis
x_lim = get(gca, 'XLim');

% Plot decision boundary

x1 = [min(X(:,1)),max(X(:,1))];                         % Construct x1 to span x1
x2 = -w(1)*x1/w(2) + (-w(3)+.5)/w(2);                  % Calculate x2 using eq.

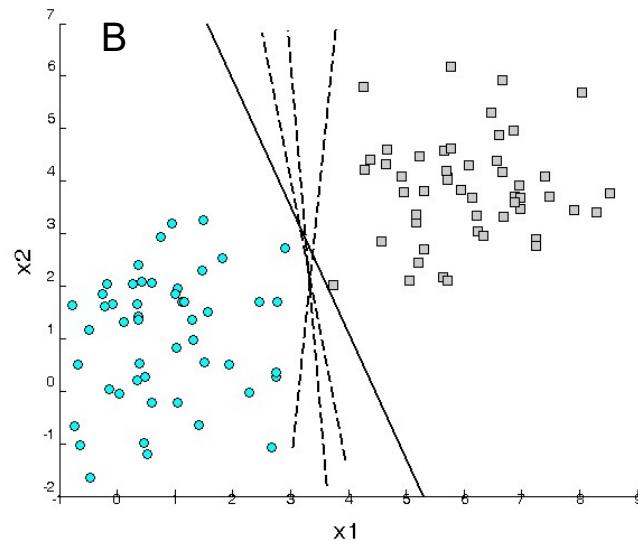
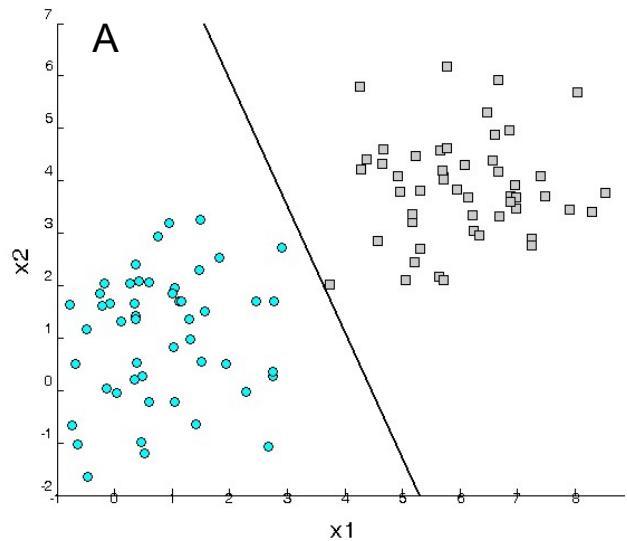
plot(x1,x2,'k');                                         % Plot boundary line
axis([x_lim, y_lim]);                                     % Limit axis

%

% Evaluate sensitivity and specificity

specificity = (tn/(tn+fp))*100; % Specificity
sensitivity = (tp/(tp+fn))*100; % Sensitivity
```

Example 1 Results



The boundary found is shown on the left accurately separates the two classes in this training set.

But many boundaries can do that as shown on the right (an infinite number).

What is the best boundary? The one that will generalize best.

Sensitivity and Specificity

Sensitivity: Percent correct detections

$$Sensitivity = 100 \frac{True\ Positives}{True\ Positives + False\ Negatives} = 100 \frac{True\ Positives}{Total\ Abnormal}$$

Specificity: Percent correct rejections

$$Specificity = 100 \frac{True\ Negatives}{True\ Negatives + False\ Positives} = 100 \frac{True\ Negatives}{Total\ Normal}$$

The Receiver Operation Curve (ROC)

Sometimes it is possible to vary the decision boundary to increase or decrease the detection of abnormal values.

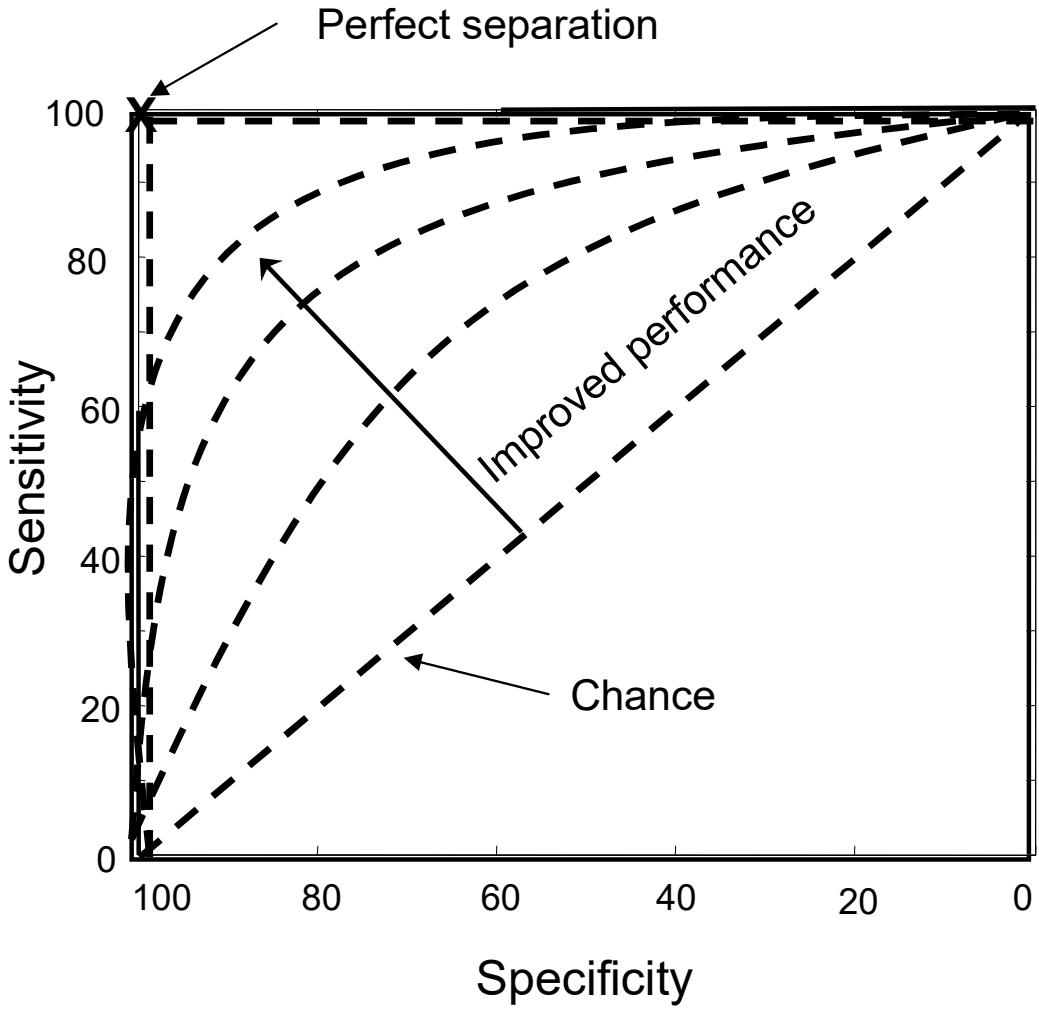
Increasing the detection of true positives will usually increase the number of false positives and lower the number of true negatives.

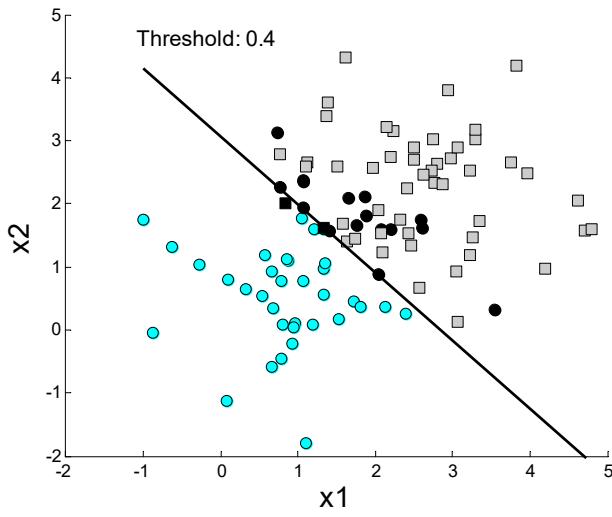
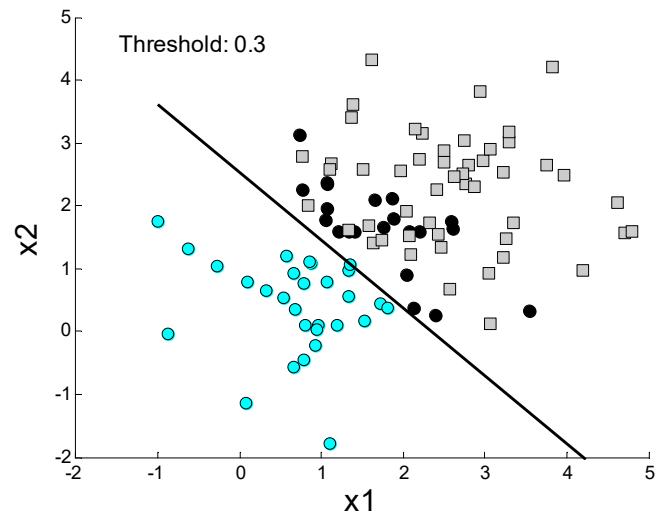
There is a trade-off between sensitivity and specificity.

A curve showing this tradeoff between sensitivity and specificity is called the ROC (receiver operator characteristics) curve.

Example ROC Curve

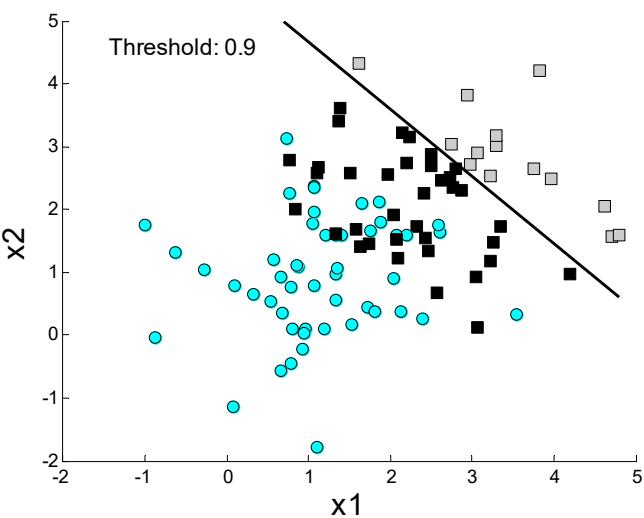
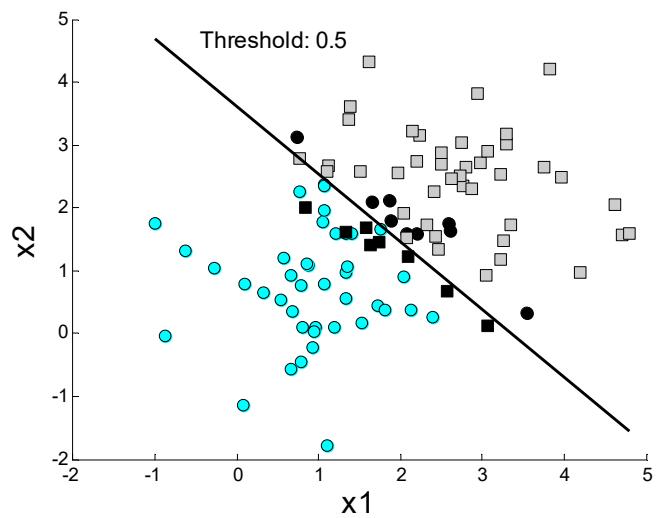
Note that Specificity is plotted in reverse

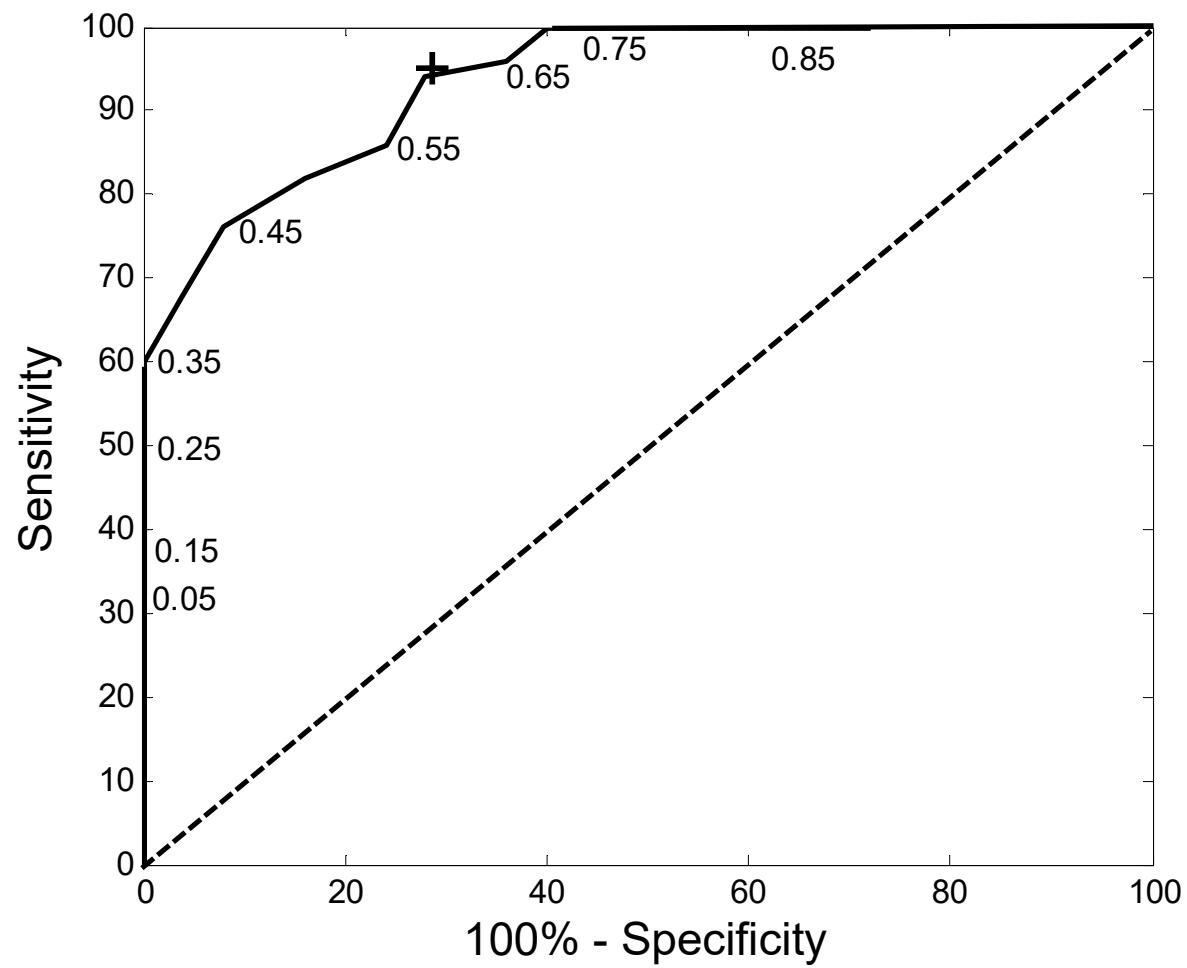




Circles are Diseased

Moving Threshold adjusts
sensitivity specificity tradeoff





The data consisted of two closely spaced Gaussian distributions,
The classifier was linear least squares.

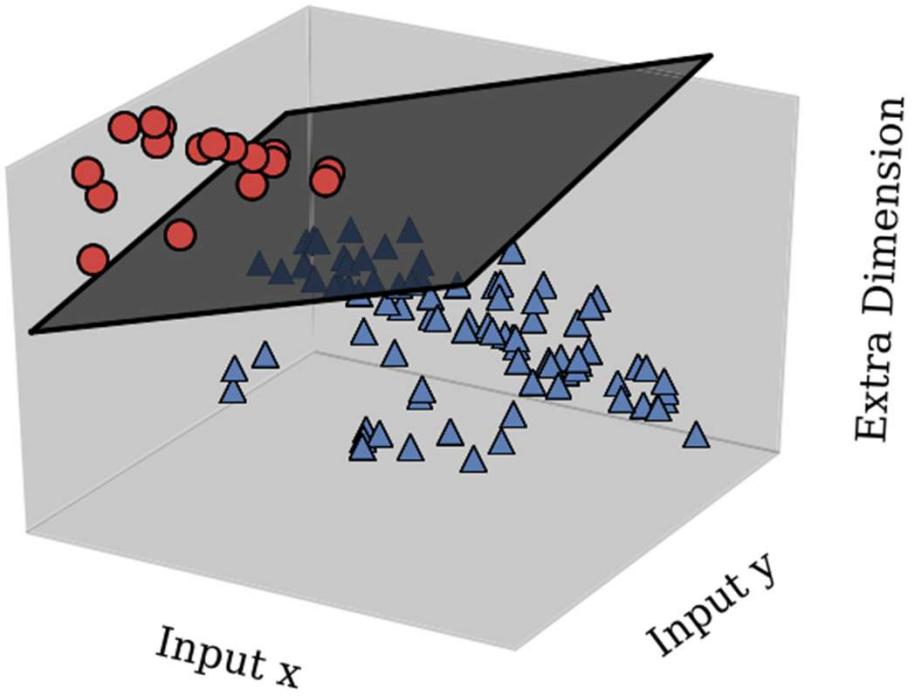
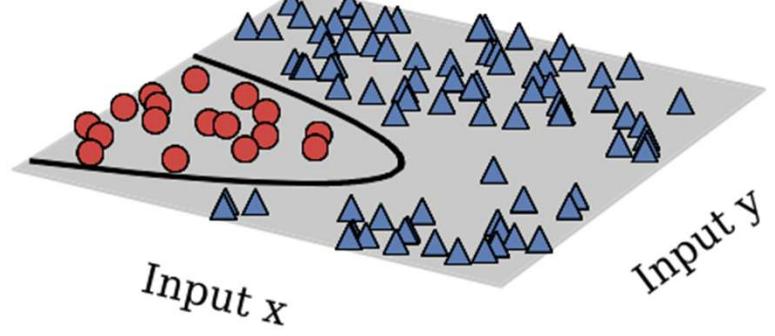
Higher Dimensions – Kernel machines

Many classification problems involve data that are separable, but not by a single straight line

In more complex data sets, it is still possible to separate the classes using a linear boundary if the data are transformed into a higher-dimensional space.

Cover's Theorem (1965)

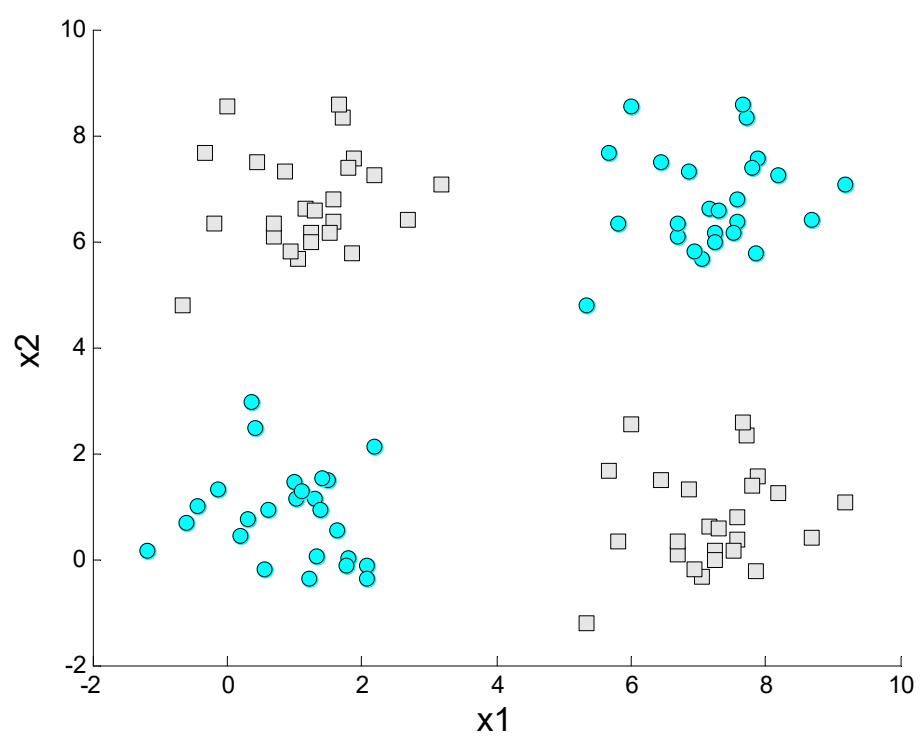
- if the number of dimensions is high enough, you can always find a linear boundary that will separate the data without error
- hyperplane



<https://i.stack.imgur.com/fylvr.png>

Example

A two-class data set that is clearly separable, but not by a single linear boundary.



Example

Classify the data in the last slide using the least squares linear classifier after the data are transformed into a higher dimensional space.

Solution: The quadratic kernel can be used to separate these data.

Take it from 2 features (x_1, x_2) to 5:

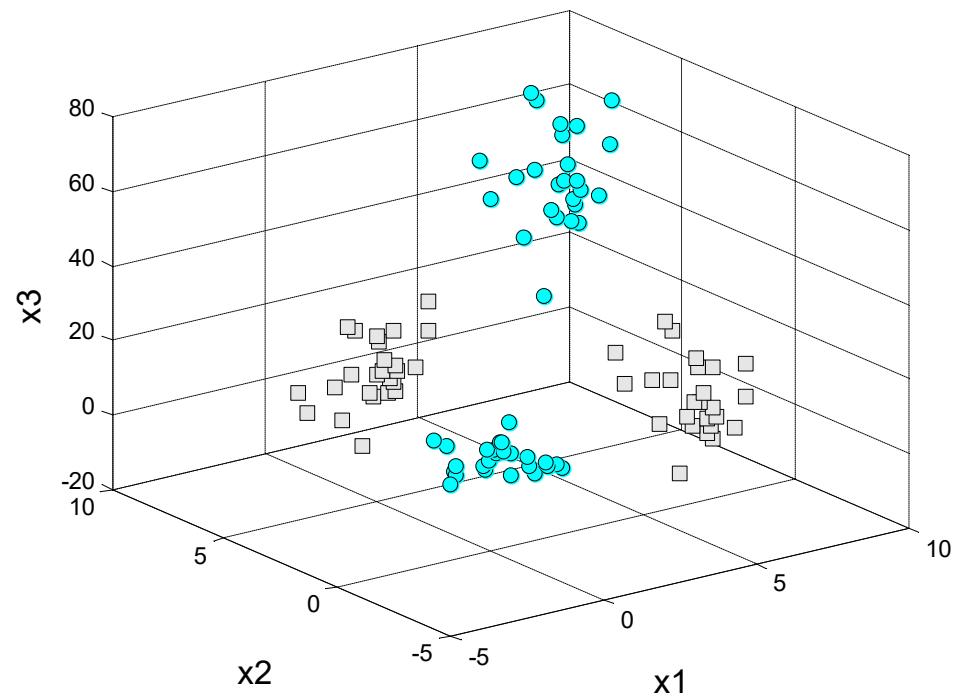
$$x = x_1 + x_2 + x_1 * x_2 + x_1^2 + x_2^2$$

Since the data are so widely spaced it is only necessary to use the cross product term, $x_1 * x_2$, to obtain perfect separation.

This allows us to implement the classifier in three dimensions (as opposed to five) which is easier to demonstrate.

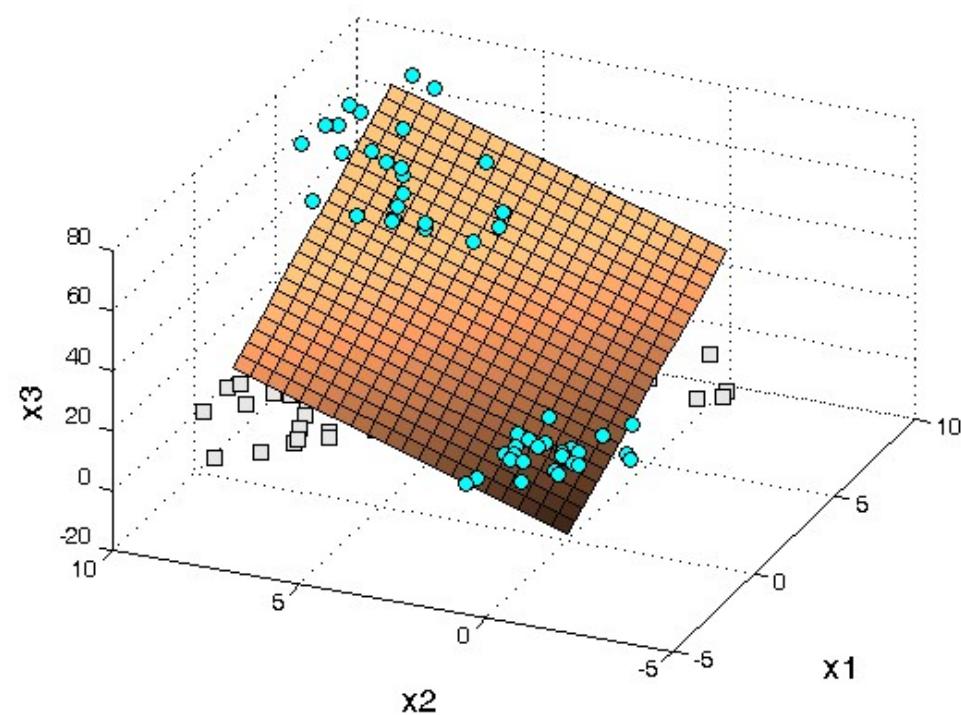
Example

The data in the previous slide transformed into three-dimensional feature space by adding a cross-product term.



Example Results

A plane gives perfect separation of the transformed data set.



Support Vector Machines

Support Vector Machines

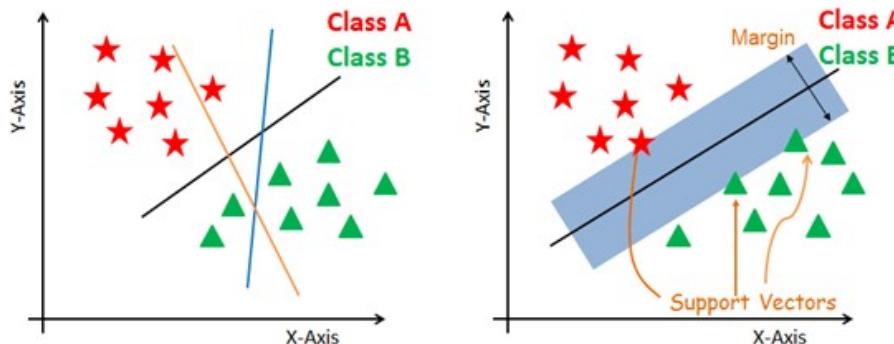
- One of the most robust prediction methods
- Uses binary linear classification
- SVM maximize the distance between the critical data points (support vectors)
- Looks at points in space and tries to maximize the gap between groups
- New data is sorted based on which side of the gap they are on
- Can use kernels to map in higher dimensions'
 - Kernel – similarity function

Support Vector

The data points that lie closest to the decision surface/line/plane

They are the points that are the most difficult to classify

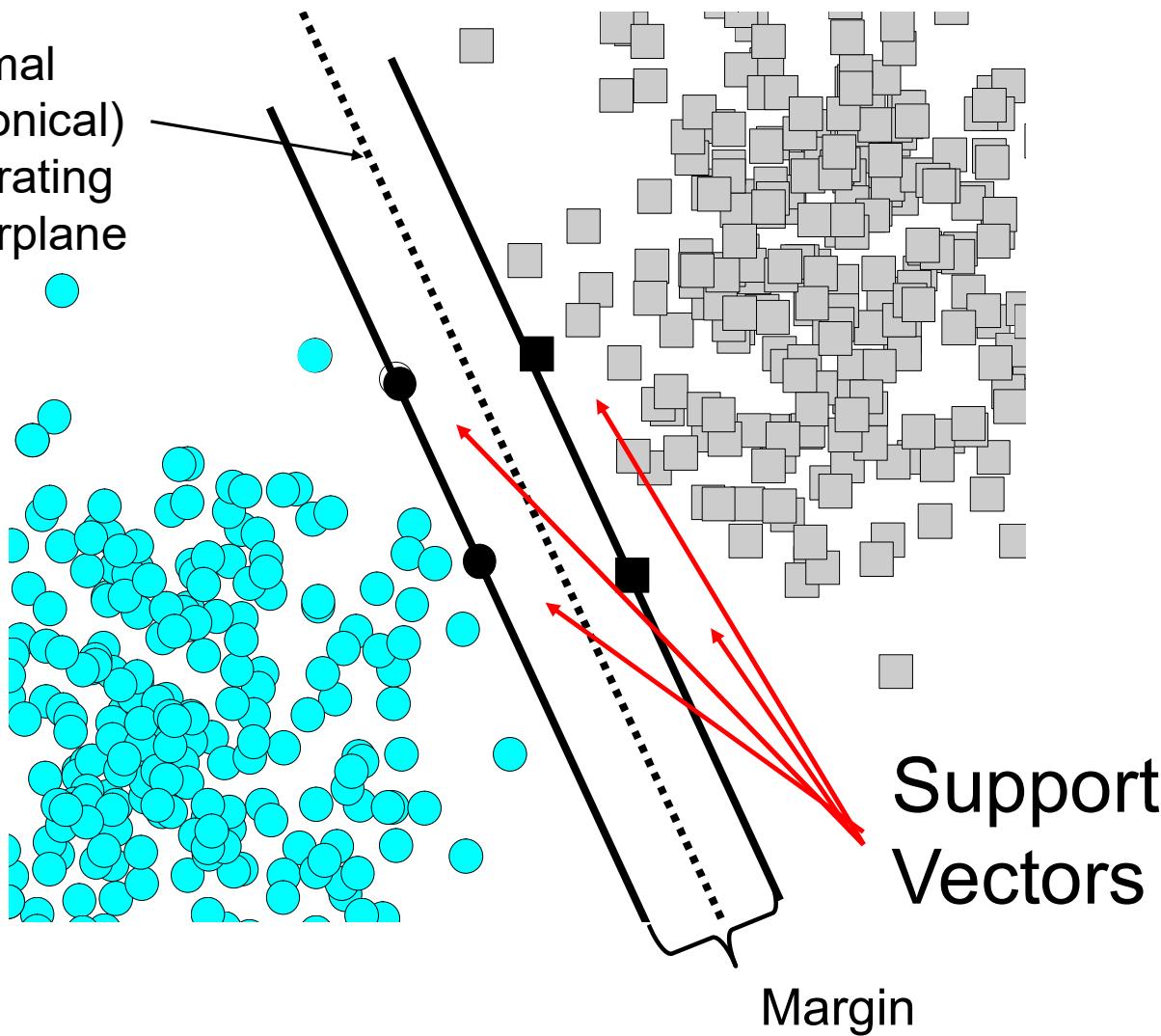
They have a direct impact on the location of the decision boundary



<https://www.datacamp.com/community/tutorials/svm-classification-scikit-learn-python>

Optimal
(canonical)
separating
hyperplane

Previous
arguments will
hold even if
points overlap



Determining the margin

To determine the equation for the margin, M, note that the distance of any hyperplane, $\mathbf{x}^T \mathbf{w} + b = 0$, to the origin is:

where $\| \mathbf{w} \|$ is the norm of \mathbf{w} and is equal to:

or in matrix notation:

If the hyperplane is equal to ± 1 , then the distance to the origin is just:

Boundary Equations

For the line separating class 1: $y_i \geq 1$, the distance to the origin is:

$$d_o = \frac{(1 - b)}{\|\mathbf{w}\|}$$

For the line separating class -1: $y_i \leq -1$, the distance to the origin is:

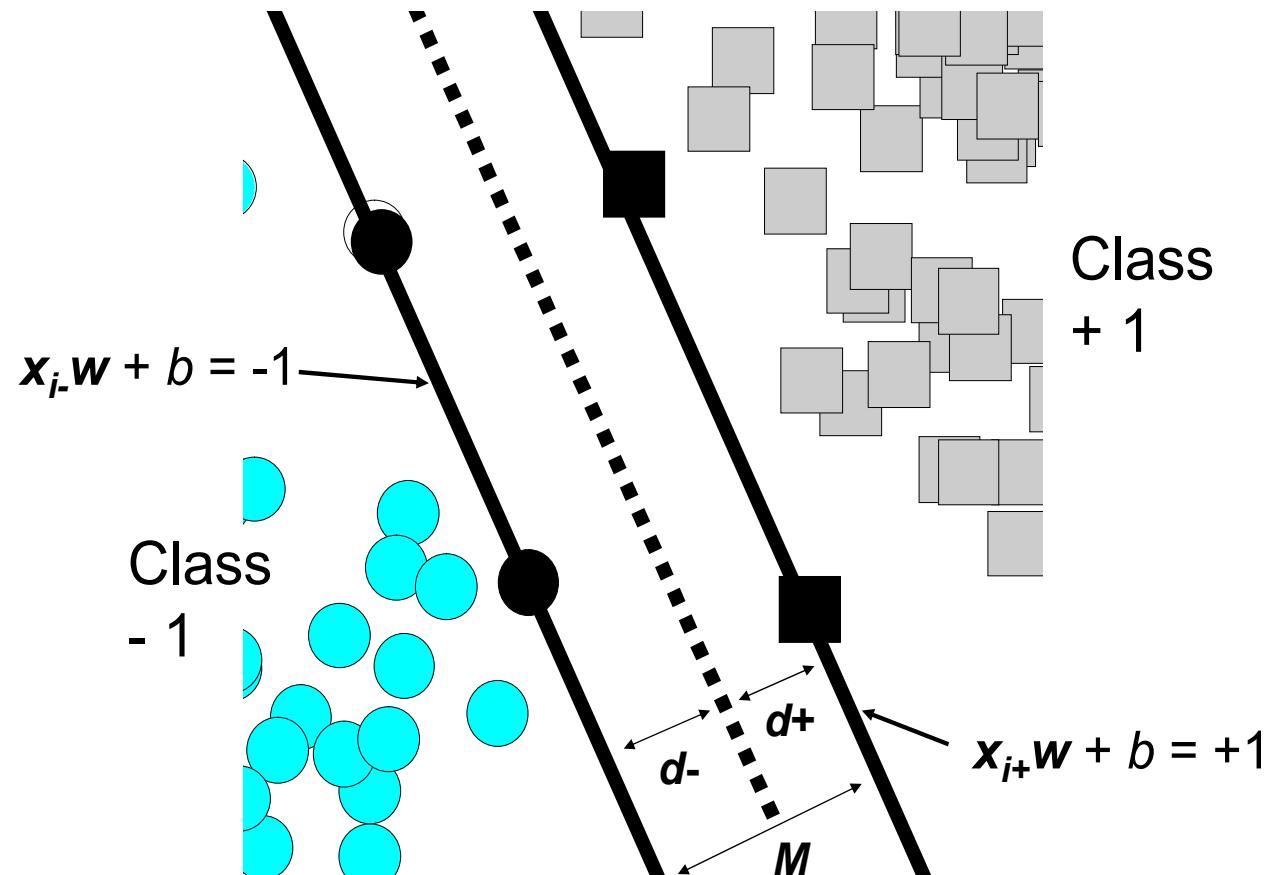
$$d_o = \frac{(-1 - b)}{\|\mathbf{w}\|}$$

The difference between the two lines, which is the margin, is obtained by subtracting:

$$M = \frac{(1 - b)}{\|\mathbf{w}\|} - \frac{(-1 - b)}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|}$$

The two lines passing through the support vectors mark the boundary for $y \geq \pm 1$ and this determines the equation for these lines.

$$M = \frac{2}{\|w\|}$$



SVM Margin Maximization

- The maximum margin is obtained by minimizing $\|w\|$ (or by minimizing $\|w\|^2$) which is equal to w^Tw and is somewhat easier to perform.
- The minimization must be done subject to the constraint imposed by $y_i(x_i w + b) \geq 1$ which ensures that the boundaries are on the correct side.
- This type of minimization problem is known as a quadratic programming or *QP* optimization problem and there are a number of routines available to solve this.

Support Classifier Algorithm Development

Find the boundaries that maximize the margins

Find the hyperplane that maximizes the margin

Constrain data points to the appropriate side of the boundary.

Classes are assumed to be ± 1 as this simplifies the mathematics.

Support Classifier Algorithm Development (cont)

Since the decision boundary is at $y = 0$ (between ± 1 , the equation for the boundary is:

$$y = \sum_{i=1}^N w_i x_i + b = \mathbf{x}_i \mathbf{w} + b = 0$$

where x_i are the input values

w is the weight vector

b is the offset or bias

(recall w and b define the classifier)

Support Classifier Algorithm Development (cont)

Since the two classes are defined by $y = \pm 1$, the value of y must be ± 1 at the closest points (i.e., the support vectors)

So the equations for the lines that go through those support vector points must be:

$$\mathbf{x}_i \mathbf{w} + b \geq 1 \quad \text{when } y = +1$$

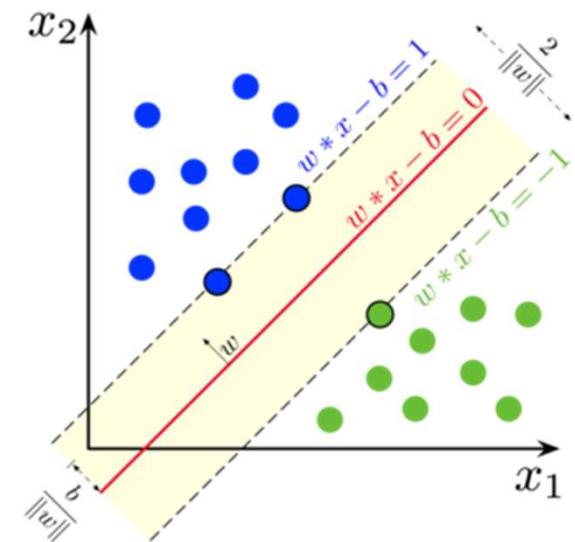
and

$$\mathbf{x}_i \mathbf{w} + b \leq -1 \quad \text{when } y = -1$$

These can be combined into a single equation:

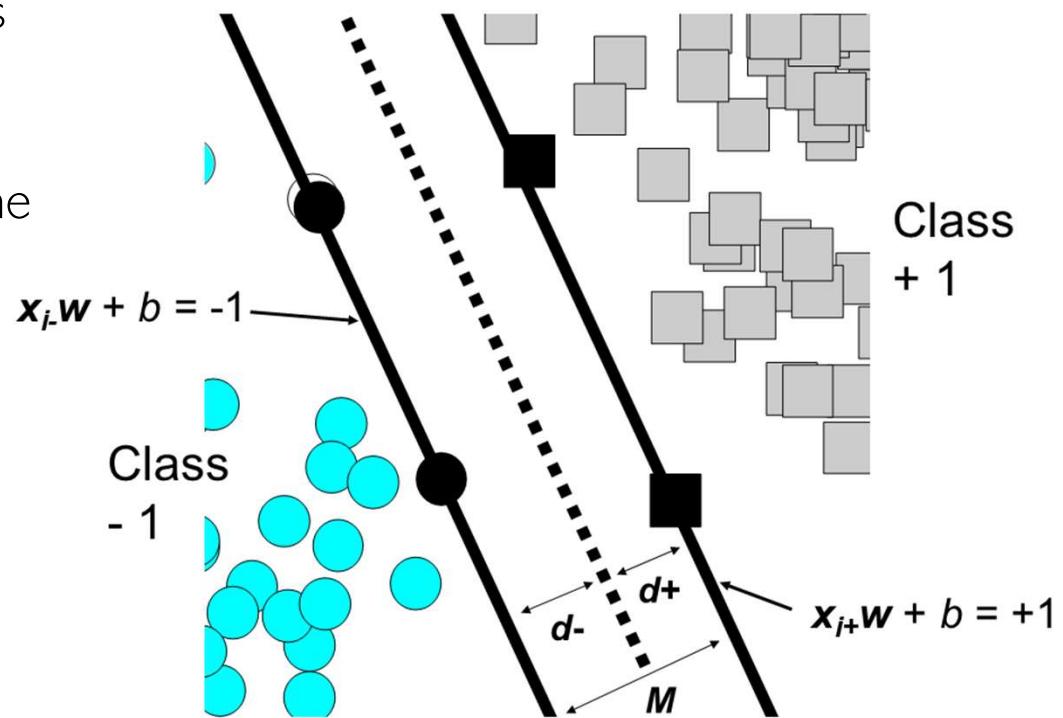
$$y_i(\mathbf{x}_i \mathbf{w} + b) \geq 1$$

This equation is interpreted in the next slide.



Support Classifier Algorithm Development (cont)

The equation simply states that w and b should be such that the two classes fall on opposite sides of the support vector lines.



Support Vector Machines

If the data are not linearly separable:

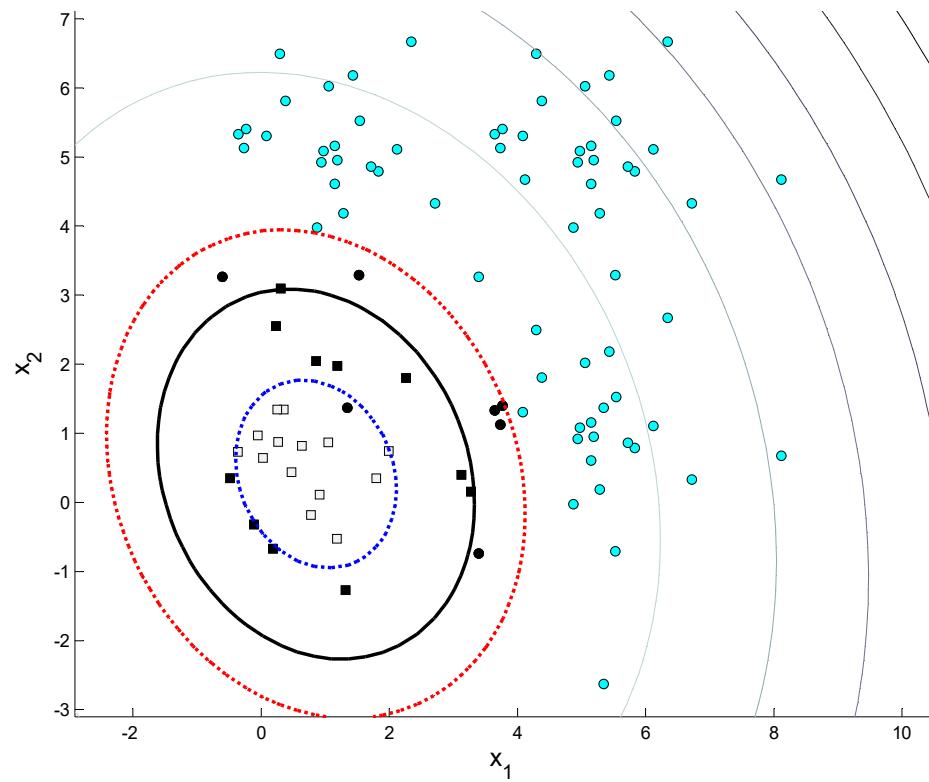
- the points overlap
- the optimization process still maximizes M ,
- relaxed constraint so that some of the points can be on the wrong side of the boundary.

This approach can be used to produce linear boundaries and is termed linear support vector machines (LSVM).

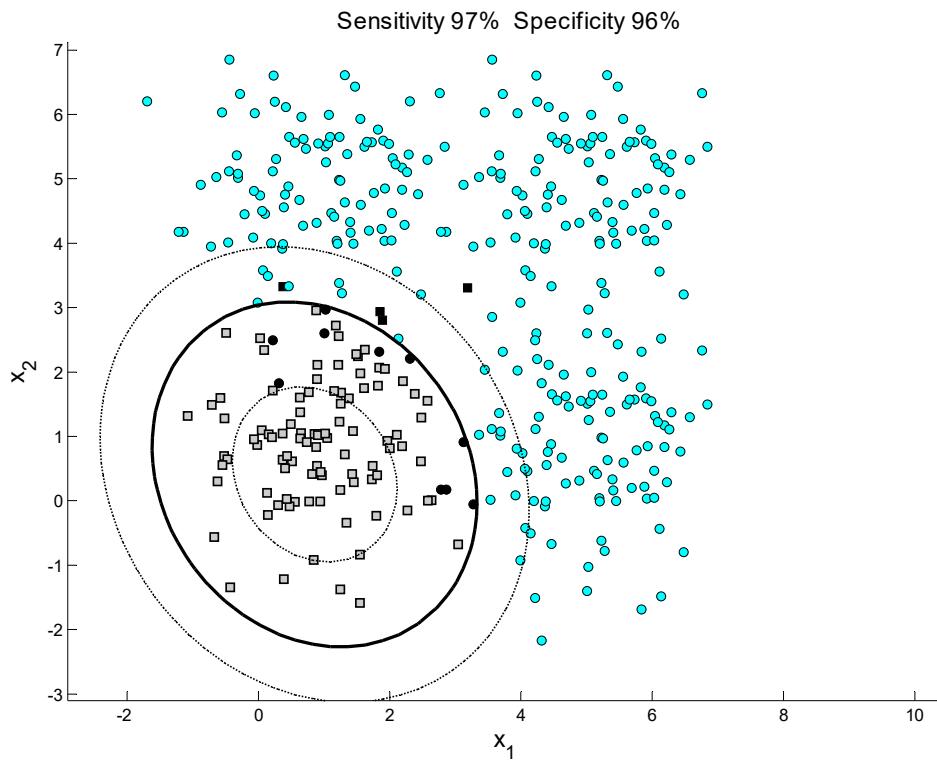
To generate nonlinear boundaries:

- a kernel can be used to transform the data into higher dimensions
- these classifiers are termed support vector machines (SVM).

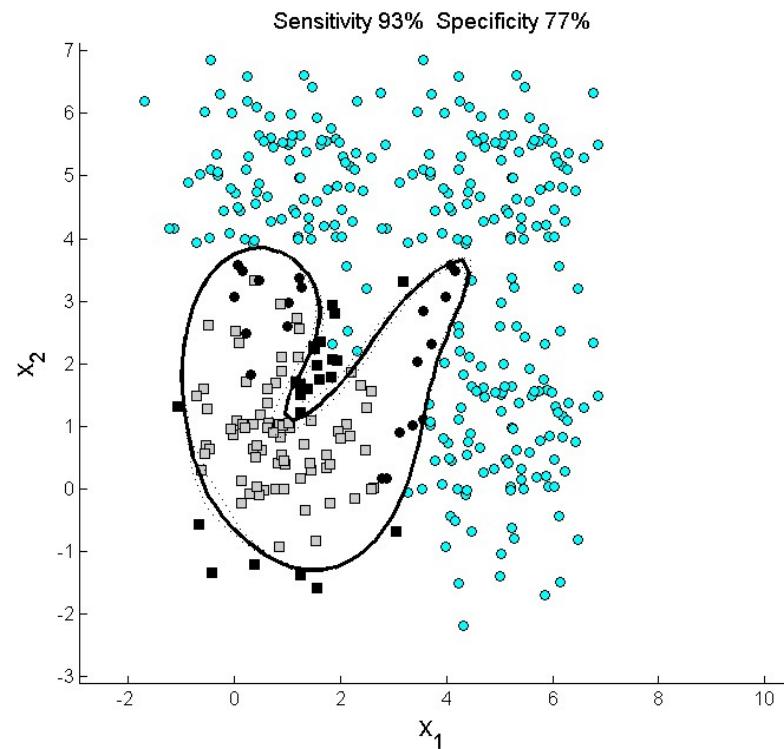
Example Kernels



Example Kernels



Example Kernels: Overtraining



Too Much Machine Capacity

The capacity of the classifier used above was increased by increasing the polynomial order to 6.

The boundary is complicated due to overtraining on the training set.

Test set performance is not nearly as good.

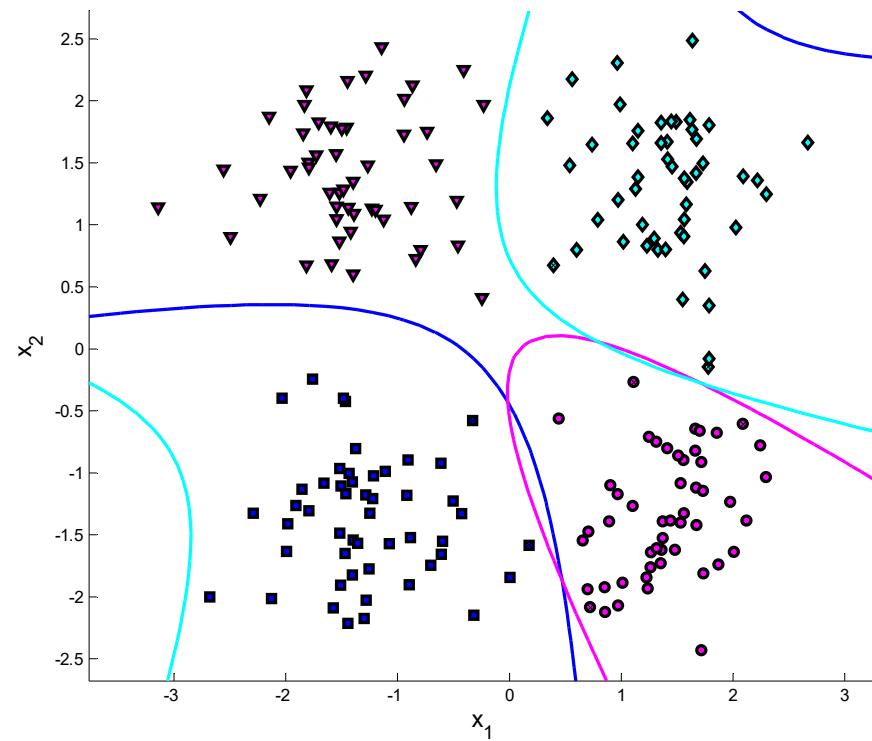
Multiple Classes

To classify data with more than two classes, implement the classifier on each class separately and combine results.

With multiple classes

- the correct classification is indicated by a matrix.
- Each class is indicated by a column which contains +1 if the associated pattern indicates that class.

Example Results: All classes are correctly identified.



K-nearest neighbour

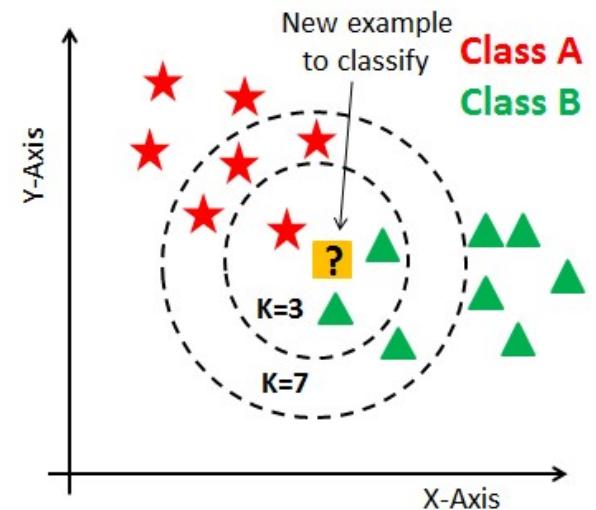
k-nearest neighbor classifier

Operates directly off the testing set and does not need prior training.

Takes each test set point and determines the distance to the k nearest training points, where k is a constant.

It then takes the class values of these nearest points and assigns the test point to the majority class.

This approach can be used for any number of classes and any number of input variables.



<https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn>

Distance Measurement

Distances between points can be measured using a variety of metrics but the most common and straightforward is the Euclidean distance. The Euclidean distance between two points is:

$$\begin{aligned}d(a,b) &= ||a-b|| \\d(a,b) &= \sqrt{(a-b)^2} \\d(a,b) &= \sqrt{(ax-bx)^2 + (ay-by)^2}\end{aligned}$$

where a and b are vectors and $||a-b||$ is the norm of the vector that results after subtraction.

Example

```
load fisheriris;
X = meas;
Y = species;
Mdl = fitcknn(X,Y, 'NumNeighbors',5, 'Standardize',1)
Mdl.ClassNames

figure();
gscatter(X(:,1),X(:,2),species)

figure();
label = predict(Mdl,X);
gscatter(X(:,1),X(:,2),label)

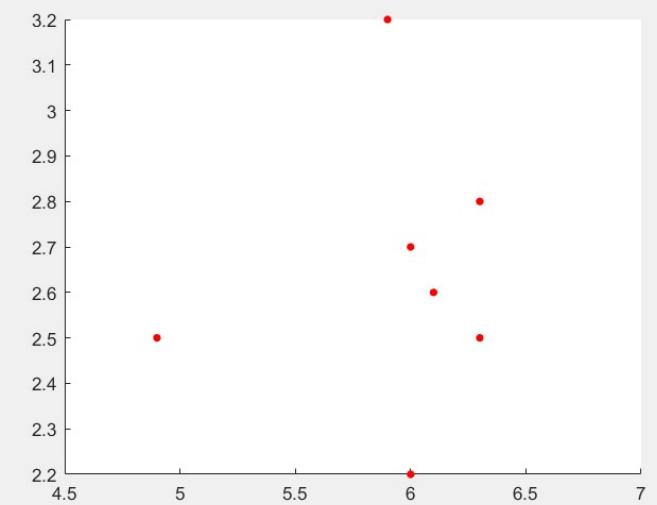
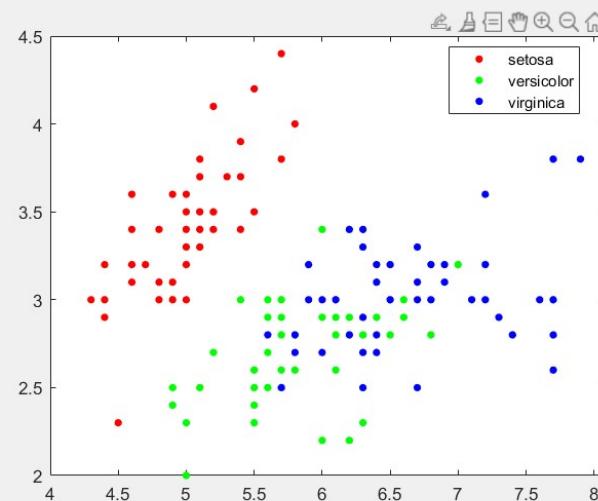
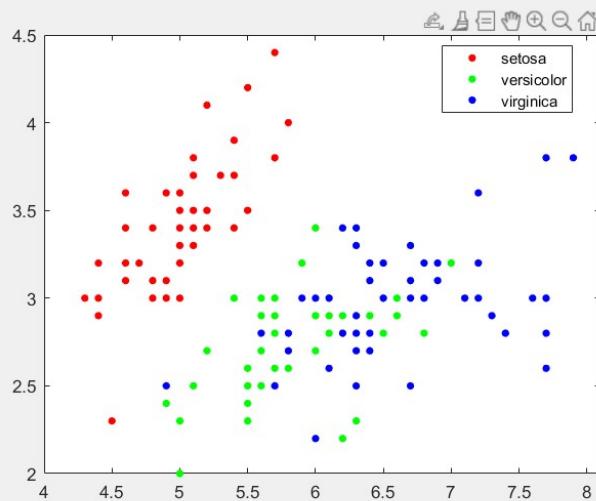
figure
hold on;

correct = 0; incorrect = 0;
CorrSet = 0; CorrVer = 0; CorrVirg = 0;
InCorrSet = 0; InCorrVer = 0; InCorrVirg = 0;
```

Example Results – k=5

Setosa Versicolor Virginica

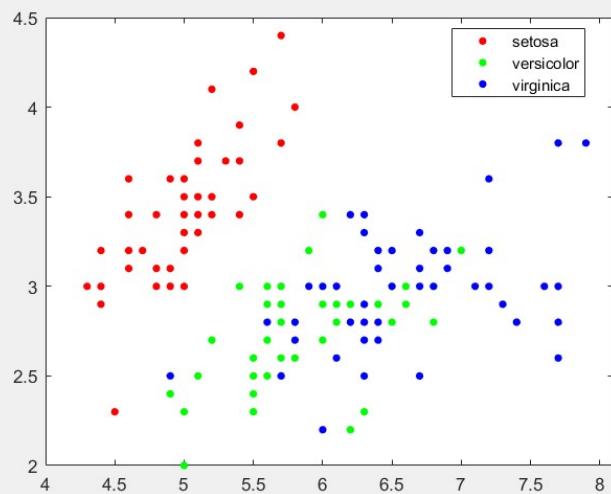
50	47	46
0	4	3



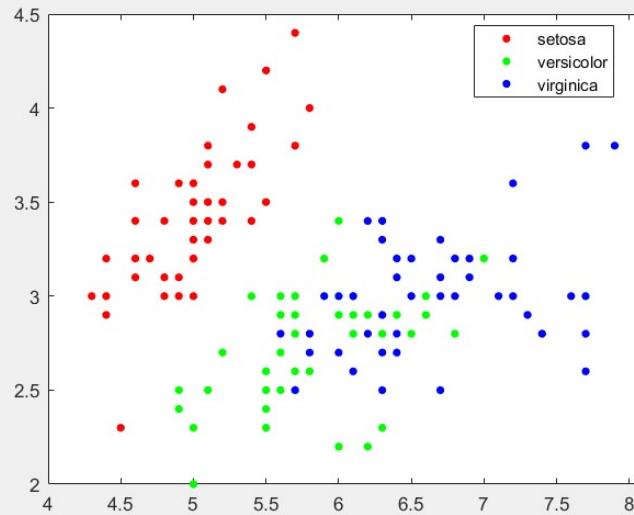
Example Results – k=20

Setosa Versicolor Virginica

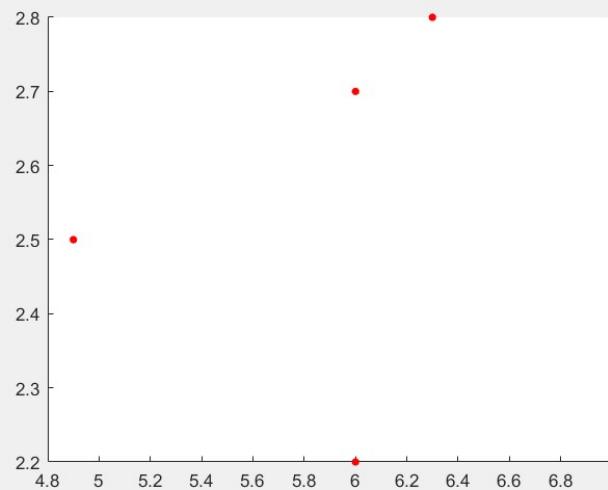
50	49	47
0	3	1



Original



Classification



Error

Confusion Matrix from Example

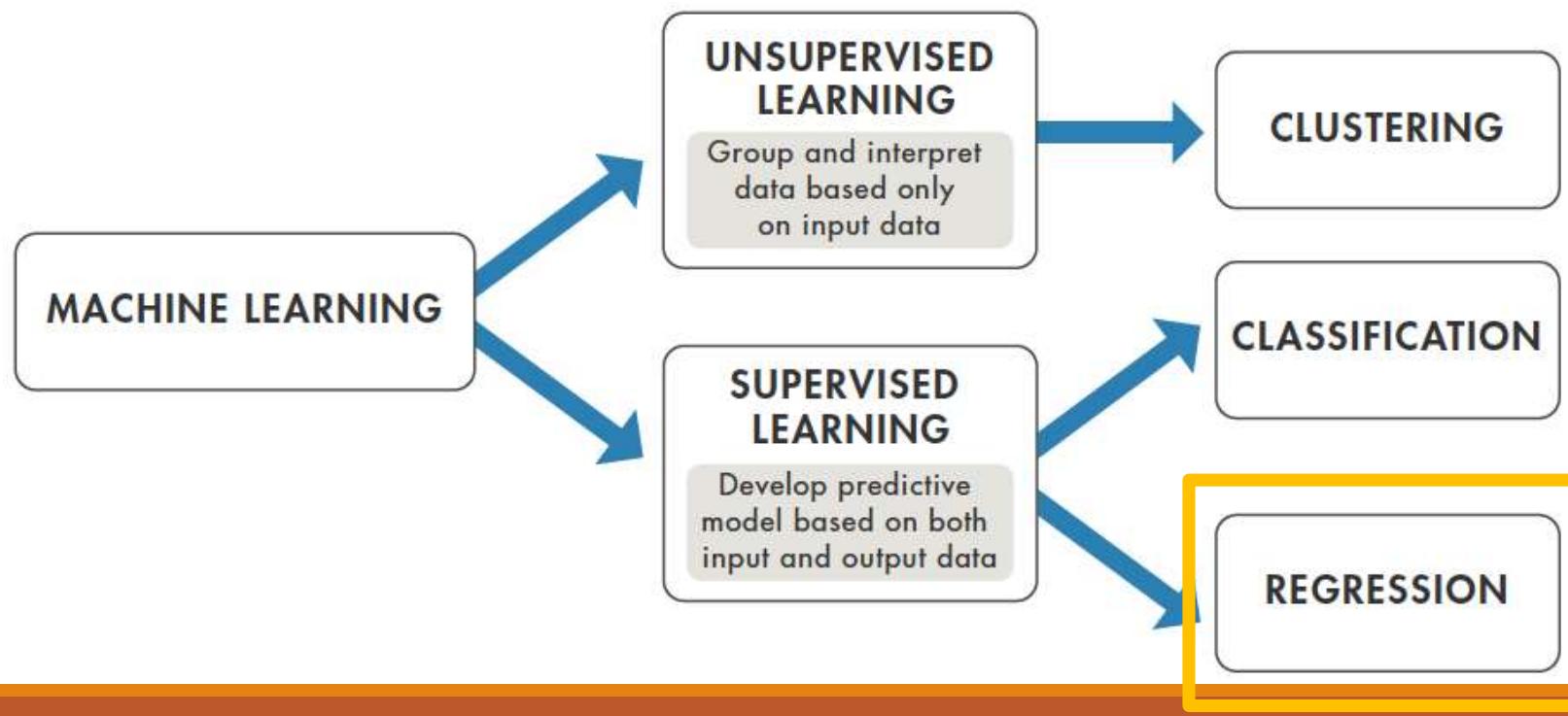
	Setosa	Versicolor	Virginica		Setosa	Versicolor	Virginica
Correct	50	47	46		50	49	47
Incorrect	0	4	3		0	3	1

The performance is slightly better when $k = 20$.

A larger k improves generality but, if k is too large, boundary points can be misclassified if they are close to a large group of the other class.

Smaller values of k can lead to misclassification due to a small number of outliers in the other class.

Types of Learning



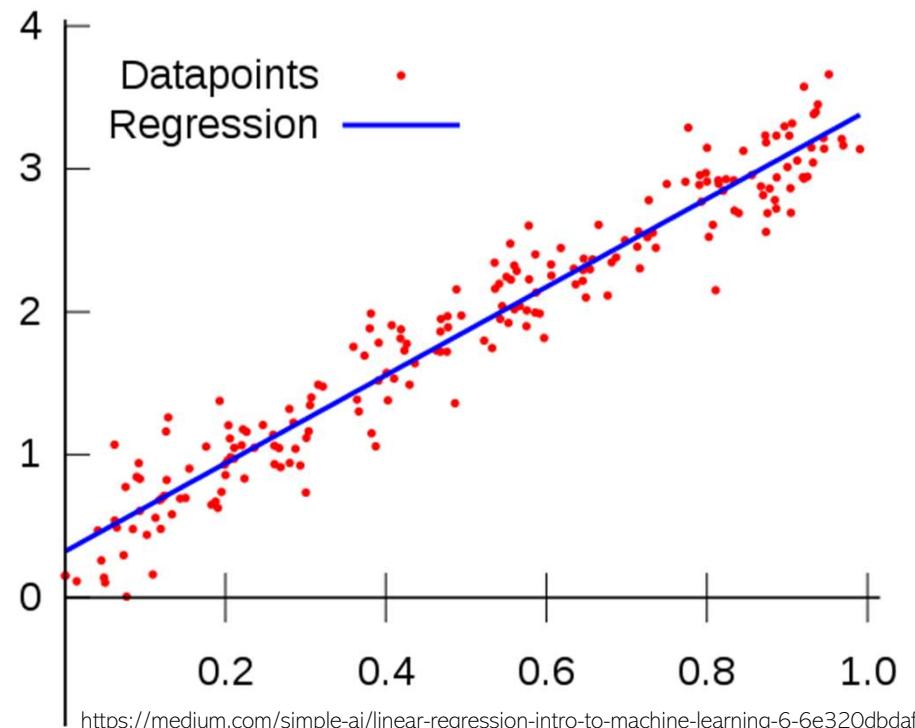
Regression

An algorithm that attempts to predict outcomes based on predictors

Can be linear, logistic, polynomial, ridge, lasso etc.

Important to consider

1. Multicollinearity
 - Predictors are correlated
2. Outliers
 - May have to pre filter for outliers
3. Heteroscedasticity
 - Unequal variability



<https://medium.com/simple-ai/linear-regression-intro-to-machine-learning-6-6e320dbdaf06>

Decisions Trees

Each node has different classification feature

Outputs of node are the possible outcomes of that classification feature

Outcome can be:

- Decision based on different classification feature
- Leaf with final classification and probability distribution

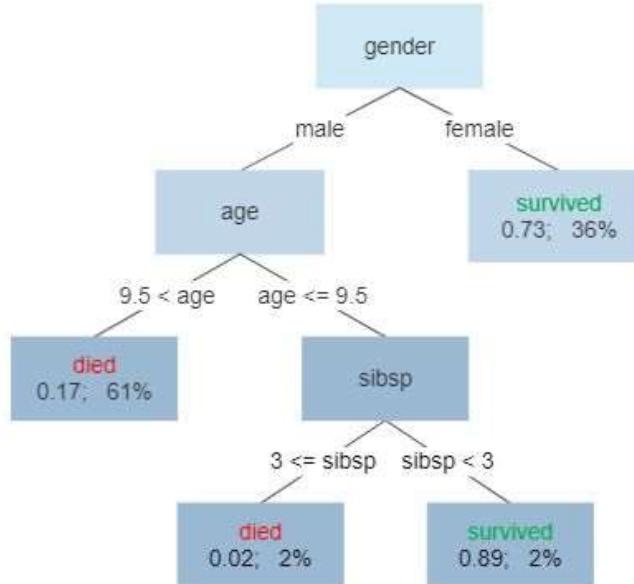
Classification Trees

- Predict discrete class i.e. yes, no
- Unordered values using dependent variables
- Titanic passenger survival

Regression Trees

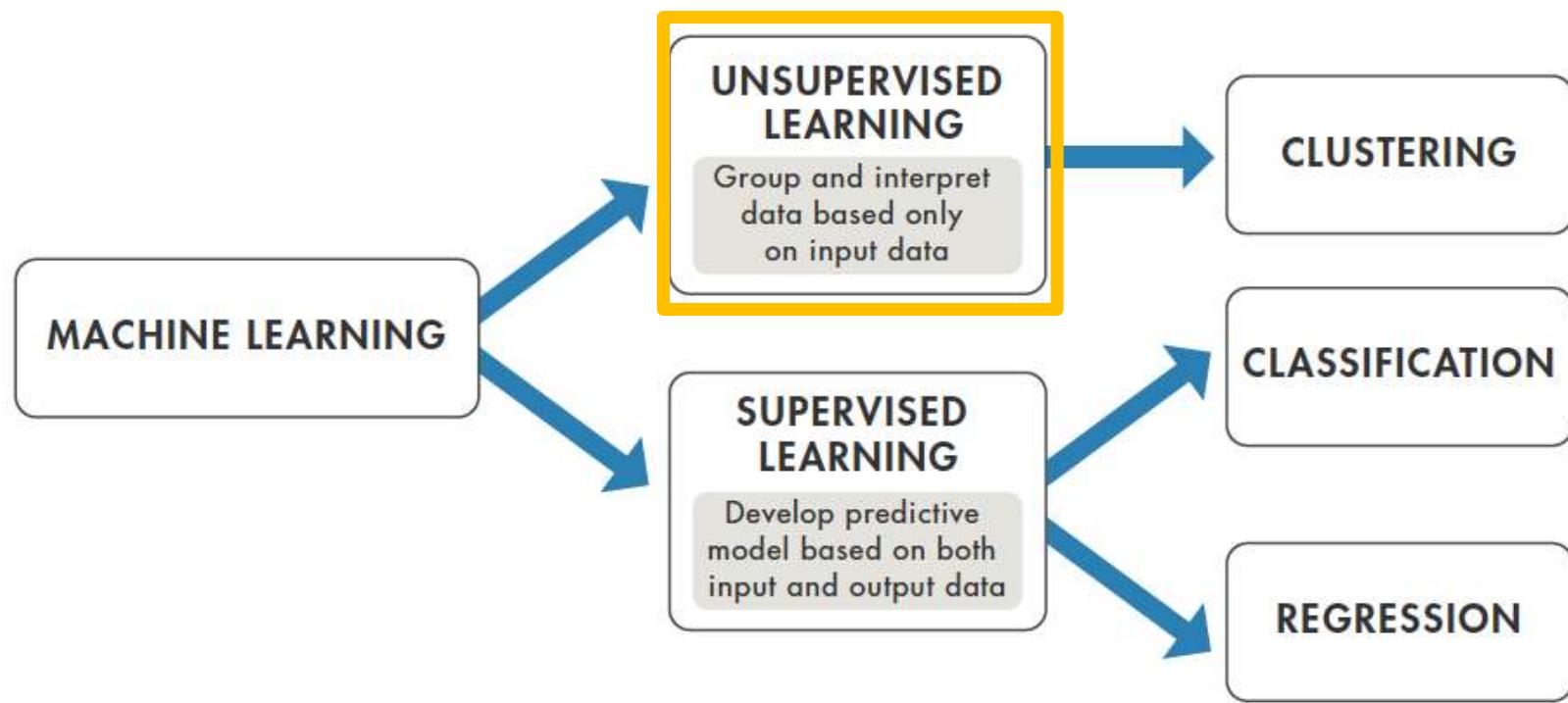
- Predict value on a continuous scale
- Ordered values
- E.x. Guess house price based on size, area etc.

Survival of passengers on the Titanic

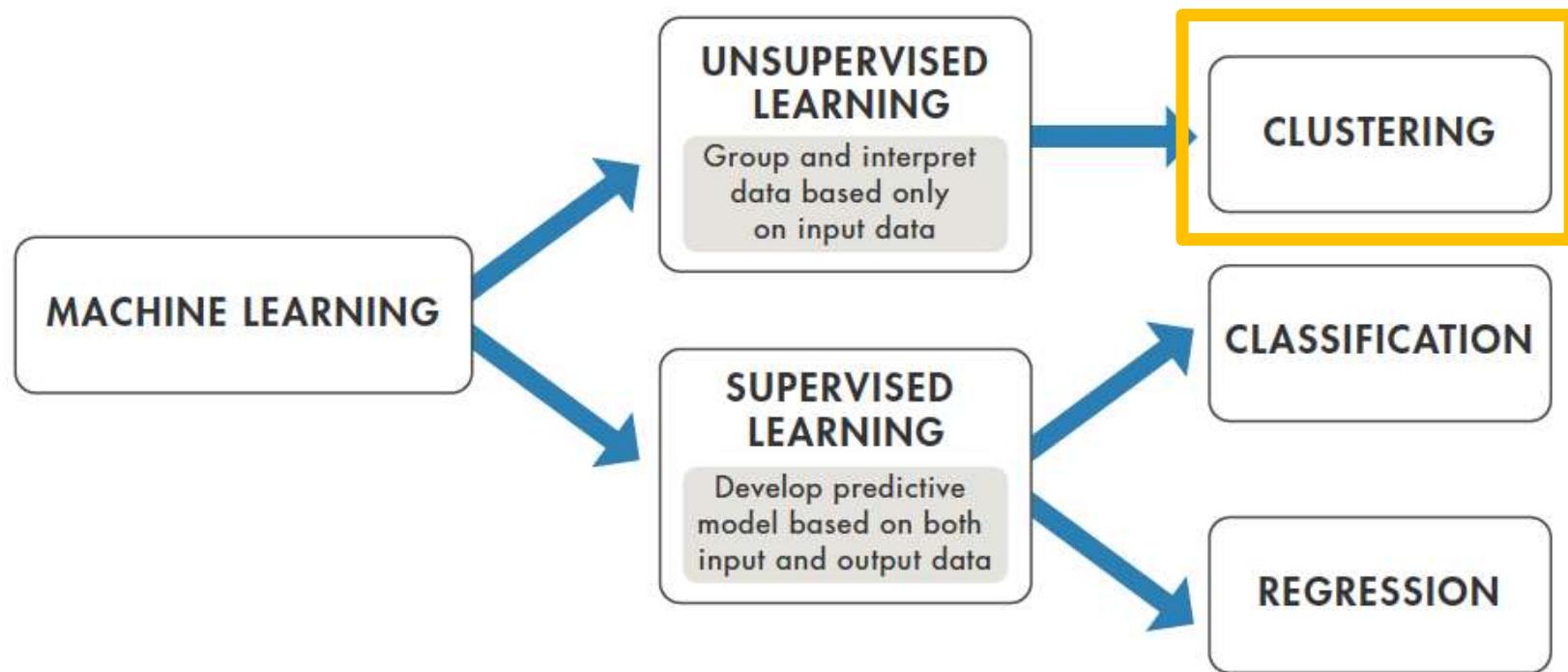


https://commons.wikimedia.org/wiki/File:Decision_Tree.jpg

Types of Learning



Types of Learning

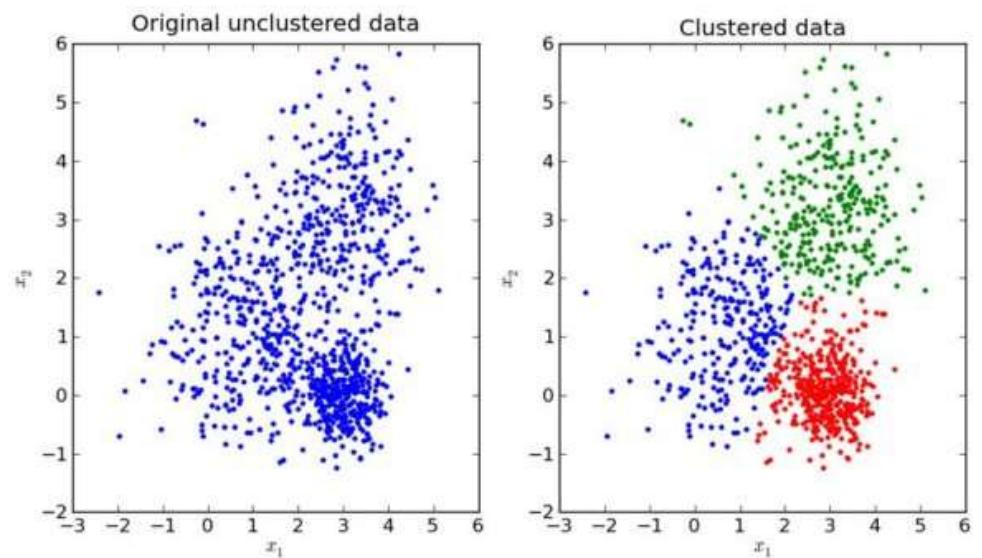


Clustering

Used to group data that is untagged/unlabeled

Try to make groups such that members of the group are more similar to each other than the other groups

Can use a variety of algorithms to achieve this



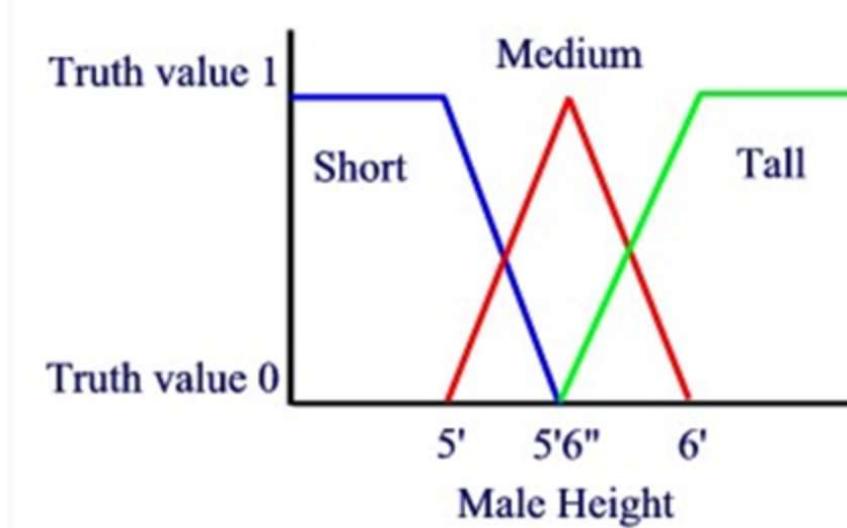
<https://www.kdnuggets.com/2017/05/must-know-most-useful-number-clusters.html>

Fuzzy Means

Data can belong to one or more clusters

Each point gets a probability to fitting into a cluster

The closer a point is to the middle of a cluster the more likely it is to belong to that group



<https://medium.com/analytics-vidhya/fuzzy-sets-fuzzy-c-means-clustering-algorithm-ac5c4386396b>

The k-means Clustering Classifier

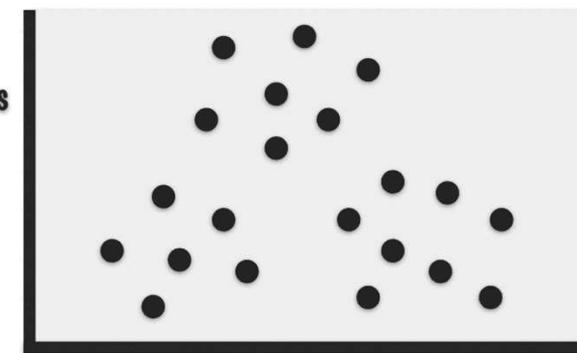
represent the training data with a number of data centers known as prototypes.

training data are reduced to a set of k representatives.

the test set point are compared to the prototypes.

Once these prototype centers are established, the test data are assigned to the class of the closest prototype.

1. Initialise random centroids
2. Until convergence:
 - Assign step
 - Update step 
3. End



<https://towardsdatascience.com/k-means-a-complete-introduction-1702af9cd8c>

The k-means Clustering Classifier (cont)

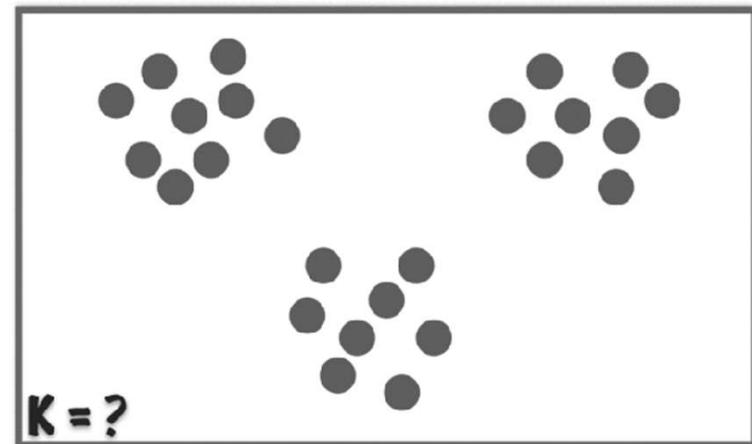
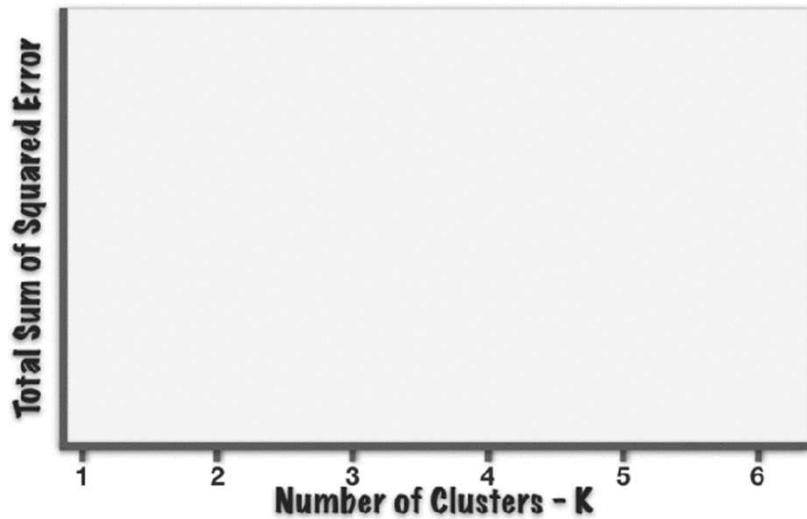
In this approach the k stands for the number of prototype centers.

The position of the prototypes determines the boundary, and the number of prototypes chosen to represent each class determines the complexity of the boundary.

The larger the value of k, the more complicated the boundary. The value of k is directly related to machine capacity.

The number of prototypes is selected by the user and the prototype centers are positioned during a training period.

Finding the “Elbow”



<https://towardsdatascience.com/k-means-a-complete-introduction-1702af9cd8c>

Prototype Positioning - LVQ method

There are several different methods for training the prototypes. The method described here is the learning vector quantization

In the LVQ method

- initial prototypes are placed randomly within each class.
- A random training data point is selected and the closest prototype is found.
- If that prototype is of the same class as the training point, the prototype is moved toward the training point.
- If not, the prototype is moved away from the training point.
- The amount of movement is proportional to the distance and a learning rate constant.

The procedure begins with a smaller learning rate constant and continues until the learning rate constant is zero.

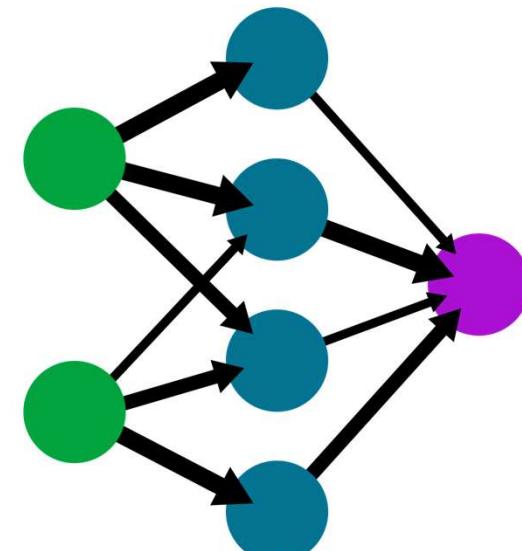
Deep Learning & Neural Networks

Neural Networks

- Artificial neural networks
 - Mimic the way neurons in the brain work
- Hidden layer contains weights, interconnected nodes, inhibitory and excitatory pathways
- Nodes are all connected to each other
- Every node is connected to next node
- Backbone for deep learning

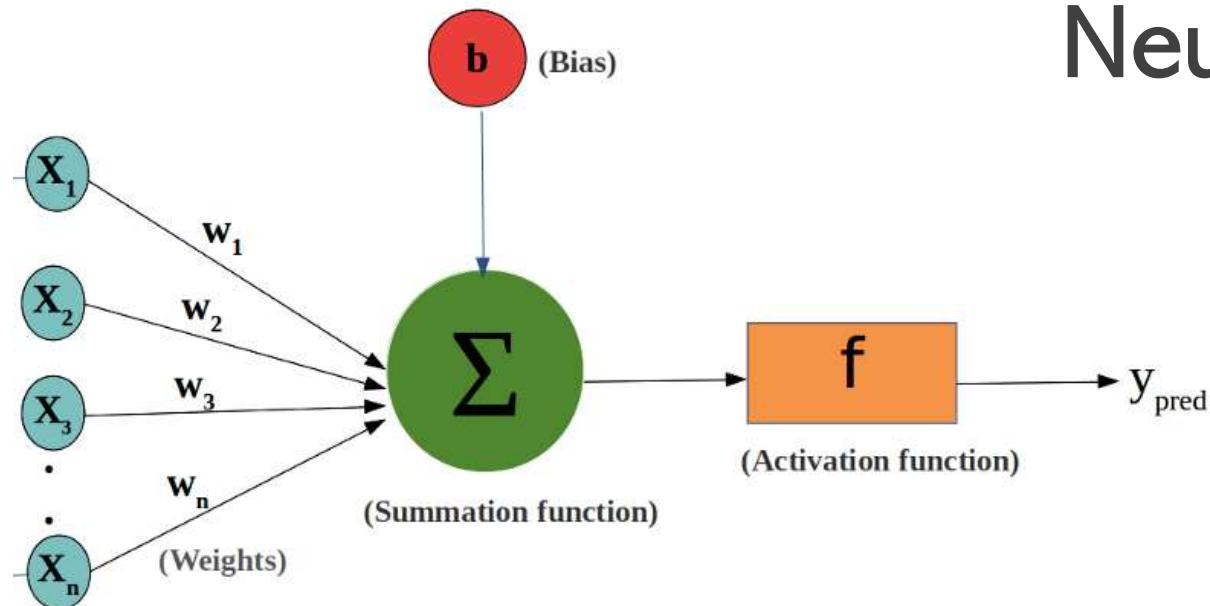
A simple neural network

input layer hidden layer output layer



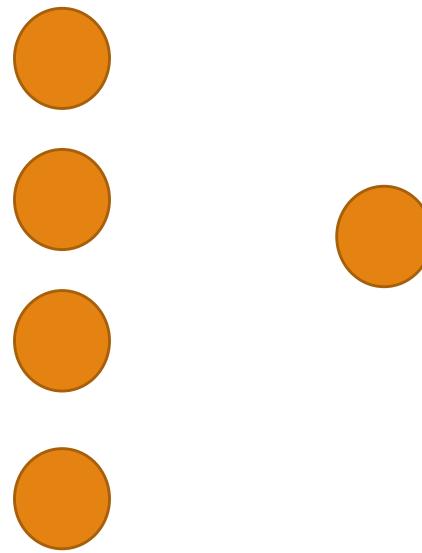
https://en.wikipedia.org/wiki/Neural_network

Neural Networks



- Connections all have weights and biases w, b
 - Weight and bias terms are what the model optimizes
 - Each level has a non-linear activation function f or σ
 - $f(X^*w + b)$
 - Forces nonlinearity into the system
- Allows for complicated input output relationships

Example





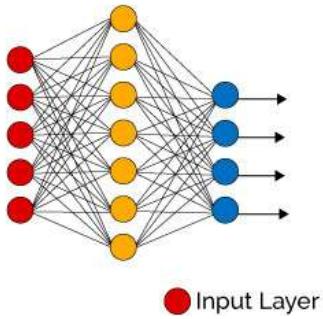
https://youtu.be/AissM0v__5s?t=3123

Neural Networks

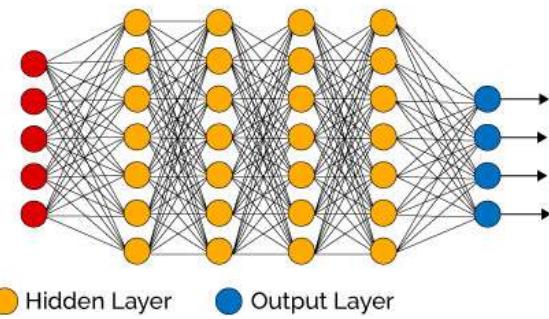
- Can do anything with just one hidden layer as long as there are enough nodes
 - Like a piecewise function to approximate a continuous function
 - if you have enough pieces you can make anything
- Node in layer uses all nodes in previous layer as input
- If we increase the number of nodes, we increase the computation time drastically...
- What if we add layers instead of adding nodes?

Deep Learning

Simple Neural Network

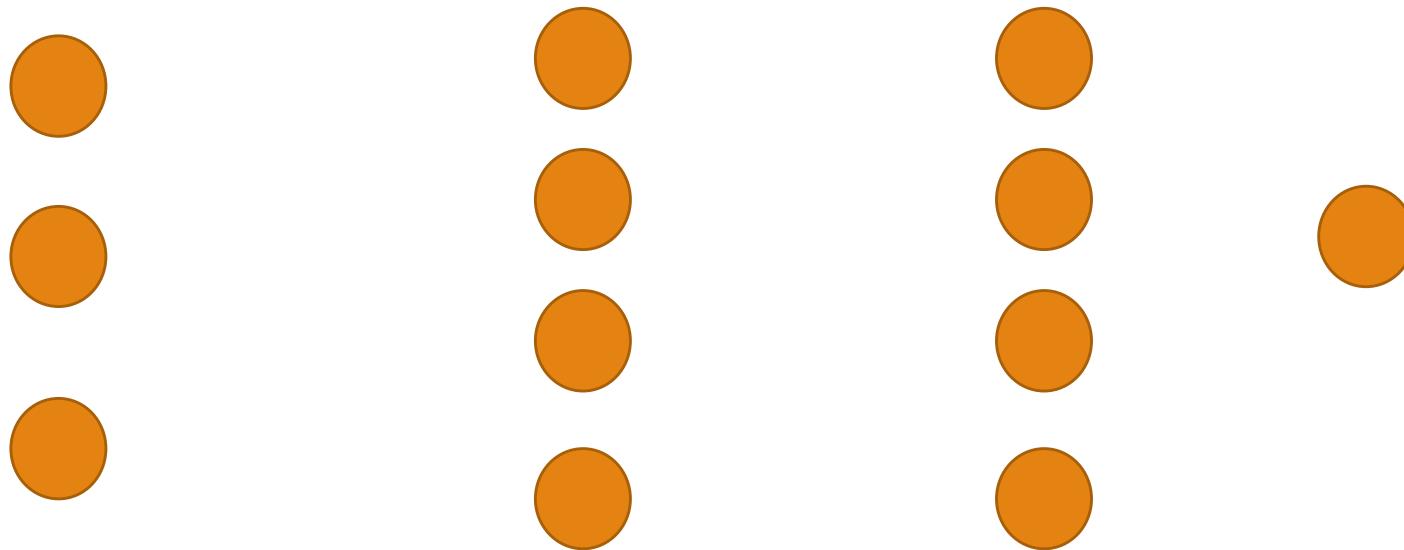


Deep Learning Neural Network



- Can be used in unsupervised or supervised settings
- Good for large data sets
- Layers cascade
 - $\sigma_1(X^*w_1 + b_1) \rightarrow \sigma_2(\sigma_1(X^*w_1 + b_1) w_2 + b_2)$
- Can also be looked at as matrix transformations occurring at each layer
 - $\phi = W^t x \rightarrow \phi' = \phi U^t$
- More hidden layers allow for reusing functions which can simplify the algorithm
- Each weight vector (node) learns differently

Example



Example

Create an algorithm to sort photos of leaves vs faces

Normal supervised learning

- give it factors
- we have to know what factors are useful to start with

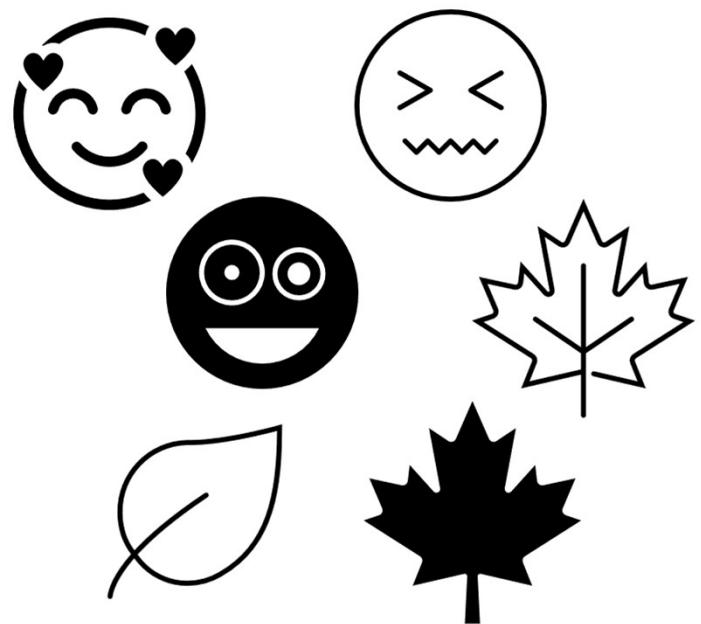
What factors do we give it?

- fractal dimension
- colour
- histogram

What if we have 1000 classes with thousands of images in each...

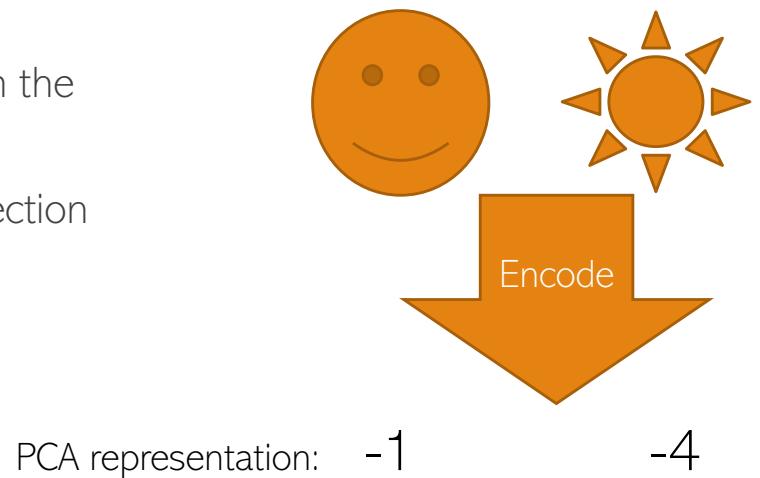
Deep learning creates a nonlinear model that can generalize the data well

Small cost – just need the labelled set, or a labelled subset



Example: Auto Encoder

- High dimensional encoder → Compression latency representation → reconstruct using decoder
- Objective: Algorithm optimizes itself to reconstruct data from the reduced dimensionality PCA data
- Self-training and can be used for unsupervised anomaly detection
- Take data down to latent small dimension representation
- Normal vs anomaly may be more obvious at this level
 - Faces may fall between -1 and 1
 - Sun may be ~ -4



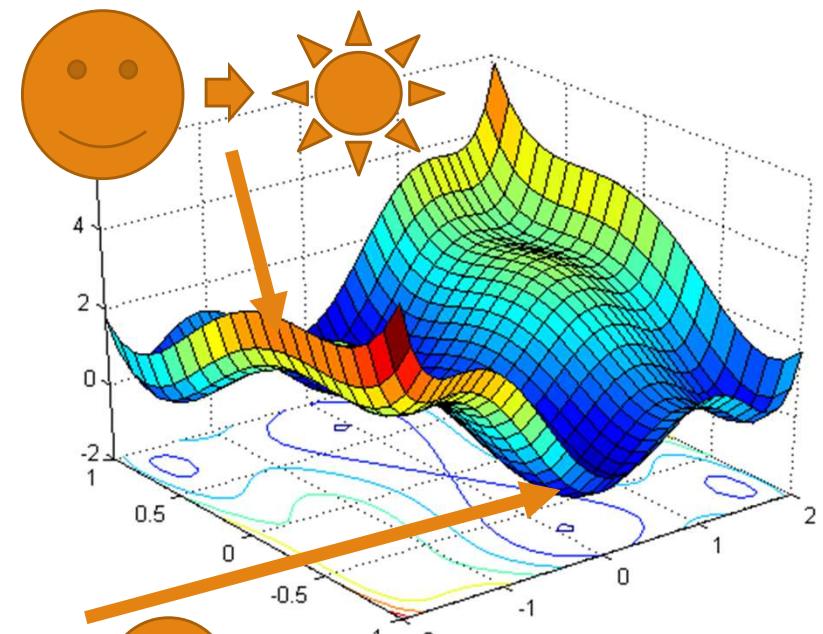
Loss Function

Mathematical Goal?

- Minimize the mean squares error
- i.e. minimize the loss function
- NN keeps trying to optimize to step closer to local minimum in loss function

Loss function is a function of weights and is dependent on the input

- In general, deep learning algorithms have complicated loss functions
 - many high and low points
 - Want to find deepest points which can be done using many different random inputs
- Stochastic gradient descent
 - Iterative model
 - Start with random values and slowly tailor your transformation matrices as you go
 - Move along loss function towards local minima
 - Uses calculus to do this



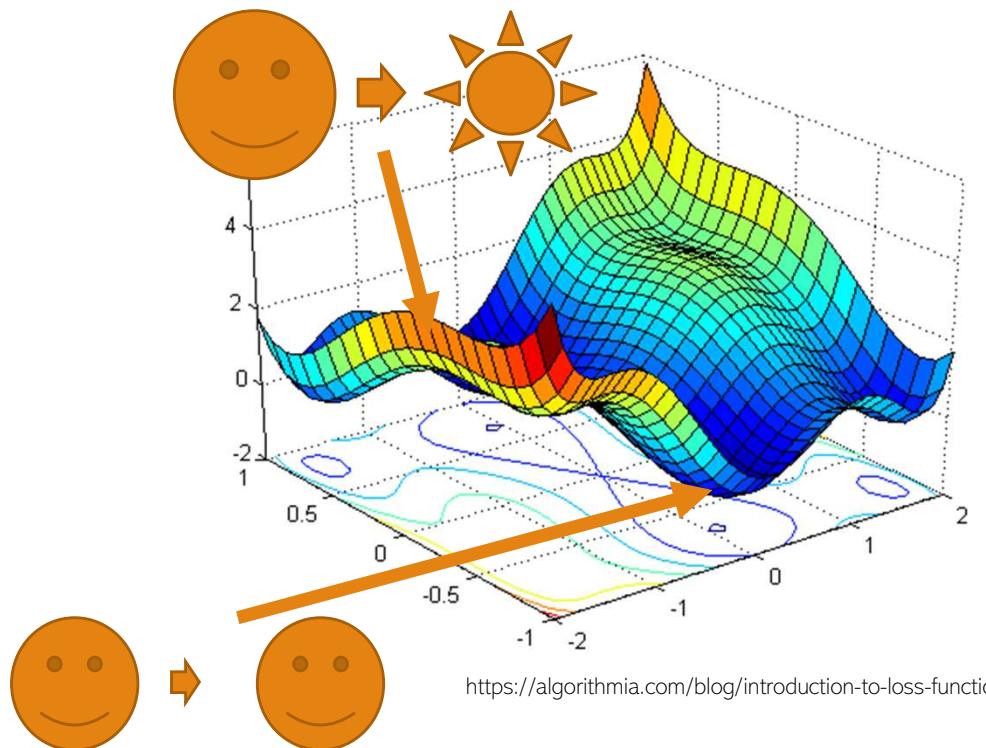
<https://algorithmia.com/blog/introduction-to-loss-functions>

Loss Function

Mathematical Goal?

- Minimize the mean squares error
- i.e. minimize the loss function
- NN keeps trying to optimize to step closer to local minimum in loss function

Loss function is a function of weights
and is dependent on the input



Picking a ML method

Have to understand your model to know if it is really working for your application

Need ML model to be generalizable

Model should work for data from many different sets and sources

Different models work best for different sizes of data sets

Is the cost of labelling worth it?

Is your data homogenous?

Using Supervised Learning to Predict Heart Attacks

Cardiologist wants to predict whether someone will have a heart attack within a year.

- 1) They have data on previous patients, including age, weight, height, and blood pressure, 6 lead ECG and HRV measures
- 2) They know whether those previous patients had heart attacks within a year.
- 3) So the problem is combining the existing data into a model that can predict whether a new person will have a heart attack within a year.

Modelling steps

- 1) Do we have patient consent to use their data?
- 2) Do we have a homogeneous data set?
- 3) CRD? RCBD?
- 4) Do we need all this data?
 - ECG → ICA
 - HRV → just use Average N N intervals?
 - PCA → Maybe blood pressure has limited variance between patients
- 5) Do we want more data?
 - STFT or wavelet transform ECG to look for changes to frequency
 - Cosinor to look for changes in BP throughout the day?
 - Is there fractal self similarity to the HRV throughout the period?

Modelling steps

- 6) Model
 - Linear regression?
 - Logistic regression?
 - Machine learning?

- 7) Last but not least...
 - Statistics