APPENDIX A

Math Glossary

---

affine transform in a plane

Depending upon the desired complexity, one of the following matrices is traditionally employed:

$$r' = \begin{bmatrix} s_x \cos\theta & -k_y \sin\theta \\ k_x \sin\theta & s_y \cos\theta \end{bmatrix} r$$

$$r' = \begin{bmatrix} s_x \cos\theta & -k_y \sin\theta & t_x \\ k_x \sin\theta & s_y \cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} r$$

where
k shearing
s scaling
t translation

Transforms, in this form, are appended *on the left*. For example

$$A_3 A_2 A_1 Moon$$

Where A_1 rotates the Moon about its axis, A_2 puts it in orbit (at a distance and current angle in its orbit) around a planet; and A_3 puts the Moon & planet into orbit around the sun, and so forth. Note that parallel lines remain parallel in this transformation.

inferring

Can be found by using an FFT as a correlation form

See also *FFT*

items

The items that can be transformed include:

- A point, either $\begin{bmatrix} x \\ y \end{bmatrix}$ or $\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$ (in the event of that the translation style matrix is employed)
- A line or curve
- An image
- A sub-frame – and the affine transform to translate it to the current frame. Recall that any translation on the current frame is on the left of the sub-frame's affine transform.

Cramer's rule

This solves an affine transform problem of the form:

$$\mathbf{A}v = u$$

Where **A** and u are known, but v is not:

$$v_x = \frac{\begin{vmatrix} u_x & A_{1,2} & A_{1,3} \\ u_y & A_{2,2} & A_{2,3} \\ u_z & A_{3,2} & A_{3,3} \end{vmatrix}}{|\mathbf{A}|}$$

$$v_y = \frac{\begin{vmatrix} A_{1,1} & u_x & A_{1,3} \\ A_{2,1} & u_y & A_{2,3} \\ A_{3,1} & u_y & A_{3,3} \end{vmatrix}}{|\mathbf{A}|}$$

$$v_z = \frac{\begin{vmatrix} A_{1,1} & A_{1,2} & u_x \\ A_{2,1} & A_{2,2} & u_y \\ A_{3,1} & A_{3,2} & u_z \end{vmatrix}}{|\mathbf{A}|}$$

See also *Matrix (left inverse)*

algebra

See also *division algebra*

alligation

$$P_{xy} = \frac{xX + yX}{x + y}$$

where
x,y are proportions of two ingredients
X,Y are values of property Pxy in the two ingredients

approximation

Sufficiently 'good' calculation as other methods are too expensive or not current feasible. Techniques or other names include:

- Curve fitting
- Interpolation
- Numerical integration
- Estimate of how well the approximation works
- Super optimizers
- Characteristic function
- Linear approximation tables
- Polynomials
- Multiple variable regression
- Solving differential equations

Characterized by an error figure ("residues")

term	*synopsis*
approximation	function that produces results close to the value of a reference function or relation;
estimation	Tries to approximate the value of variable that is not observed / measured, but is closed to the value if it were measured.
forecast	success value in sequence (usually time based) from past values
prediction	
simulation	class of techniques to estimate or forecast a value
trend	

Table 1: *Distinction between different related terms*

area
signed

If the area is negative, the traversal around its outer edge was clockwise; otherwise the traversal was counterclockwise

arrival rate
average

$$\langle \lambda \rangle = \sum \lambda_i p_i$$

see *trends (predicting)*

average see *central tendency, mean*

Back solving Removing unnecessary information in a Gaussian eliminated system to further solve it.

See also *Affine transform (Cramer's rule), Matrix (inverse)*

basis The atoms a description, structure, system or vector are defined in terms of. With a vector, the basis elements can be the axis:

$$v = x\hat{i} + y\hat{j} + z\hat{k}$$

Note that they do not have to be linearly independent.

see also *coordinate system*

basis function The things a computation is defined in terms of, taken as a given. A system of computation provides the basis functions, which are usually designed to:

- Meet the criteria that anything should be computable – even if at great cost – in terms of the functions, and
- Either speed execution or appear to be a natural way to express computation.

Bayes Theorem Bayes theorem is used to estimate the probability of an event, in the current (or hypothetical) context, given its likelihood and that of its precedents. This is typically represented using *conditional probability* notation. The Bayesian probability of A is computed in the simplest (*Boolean)* case is event A occurring, or not occurring:

$$P(A|s) = \frac{P(s|A)P(A)}{P(s|A)P(A) + P(s|\neg A)P(\neg A)}$$

The general case is from the *mutually exclusive* events that A_1 .. A_n that exhaust all other possibilities of getting into the given s state. This latter formulation is more suitable for matrix and vector representations of systems.

$$P(A|s) = \frac{P(s|A)P(A)}{P(s|A)P(A) + P(s|A_1)P(A_1) + \ldots + P(s|A_n)P(A_n)}$$

see also *Chapman-Kolmogorov theorem, conditional probability, fault trees, Gaussian distribution, inference, Poisson model, probability distribution, stationary distribution, transition matrix, Vitterbi method*

identity

$$\mathrm{I} = \frac{P(A|s)P(s)}{P(s|A)P(A)}$$

naïve Bayesian Assumes that the simple presence of an element is enough to predict the outcome or select the best hypothesis. Usually the order, sequence, how the elements are combined – and other factors – are important to identifying the outcome in greater accuracy.

Bayes factor Used in the question of selection competing hypothesis given the same evidence.

$$= \frac{P(Hypothesis_1 x)}{P(Hypothesis_2|x)}$$

If the P(x|Y) is symmetrical around x, but no specific distribution, the Bayes factor is: *Berger and Jeffreys, 1992*

$$\leq \sqrt{\frac{2}{\pi}}\left(|D_f| + \sqrt{2\ln\left(|D_f| + 1.2\right)}\right)e^{-D_e^2/2}$$

D_e = the number of standard deviations measured value deviates from prediction H1

D_I= the number of standard deviations measured value deviates from prediction H2

See also *Chi-squared test.*

classification

A variant of selecting between hypothesis, or atleast formed in that manner. Each of the hypothesis is of the form:

The given item could have been generated by the rules of category y

The selection between them is a matter of finding the hypothesis with the max probability.

Benford's law

The empirical observation that the first digit of a decimal number is not equally likely to be 1..9, rather it follows a decreasing probability:

1 30%
2 18%
…
9 4.6%

Benford's Law is often useful in spotting fraud.

see also *Newcomb's law, Hill's theorem.*

Besicovitch dimension formula

$$D = \lim_{e \to \infty} \frac{\log N(e)}{\log \frac{1}{e}} = -\lim_{e \to \infty} \frac{\log N(e)}{\log e}$$

where
D= Besicovitch Dimension
e = element length
N(e) = number of elements need to cover object

Bessel's Equation

$$x^2 \frac{d}{dx^2} + x \frac{d}{dx} + \left(x^2 - n^2\right) y = 0$$

n = order

Solution is a Bessel function of the first kind:

$$y = AJ_n(x) + BY_n(x)$$

$$J_n = \sum_{k=0}^{\infty} \frac{-1^k}{\Gamma(k+1)\Gamma(k+n+1)} \left(\frac{x}{2}\right)^{n+2k}$$

$$Y_n = \begin{cases} n \text{ is an integer} \\ J_{-n}(x) \end{cases}$$

see also *Ordinary Differential Equation*

Beta distribution

aka Pearson distribution

Beta function

aka Euler's first integral

$$B(r,s) = \int_0^1 dt\, t^{r-1}(1-t)^{s-1}$$

$$B(r,s) = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$$

$$B(s,r) = B(r,s)$$

Beziér curve

binary tree

signature of an item is the path of left-right choices. Each step along the way can have a different threshold, possibly fixed or a median divisor. Note: the partitions must not loop back or cross self when dividing space.

binary tree: divided by point or median

kd tree: cycles thru each axis on per level) usually a point along the axis that is a median divide.

support vector: linear partition thru the space (a tree of these is called?) knowing that

a point is on the partition, and constraints on that, allow inferring the other parameters.

neural network: non-linear partition.

binomial distribution

The binomial distribution is used when items are divided between two bins. For example, if 2/3's of a set of buckets are 'full' (p=0.667), with a sample of 5 buckets (n=5), how many of those will be 'full'? P(x=3). To describe the probability that an event with probability *p* occurs every *x* times in *n* trials, the first bin is for the event of interest, the second for the rest.

A binomial distribution can be replaced with a normal distribution when:

$$n > 30$$
$$np > 5$$
$$n(1-p) > 5$$

see also *combination, normal distribution*

***Table 2:** Binomial distribution parameters*

Parameter	*Description*
p	Probability of an event; or the proportion of items that are in the first category
n	The number of trials examined, or the number of samples in the measured set.
x	The number of times the event is hypothesized to occur; or the time number of items hypothesized to be in the first category.

***Table 3:** Binomial distribution properties*

Property	*Form*
probability density function	$P(x) = {}^{n}C_{x} p^{x} (1-p)^{n-x}$
mean	np
standard deviation	$\sqrt{\frac{p(1-p)}{n}}$
confidence intervals	$p \pm (z \cdot stddev)$

binomial expansion

Expands equations of the form $(a+b)^N$. Typically solved using Pascal's triangle.

$$(a+b)^N = \sum_{n=0}^{N} {}^{N}C_{n} a^{n} b^{N-n}$$

$$(a+1)^N = \sum_{n=0}^{\infty} {}^{N}C_{n} a^{n}; N < 0, |a| < 1$$

C is the "combination" function

See also *combination*

Boltzmann constant

$k = 1.3906\ 10^{-27}$ joule/Kelvin

often used in entropy calculations and probability distributions

distribution

$$P(A \mid B) = e^{\frac{1}{kT}(E(B)-E(A))}$$

boolean table

Karnaugh maps for up to 6 variables, then switch to Quine-McCluskey algorithm or Espresso minimizer.

bounding

box, sphere, cylinder

bundle adjustment

non-linear least squares optimization. Use Levenberg-Marquadt algorithm to solve.

canonical transformation

A transformation of coordinates, but preserves key physics relationships

$$L_1(q,q') = L_2(Q,Q') + \frac{dS}{dt}$$

where

L1 – Lagrangian function in original set of coordinates

L2 – Lagrangian function in transformed set of coords

Catalan

$$C(n) = \begin{cases} n = 1 & 0 \\ n > 1 & \sum_{i=1}^{n} C(i-1)C(n-i) \end{cases}$$

$$C(n) = \frac{2n!}{n!(n+1)!}$$

The number of triangulations for a polynomial is C(n-2) where n is the number of sides

see also *combination*

Cauchy distribution function

$$= \frac{1}{\pi(1+x^2)}$$

This is very similar to a Gaussian distribution (with fatter tails), and should be used only when the Gaussian does not apply.

see *distribution*

center

use the determinant method to find the center of a circle, triangle, or rectangle.

center of mass

$$= \frac{\sum_{i=0}^{T} i \cdot hist(i)}{\sum_{i=0}^{T} hist(i)}$$

central limit theorem

"the sum of different independent random components is a random value distributed according to the normal distribution law." It does not matter what the distribution law is for the random components.

see also *probability distribution*

central tendency

Includes mean, median, mode, negative skew and positive skew. There are no rules or mechanical methods to say whether median, mean, or mode is most appropriate description of the central tendency.

see also *normal distribution*

centration bias

Tendency to focus on just one feature of a problem.

Chapman-Kolmogorov

Forward chain of inference, see *conditional probability*

Chebyshev approximation

Typically a region that is defined by a polynomial divided by another polynomial, and a rule how transform other regions into that.

Chebyshev inequality

Most of the values in a probability distribute are near the mean. $1/k^2$ values are more than k standard deviations from the mean.

$$P(|X - \mu| \geq \alpha) \leq \frac{\sigma^2}{\alpha^2}$$

$$\alpha = k\sigma$$

X^2 test

$$X^2 = \sum \frac{(O-E)^2}{E}$$

O= observed frequency in that bin
E= expected frequency in that bin

testing a

Used to test a model that predicts the future. Let's say that it is a table of sales figures by

hypothetical model

region for each quarter, with a projected table and the actual results.

Parameterization:

- Number of classes: # regions
- Number of constraints: # of rules to create model
- Degree of freedom= #Classes - # Constraints

Steps:

1. Compute X^2 for the what was expected from the model, and what happened.

$$X^2 = \sum_{region,qtr} \frac{\left(\left|O_{region,qtr} - E_{region,qtr}\right| + \frac{1}{2}\right)^2}{E_{region,qtr}}$$

2. Lookup X^2 value in table for Degree of Freedom.
3. If computed X^2 >= value in table, the hypothetical model is invalid.

classifiers

Grouping similar items for the purposes of classification or categorization. Conversion between specific forms of classifiers.

ID3, ID4, C4.5, CART

Classification and Regression Trees
L Breiman, J Friedman, R Olshen, C Stone 1984
Wadsworth, Belmont

linear

Groups items based similar features; the classification is based on the value of the linear combination of features. Techniques include: linear discriminators, naïve Bayesian classifiers, neural networks, kernel machines, and linear regressions.

Linear classifiers work by partitioning a space, those on one side are in group X, those on the other are in group Y. (Several partitions are needed to classify into more than one group). This line can serve as a linear regression, and used to infer other properties, akin to an associative memory. The process involves find the point(s) on the line coordinated by the givens, and rest are specified by the line's characteristic.

See also *co-linearity, latent semantic analysis, regression*

statistical

Groups items based on quantitative information; often needs a training set

clustering

Other: k-means See *classifiers, correspondence analysis*

hierarchical algorithms

Create clusters by merging or dividing (agglomeration / division). Agglomeration forms clusters by merging and treats each instance in the population as an instance of a cluster. Divisive starts with one cluster, splitting until a stopping point has been reached.

coefficient of determination

Variation of variable that can be predicted from another variable

$Cor(x,y)^2$

collar variables

A small set of key variables that determine a systems behaviour. The system may be controlled by controlling these. Similarly the system state may be described primarily with these.

aka *back doors, master variables, narrows, variable subset selection*

Collatz problem

$$x_{n+1} = \begin{cases} x_n \text{ is odd} & \frac{1}{2}x_n \\ x_n \text{ is even} & 3x_n + 1 \end{cases}$$

collinearity

The easiest method of telling if three points are on a line is compute the area of their triangle, and confirm that it is zero.

$$0 = \frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}$$

$$0 = \frac{1}{2}\left(x_1 y_2 + x_2 y_3 + x_3 y_1 - y_1 x_2 - y_2 x_3 - y_3 x_1\right)$$

See also *polygon (area)*

combination₁ Which elements are used together; their order (in a sequence) is not important. The number of combinations are:

$${}^{n}C_{r} = \frac{n!}{r!(n-r)!} = {}^{n}P_{r}\frac{1}{r!}$$

The evaluation of this can be sped: ${}^{n}C_{r} = {}^{n}C_{n-r}$

The recursive definition is: ${}^{n}C_{r} = {}^{n-1}C_{r-1} + {}^{n-1}C_{r}; 0 < r < n$

where
r is the number of objects used in the combination
n is the set of possible objects that might be selected for use in the combination

see also *catalan, counting, permutation*

combination₂ see *counting, permutation*

commute Operators commute if the order in which they are performed does not matter AB == BA

commutative algebra Typically we are most interested in whether or not *multiplication* operations are commutative. Non-commutative algebras can be self-consistent. Efficiently evaluating long sequences of multiplications typically involves re-ordering multiplication steps so that they can be pre-multiplied, and *parenthesization* to improve the evaluation order. The later can be applied to non-commutative, but often has to be done dynamically, unless the variable shapes (matrix sizes) are known ahead of time.

See also *parenthesization*

composition function $(f \circ g)(x) = f(g(x))$

computer algebra system Mechanically solves and/or manipulates systems of symbolic equations and concrete numbers. May involve symbol pattern recognition.

condition number An indicator of how the errors in `b` affect the solution of A**x**=b; with higher the number, the less stable the solution. The condition number is computed differently for each of the matrix norms.

see also *matrix norm*

confounding of variables When two or more variables vary together in a way that makes it hard to sort out their independent effects.

constraint satisfaction Faster than brute force for certain. Or approximate such a problem. see also *simplex method, unification*

convolution

$$(f * k)(x) = \int_{-w/2}^{w/2} d\delta\, f(x)k(x-\delta)$$

k is the convolution kernel. w is the width of the kernel – -w/2 to w/2 is called the support.

see also *edge filter, FIR*

convolution filter There are several: blur, edge filter, sharpen, and unsharpen mask
convolution

transform

How to form a function into a convolution

How to build a convolution by specification

See also *edge filter*

coordinate system

A coordinate system needs to define

- Scale factors (default to 1)
- Displacement vector
- Arc length
- Volume element
- Gradient
- Divergence
- Curl
- Laplacian.

see also *area, covariance (general), volume*

Table 4: *Coordinate Systems*

Coordinate System	*Form*
cylindrical	$ds = \hat{e}_r dr + \hat{e}_\theta d\theta + \hat{e}_z dz$
	$ds^2 = dr^2 + r^2 d\theta^2 + dz^2$
polar	$ds = \hat{e}_r dr + \hat{e}_\theta d\theta$
	$ds^2 = dr^2 + r^2 d\theta^2$
rectilinear	$ds^2 = dx^2 + dy^2 + dz^2$

homogenous coordinate system

Adds a "w" coordinate. <xy,z,w> == <x/w,y/w,z/w>. 0 is a direction vector, others are points. Matrices need a (eg 4x4) need a bottom row of 0,0,0,1

correlation

"We may now depict the correlation structure among the [n] measures in two ways:

Stephen Gould, The Mismeasure of Man*, p240-3*

1. "We may gather all correlation coefficients between pairs of measures into a single table, or matrix of correlation coefficients… The line from upper left to lower right records the necessarily perfect correlation of each variable with itself. It is called the principal diagonal, and all correlations along it are 1.0. The matrix is symmetrical around the principal diagonal, since the correlation of measure 1 with measure 2 is the same as the correlation of 2 with 1. Thus the three values either above or below the principal diagonal are correlations we seek…
2. "We may plot the points for all individuals onto an [n] dimensional graph. [When] the correlations are all positive, the points are oriented as an ellipsoid (or football). (In two dimensions they form an ellipse.) A line running along the major axis of the football expresses the strong positive correlations between all measures."

See also *factor analysis*

pearsons

A standardized measure of how two variables affect each other; a numerical index of the degree of relationship that exists between two variables.

"Pearson's *r* is not an appropriate measure for all kinds of correlation, for it assesses only what statisticians call the intensity of linear relationship between two measures – the tendency for all points to fall on a single straight line. Other relationships of strict depedence will not achieve a value of 1.0 r."

see *Pearson Product Moment Correlation*

interclass correlation

$$InterclassCorrelation = \frac{\sigma^2}{|\sigma|^2}$$

$$\sigma = \begin{pmatrix} \sigma_a \\ \sigma_e \end{pmatrix}$$

$$\sigma^2 = \sigma \cdot \sigma$$

$$Y_{ij} = \mu + a_i + e_{ij}$$

Y_{ij} The overall response of element *j* within cluster *i*
μ The mean
a_i The variation (error) that comes from the cluster *i*; it has a mean of 0, and a variation of σ_a^2

e_{ij} is the (unit effect) error from element *j* within cluster *i*; note this means the same element has a different coefficient it moves to another trial or class. This error has a mean 0, and a variance of σ_e^2.

linear

$$y_n = L\{x_n\} + e_n$$

Where

$L\{\}$ is a linear transform of x_n

x_n is the signal

e_n is the error (independent of x_n)

$$r^2 = \frac{\text{cov}(x, y)}{\text{var}(x)\,\text{var}(y)}$$

r^2 is a measure of the total variance

$|r| \cong 1$ high linear relationship

$|r| \cong 0$ no linear relationship

J Pijn's method

k data divided up into L equal-sized bins

p_i = mid point of values in bin i

q_i = average y amplitude in bin i

connect $<p_i, q_i>$ as a line and perform piecewise linear regression

$$h^2 = \frac{\left(\sum_i y_i - \langle y \rangle\right)^2 - \sum_i \left(y_i - f(x[i])\right)^2}{\sum_i \left(y_i - \langle y \rangle\right)^2}$$

$r^2 = h^2$ if x,y have a linear relationship

correspondence analysis

continuous data: principal component analysis

categorical data: latent semantic analysis

correspondence problem

Given two or more images, which points correlate in one with another?

How to find this correspondence efficiently?

1. Points are limited to a set of "interesting" points. These are based on their being likely unique, and found with an procedure, such as SIFT
2. Points that are likely related are found (as a subset) via an approximate nearest neighbor search.
3. The Fundamental Matrix F relates two such points from images *j* and *k*:

$$x_{ij}^T F x_{ij} = 0$$

see also *fundamental matrix*

counting

How many ways of choosing x items from a counted set S? This depends on:

- Is duplication of items allowed? Can the same item be selected multiple times?

(That is, is the system *degenerate?)*

- Is the order of important? Is it a sequence or an unordered set?

see *combination, counted set, multiples, permutation, scheduling*

Duplication allowed?	*Ordered?*	*Method*	*Symbol*
Yes	-	**Multiples**	${}^{n}M_{r}$
No	Yes	**Permutation**	${}^{n}P_{r}$
No	No	**Combinations**	${}^{n}C_{r}$

Table 5: *Methods of counting the number of ways one can choose x items from a counted set*

covariance (x,y) A scalar that estimates how similar two processes are (how much their values moved together).
$= 1/n \sum_{i}^{n} x_i y_i - \langle x \rangle \langle y \rangle$

general covariance The equations retain the same form in all coordinate systems.

Cramer's rule See *Affine transform (Cramer's rule), Matrix (left inverse)*

critical points The minima, maxima (including those at the boundaries), roots and poles of a function.

cross-sectional study Compares groups of subjects (who differ in some manner) observed at the same point in time.

curve
simple A curve that does not cross itself

parameterization see

decision tree Results: a thing of two forms:

It is passed a thing (oracle) that will answer questions as the tree is traversed.

All of the relevant facts are passed as a parameter.

"Given enough data, a decision tree can model virtual any relationship, no matter how complex.. It can also handle incomplete data.. much more easily than can regression analysis."

Robert Pool "If it ain't broke, fix it" Tech Review Sept 2001

algorithms: ID3, C4.5

how it works Selects attribute that has closest coupling with output. This coupling is traditionally measured as information entropy, to use the column with the least variation. The selection may be done in a subset of circumstances (i.e. only when the outlook is overcast and we need more information).

"The decision tree begins by slicing the data into two or more subsets that best correlate with the two divergent outcomes. Each of the resulting two subsets of data points is then split by the same best-predictor criterion (etc).. until further divisions do not improve the predictive value." (Pool)

quantitative Simple trees
trees with independent conditions
trees with dependent conditions
decisions with conditions.

See also *entropy*

node

- next: go this on fail or alternation
- item type to match
- comparison: comparison delegate used to check if thing
- match action: what to do if there is a match

- sub: next node on successful match

perhaps could use a unification style model.

decomposition To break down into simpler part.

See *matrix decomposition, FFT, trend decomposition*

degenerate Characteristic functions with the same characteristic values.

Dempster-Shafer Useful for handling ignorance

Estimate degree of belief for one proposition given the subjective probabilities of a related proposition.

Take the orthogonal sum of each possible proposition to combine (when based on independent evidence). Use belief figure, or measure of how much we know.

A P Dempster, "A generalization of Bayesian inference" Journal Royal Statistical Society B, *Vol 30 n 2, p205-247, 1968*

G Shafer A mathematical theory of evidence, *Princeton NJ, Princeton University Press 1976*

derivative
method of differentiation

Ridder method

See also *integration*

differential equations
solving

Given a differential equation, any method or heuristic may be used to obtain a solution if you can show that the solution (1) satisfies the differential equation, and (2) satisfies the boundary conditions.

n: the number of constants.

The number of constants – highest order of differentiation.

For linear differential equations, then if y=y1(x) satisfies the equation, and y=y2(x) satisfies the equation, then so does y=c1y1(x)+c2y2(x)

see also *ODE*

Linear superposition

The sum of two solutions is also a slution

multiplying any solution by a constant is a solution.

differential operators
key

Key differential operators – these have coordinate system specific analytic definitions.

	Operator	*Interpretation*	*Integral theorem*
$\nabla\phi$	grad ϕ	Maximum rate of change of ϕ; its direction is the maximal direction	$\int_p^q \nabla\phi \cdot dR = \phi(q) - \phi(p)$
$\nabla \cdot F$	div F	Net out flux of F per unit volume	$\iiint_D \nabla \cdot F \, dv = \iint_s F \cdot n \, ds$
$\nabla \times F$	curl F	Swirl of F per unit area	$\iint_s \nabla \times F \cdot n \, ds = \int_c F \cdot dR$

Table 6: *Key differential operators*

differentiation The process (or method) of obtaining a derivative. There are symbolic and numerical methods.

See also *ODE (ordinary differential equation), PDE (partial differential equation)*

Dirac Notation Breaks the computation of the system's expectation value into an initial and final state. Using vectors to represent these states, and a set of combination rule, the expectation value can be computed.

The expectation value:

$\langle \alpha | \beta \rangle$

α is the intial state

β is the final state

$$\langle\alpha|\beta\rangle = \langle\beta|\alpha\rangle^*$$
$$\langle\alpha|\alpha\rangle \geq 0$$
$$\langle\alpha|\alpha\rangle = 0 \leftrightarrow |\alpha\rangle = 0$$

discriminant function

"a function of a set of variables that is evaluated for samples of events or objects and used as an aid in discriminating between or classifying them." *Websters*

distance

Euclidean

Measures magnitude of difference in two vectors. This is often used in geometry and to estimate text similarity.

Manhattan

Primarily used to estimate text similarity, but also used to estimate similarity non-numerical data sets.

distribution

A distribution has three forms of functions:

- A density function,
- A distribution function, and a
- Random selection generator that randomly selects items according to the distribution

Many distributions are formulated to support probability calculation. Described by mean, mode, median, and standard deviation; shape summarized by skew and kurtosis. Standard distributions, with their own parameterized shape, include:

- Beta-distribution, for modeling the completion times of activities
- Boltzmann distribution
- X^2 (Chi-squared) distribution,
- Exponential distribution,
- Fisher's distribution,
- Gamma distribution,
- Geometric distribution,
- Gumbel distribution
- Hypergeometic distribution,
- Log normal distribution,
- Logistic distribution,
- Logit distribution
- Negative binomial distribution,
- Student's distribution[1],
- Weibull distribution

see also *central limit theorem, power law, probability*

Table 7: *Distributions, normalized*

Distribution	*Density* $f(x)$	*Cumulative* $F(x)$	*Special properties*
Binomial	${}^nC_x p^x (1-p)^{n-x}$		For modeling the distribution of items that are in one of two bins.
Cauchy	$\frac{1}{\pi c + \frac{\pi x^2}{c}}$	$\frac{1}{\pi}\arctan\left(\frac{x}{c}\right)+\frac{1}{2}$	Similar to the normal distribution, with fatter tails; use the normal if possible.
Exponential	$\lambda e^{-\lambda x}$	$1-e^{-\lambda x}$	Modeling response time
Log-odds			
Normal	$\frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-x^2}{2\sigma^2}}$	$\frac{1}{2} erf\left(\frac{x}{\sigma\sqrt{2}}\right)+\frac{1}{2}$	Also called the Gaussian distribution
Poisson	$\frac{\lambda^x}{x!} e^{-\lambda}$		Modeling arrival or service time

[1] The Student's distribution was published by Gosset under the pseudonym Student.

Uniform $\begin{cases} a..b & \frac{1}{a-b} \\ else & 0 \end{cases}$ $\begin{cases} x < a & 0 \\ x \le b & \frac{x-a}{b-a} \\ x > b & 0 \end{cases}$

---

lambda = vt

v = the arrival rate average

x = the number of arrivals in time t

comparing distributions

If the units are different, convert the standard deviation to a ratio of the mean (eg a percentage of the mean).

selecting a distribution

- Is the independent variable is categorical?
  - Does it have only two possible categories? If so, use a Binomial distribution, or possible a Normal distribution
  - Otherwise use a power law
- Is the independent variable a task duration or completion time? If so, use a Beta-distribution
- Is the number of samples less than 30? Use a Student's t-distribution
- Is the distribution the sum of variables from distributions? Use a Normal distribution
- If there is no better distribution available, use a Normal distribution.

special properties

In the Normal distribution (and most other distributions) 99% of all data is in the range of $\langle m \rangle \pm 2.58\sigma$. This, and other such ranges, are list in the table below.

$P(-1\sigma < m - \langle m \rangle \le 1\sigma)$	68.27%
$P(-1.28\sigma < m - \langle m \rangle \le 1.28\sigma)$	90%
$P(-1.645\sigma < m - \langle m \rangle \le 1.645\sigma)$	80%
$P(-1.96\sigma < m - \langle m \rangle \le 1.96\sigma)$	95%
$P(-2\sigma < m - \langle m \rangle \le 2\sigma)$	95.45%
$P(-2.58\sigma < m - \langle m \rangle \le 2.58\sigma)$	99%
$P(-3\sigma < m - \langle m \rangle \le 3\sigma)$	99.73%

two distributions

When there are two different groups with similar distributions except that the centers differ slightly. Although similar, the difference in centers is small, the difference in the extremes is much larger. (A will be the one with the center to the left of B).

When dividing the distribution – everyone below threshold p goes into slot 1, everyone above threshold r goes into slot 3, and the rest go into slot 2.

< p – The # of elements in A less than p will be greater than the # of elements in B less than P; the difference is in the curve.

> r – The #'s are opposite, for the right hand side, due to the similarity.

If the members are assigned to bins based on the attribute X. A will outnumber B in those bins. Increase the range (number of gins) will increase the ratio of A to B. Note #Bins could be headcount.

- To increase the proportion of A in lower slots, lower the threshold
- To increase the proportion of B in lower lost slots, increase threshold (although it has a max proportion it can achieve)
- To increase the proportion of A in the higher slots, lower the upper threshold; but A has a max proportion that it can achieve.
- To increase the proportion of B in higher slots, increase the threshold.

division algebra

An algebra that supports some notion of divisions or reciprocal. Every non-zero element (x) in the algebra has an inverse (x^{-1}) that, when multiplied against the original) produces an identity element. There is no requirement that there be a defined procedure to find such an element.

duration
estimating

To estimate the duration of a task, or process – such as waiting for a certain number of people to arrive. Use a model like Poisson model.

see *Beta distribution, Poisson model*, and *trends (predicting)*.

dynamic programming

Any of a set of techniques for finding a sequence (or schedule) given constraints and a set of items.

See also *linear programming, programming*

dynamic system

Can exhibit one (or more?) of the following behaviours: amplification, noise reduction, memory, and oscillation. These can be used in combination to construct sophisticated mechanisms.

edge filter
Canny Edge Detector

1. Smooth the image using a Gaussian filter, to reduce noise
2. Find the intensity gradient, using a horizontal and vertical edge differential filters. This creates Gx, and Gy
3. Remove all edges whose |G| < threshold ("non-maximum suppression")
4. Trace edges
5. Group edges by locality and angle. Then map those <x,y,θ> to <r, θ> where r is the shortest distance to origin. Perform histogram look for peaks in histogram (each peak corresponds to line along the shapes edge).
6. Locate the maxima and its directional derivatives

A convolution transform used to detect edges. It performs a spatial derivative and computes an edge mask. The most common type is to use a single edge filter like so:

Alternatively a pair of transforms – a vertical and a horizontal (or real and imaginary) edge transform – can be used in conjunction to find the direction of an edge. This is more practical for some tasks.

See also *Canny edge detection, convolution filter*

Prewitt filter

$$G = \begin{pmatrix} G_x \\ G_y \end{pmatrix}$$

$$\theta = \arctan\left(\frac{G_y}{G_x}\right)$$

	Horizontal	Vertical
Prewitt filter	$\begin{pmatrix} 0.5 & 0.5 & 0.5 \\ 0 & 0 & 0 \\ -0.5 & -0.5 & -0.5 \end{pmatrix}$	$\begin{pmatrix} -0.5 & 0 & 0.5 \\ -0.5 & 0 & 0.5 \\ -0.5 & 0 & 0.5 \end{pmatrix}$
Prewitt differences	$\begin{pmatrix} 0 & 0.5 & 0 \\ 0 & 0 & 0 \\ 0 & -0.5 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 \\ -0.5 & 0 & 0.5 \\ 0 & 0 & 0 \end{pmatrix}$
Sobel Filter	$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}$	$\begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}$
Roberts cross	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ $out = g + g^*$	$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$

Scharr operator

$$\begin{pmatrix} 3 & 10 & 3 \\ 0 & 0 & 0 \\ -3 & -10 & -3 \end{pmatrix} \qquad \begin{pmatrix} 3 & 0 & -3 \\ 10 & 0 & -10 \\ 3 & 0 & -3 \end{pmatrix}$$

eigenvalue
square matrices

Method for calculating the eigenvalue for a square matrices can be done as follows:

1. Find the characteristic polynomial of the matrix
2. Find the roots of the polynomial
3. The roots are the eigenvalues of the matrix

problems that can be solved

$\mathbf{A}x = B$

$\mathbf{X}A = \mathbf{B}$

Prices per item: x, along the diagonal of X

Quantity per item: each row in **A** is a possible basked (distribution), A

Amount paid (total): each row in B or $|\mathbf{B}|$

eigenvector

Framing problems so that an eigenvector will rank it for you. Finding eigenvector using powers method, with spectral shifting, or gradient descent on PR quotient. Discard bottom eigenvector [1,1,…] which has zero eigenvalue

Eigenvector also minimizes the quadratic form when

$$m = (I - w)^T (I - w)$$
$$m_x = (I - w)^T (I_x - w_x)$$
$$m = (I - w)^T w - (I - w)^T wx$$

emulation

Approximates a function or system, esp. in control system theory;

See also *estimation, function approximation*

endogenous variable

A variable that is influenced by at least one other variable.

entropy
Boltzman

$S = k \ln \Omega$

k is the Boltzmann constant

information

Information entropy is

$$H(x) = -\sum_{i=1}^{N} p(x_i) \log_2 p(x_i)$$

where
p_i is the probability of event I, or being in state i
N is the number of distinct possible events that might occur

The goal, with scheduling and classification, is often to maximize entropy. With decision trees, the entropy predicts the average traversal depth to determine if an event is 0, 1, … n, or to classify it as such.

H is maximized when each $p_i = \frac{1}{N}$ and minimized when one event has probability of 1 (and the others have a probability of 0).

see also *choice time*

conditional

$$H(x|a) = -\sum_{i=1}^{N} p(x_i|a) \log_2 (x_i|a)$$
$$H(x|A) = \sum_{i=1}^{N} p(a_i) H(A|a_t)$$

normalized

Normalized by the largest possible value of H(x):

$$H^n(x) = \frac{H(x)}{\log_2 N}$$

e.g. $\sum p_i = 1$

expected entropy

$$\frac{-H}{N}\sum n_j$$

k = the number of different values for the attribute
n = the number of examples currently under consideration
n_j= the number of examples currently under consideration with the *j*th value of the attribute

equality

A few stable (and slow) equals are:

$$fabs(A-B) < \varepsilon$$
$$fabs(A/B-1) < \varepsilon$$
$$fabs(A-B) < \max(fabs(A), fabs(B)) * \varepsilon$$
$$fabs(A-B)/\max(\max(fabs(A), fabs(B)), DBL\_MIN) < \varepsilon$$

Finally, the following, but it can only be performed at machine precision (which is not always the same as a double):

$$\max(fabs(A), fabs(B)) == fabs(A-B) + \max(fabs(A), fabs(B))$$

equation forms

name	*example form*	*Technique to solve systems of*
differential equation		
linear equation	$c_0 = \sum_i c_i v_i$	Guassian elimination
parametric equation		
polynomial equation	$c_0 = \sum_i c_i v^i$	Gröbner basis

equivalence relation

"A relation between elements of a set that is: symmetric, transitive, reflexive and for any two elements either holds or does not hold." *Websters, 1987*

$$\forall_{x,y} f(x,y) \vee \sim f(x,y)$$

error

Any discrepancy between a computed, observed, or measured quantity and the true, specified, or theoretically correct value or condition

error function

Related to the cumulative probability function of a normal distribution.

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x dt\, e^{-t^2}$$

complement

$$erfc(x) = 1 - erf(x)$$

error term

Used in confidence intervals:

a± error term

estimation

A function that quickly approximates a value; usually these are derived from either a symbolic function, or measurements of the natural world. To construct an estimator function:

- Select the type or model of estimation:
  - A fit to a linear or polynomial function
  - A fit to a logarithmic or power function
  - A fit to a distribution function

- Queue model
- State machine (e.g. Markov model)
- With categorical data, sort into an index and use a power law (e.g. Zipf) or a right-only distribution to map the index number to the measured quantity.

- If need be, sample the given function
- Use the sample data to compute the necessary parameters.
- Finalize the equation

Share price models will be wrong from time to time (every few months); Black-Sholes is good for only a few months.

see also *distribution function, forecasting, function approximation, inverse with empirical data, likelihood estimator, regression, trends*

point estimation — Selects a specific scalar or vector as the best estimate of a value

interval estimation — Locates the value within a region (rather than a specific point) and associates a confidence with it.

see also *probability density function*

exponentiation — Power method is to keep squaring values to form the exponents 2n, then multiply those in a binary number fashion.

extreme value theory

factor analysis — Correlations among many variables are analyzed to identify closely related clusters of variables. The most important use of factor analysis is to make problems manageable by eliminating (or reducing) variables.

Stephen Gould, the Mismeasure of Man, p246

"Factor analysis is a mathematical technique for reducing a complex system f correlations into fewer dimensions. It works, literally, by factoring a matrix, usually a matrix of correlation coefficients. (Remember the high-school algebra exercise called 'factoring,' where you simplified horrendous expressions by removing common multipliers of all terms?) Geometrically, the process of factoring amounts to placing axes through a football of points. In the 100-dimensional case, we are not likely to recover enough information on a single line down the hyper-football's long axis – a line called the first *principal component.* We will need additional axes. By convention, we represent the second dimension by a line perpendicular to the first principal component. This second axis, or second principal component, is defined as the line that resolves more of the remaining variation than any other line that could be drawn perpendicular to the first principal component… Subsequent lines would be perpendicular to all previous axes, and would resolve a steadily decreasing amount of remaining variation… If we elect to stop at 5 dimensions, we may achieve a considerable simplification at the acceptable price of minimal loss of information."

factorial
$n!$

A factorial can be accurately computed quickly by using the $\Gamma(\ )$ function:

$$n! = \Gamma(n-1)$$

Stirling approximation

A factorial can be computed by using an Stirling approximation (there are several compact approximations, but all require that n is large):

$$n! \approx \sqrt{2n\pi}\left(\frac{n}{e}\right)^n$$

$$n! = \sqrt{2n\pi}n^n e^{-n}(1 + C_n)$$

$$C_n = \sum_{m=1} -1^m \frac{2m+1}{2(m+1)n}$$

$$Error = C_n * 100\%$$

$$\ln n! = n\ln n - n + \tfrac{1}{2}\ln 2\pi m + \ldots$$

Note that:

$$n! \approx n^n$$

fast multipole method

fibonacci

$$= \left(\phi^n - (1-\phi)^n\right)\tfrac{1}{\sqrt{5}}$$

Recursion definition does it do this because of binomial expansion… Combinatorial. Use powers method

filter

point and neighborhood. See edge filter, FIR, fourier transform.

FIR
finite impulse response

see also *IIR*

Fischer Score

Used to compare features for their discrimination ability. Looks at the mean and stddev incidence for a feature *g* in two categories.

$$\frac{(\mu_0(g) - \mu_1(g))^2}{\sigma_0(g) + \sigma_1(g)}$$

fitting

Process of selecting a formula, by taking empirical data, and identifying the coefficients that best match the data. The most commonly accepted a criterion for 'best coefficients' is the least-squares method.

see also *regression*

fixed point

Keep applying the rules until nothing changes any more. This is the fixed point.

forecasting
requirements

Herbert Simon, Science of the Artificial

Requires:

- "a theoretical understanding of the phenomena to be predicted, as a basis for the prediction model, [or]
- "Phenomena that are sufficiently regular that they can be simply extrapolated."
- "reliable data about the initial conditions"

"Construct alternative scenarios for the future and analyze their sensitivity to errors in theory and data."

"Select some planning horizons"

"Concentrate .. analytical resources on examining alternative target states for the system for the short, middle and long run"

Target state: "upper bounds on the quantities of" resources used, or problems.

"Satisfy [yourself] that its realizability is not unduly sensitive to unpredictables."

"Construct paths that lead from present to futures"

Note. "Future contingencies that have no implications for present comittment have no relevance to design."

approaches

Forecasting is used to constrain planning by keeping the focused on likely possibilities (e.g. away from the unlikely) and warranted circumstances of importance. The approaches include *time series extrapolation,* a *cause-and-effect model,* and a *failure rate* model. The first step is to decompose the time series into four parts:

- The overall trend
- The cycles
- Seasonality
- Residuals – aka the error factor – the part of the forecast that one will be off by.

Project these into the future and recombine.

cause and effect model

A cause and effect model has the following properties

- It takes a long time to construct, and often is the most expensive approach.
- It requires a substantial historical dataset.
- It has the advantage of producing accurate predictions.

catastrophe theory

Used to identify future possibilities that can't be extrapolated from trends and sample data; these are intrinsically rare events.

time series extrapolation

Time series extrapolation is often used. It has the following properties

- It is faster and requires little money
- It requires only a small window into the past
- It has limited accuracy, so it shouldn't be used where accuracy is critical.

The process tracks the recent history and extrapolates from these the most likely results.

see also *trends (linear extrapolation)*

Fourier analysis
FFT

FFT can often be used to solve (or approximate) anything with a sin, cos, $e^{i\theta}$, affine translations (offset or movement), and convolutions. Simpler techniques, such as converting to polar notation and examining a histogram may be more suitable.

See also *affine transform, cancos filter, convolution problems FIR, IIR, Lanczos filter*

correlation & image matching

Useful for seeing of the images are the same, under an affine transform.

F_n = FFT of image n

Result = $FFT^{-1}(F_1 * F_2)$

Look for the largest magnitude (see below for larger uses)

registration: finding the offset of two images

the Fourier shift theorem, which guarantees that the phase of a specifically defined "ratio" is equal to the phase difference between the images. To find displacement between two images:

1. Compute the Fourier transform of the two images to give F_1, F_2
2. Compute the Ratio = A_1 conj(A_2) / | A_1 A_2 |
3. Apply Inverse Fourier transform to the Ratio= FFT^{-1}(*Ratio*)

By applying the Inverse Fourier transform to the Ratio, we get a matrix of numbers returned that is zero everywhere except for a small area around a single point. Find the x and y indices in the matrix of the max value. This is the displacement (x0,y0) that is needed to optimally register the images. The max value is the degree to which the images match. (This process can be done more efficiently with discrete cosine transform)

finding the affine transform between two images

1. let A_n= Abs(F_n)
2. Remove the low frequency noise from A_1 and A_2 using a high pass filter
3. Convert the images to polar coordinate form (square though: same number of rows & columns). use bilinear interpolation to sample the image
4. Find the offset (x', y') of the polar images

$scale_x$ = b ** (x' mod rows)

$scale_y$ = b ** (y' mod columns)

$angle_x$ = 180*(x'/rows)/columns

$angle_y$ = 180*(y'/rows)/columns

Note: This technique assumes that the images are of sufficient quality and that their differences can be approximated by an affine transform. Other techniques to find perspective warp transforms (etc) may be applied instead

function
approximation

Attempts to find a simpler function that is a close fit, or possibly using a particular type of formula. Commonly used techniques include series and polynomial approximations. (Radius of convergence is a key criterion of series approximations).

If the function has equivalency (under a various tests), or testing shows the error under various approximations is low, one can derive a form that is a polynomial.

Continuous, well understood about the normal form – regression, function fit

Discrete, probabilistic statistics – Bayes, conditional probability

Input discrete vs continuous: classification – side of partition or place on grid.

see also *estimation*

Discrete	*Continuous*
Table	Rtree

Table 8: *Methods of approximating or emulating a function, based on its range*

When both are continuous: regression

forms

When changing the representation of a formula, if it is already a polynomial, see polynomials for alternative forms;

closed form – the function is defined without recursion, integration, differentiations, relations, and set theoretics.

Cauchy form

Polynomial

primitive recursive

Taylor's form

minimization

Similar to a root finder, it tries to find a value that will make the function as close to zero as possible.

function
periodic

f(x)=f(x+p)

odd: f(-x)=f(x)

even: f(-x)=-f(x)

functional explanation

An explanation that defines a system and relates the elements to each other in terms of what each element does. (Rather than in terms of the physical mechanism that performs the action.)

fundamental matrix

Used with stereo geometry (3D scene projected onto two different 2D planes).

Estimating. Take 8 non-co-planar points. Put the points in a matrix, perform rank deprivation and take the SVD. The last column of V is used as the parameters to construct the fundamental matrix.

Assumes a pin-hole camera.

see also *stereo geometry*

Gaussian distribution

Another name for the Normal distribution

see *Normal distribution*

Gaussian elimination

If necessary interchange the first equation with another so that x appears in the first equation.

Eliminate x from every equation but the first by adding the appropriate multiple of the first equation

Temporarily ignore first equation. Solve x2…xn ->

Can use gradient descent to approximate results in many cases. Note: gradient descent finds local minima, so randomly try different places until we're not finding a better place.

gaussian mixed models

Allows taking a lumpy distribution and breaking it down into several Gaussian distributions

geodesic

The curve along a surface, which marks the shortest distance between two neighboring points.

golden rectangle

Taking a square from the rectangle will leave a new rectangle with the same proportions.

Gompertz law of mortality

$\log(death\ rate) = c_1 t + c_0$

See also *Weibull power law*

graph

compression

a small version of the graph that preserves the desired properties.

junction tree algorithm

converts a graph into a tree

Laplace matrix

From unweighted adjacency:

$$L_{i,j} = \begin{cases} \text{degree}(v_i) & i = j \\ -1 & i \neq j, adjacent(v_i, v_j) \\ 0 & otherwise \end{cases}$$

normalized Laplace matrix

$$L_{i,j} = \begin{cases} 1 & i = j, \text{degree}(v_i) \neq 0 \\ -[\text{degree}(v_i)\text{degree}(v_j)]^{-0.5} & i \neq j, adjacent(v_i, v_j) \\ 0 & otherwise \end{cases}$$

Building a Laplace from an adjacency matrix

$$D_{i,i} = \sum_j adjacency(i, j)$$

$$L = D - adjacency$$

D[i,i] = Sum_j Adjacency(i,j)

l = D - Adjacency

(^ I don't think that is the same as the first. Seems to be weighted)

Building an adjacency matrix from a Laplace matrix

number of independent graphs

1. l = convert to a Laplacian matrix
2. Compute the n eigenvalues for l. Call this lambda[]
3. Count the number of zeros in lambda[]. This the number of independent graphs.

path

Can compute a signature of a path, primarily as a hash of the

planarity

representing as a matrix

Adjacency matrix (which can have eigenmodes)

Laplace matrix

graphics

Tend to use homogenous coordinates

- Transform, Clipping, and Lighting (see shading)
- Vertex skinning
- Key frame interpolation
- Triangle setup

Gröbner basis

Used to solve systems of polynomial equations

Gumbel distribution

Used to predict extremes

Parameter	*Description*
μ	Easy to measure as the peak, or by linear regression of the data
β	Determine from the mean

Table 9: *Gumbel distribution parameters*

Property	*Form*
probability density function	$P(x) = \frac{z}{\beta} e^{-z}$ $z = e^{-\frac{x-\mu}{\beta}}$

Table 10: *Gumbel distribution properties*

Property	*Form*
cumulative function	$= e^{-z}$
mean	$\mu + \beta\gamma$ (note that y is Eulers constant)
median	$\mu - \beta \ln(\ln 2)$
mode	μ
Variance	$\frac{\pi^2}{6}\beta^2$

Helmhotz equation $\nabla^2 f + k^2 f = 0$

see also *Poisson equation*

Hermite Recurrent d/dx functions

$$H_o(x) = e^{-x^2/2}$$

$$H_n(x) = \frac{1}{\sqrt{2n}}\left[x - \frac{d}{dx}\right]H_{n-1}(x)$$

homograph matrix

$$\begin{bmatrix} r_1 & r_2 & t_1 \\ r_3 & r_4 & t_2 \\ p_1 & p_2 & 1 \end{bmatrix}$$

r: rotation elements

t: translation elements

p: perspective warp

planar homography assumes two images are on the same plane, but the plane is perceived in a 3D world. So plane #1 is the yard, and plane #2 is the flat image of the camera.

Hough transform Shape recognition. Very tricky. Feature detection is edge based.

See also *Canny edge detection*

idempotent A function that always provides the same result when given the same parameters.

IIR
infinite impulse response

IIR is a special purpose virtual machine. Low-pass filters, high-pass filters, combinations of those filters, and even equalizers, can be specified, and use very specific techniques (like compiler) to convert them into an IIR implementation. (These can be compiles into a hardware circuit).

IIRs are easy to implement - and take less CPU power than other methods. But sometimes they sound poor; if they sound too bad, you'll want to use a different technique.

Has two orders: forward and past; in some cases may require more math than a FIR.

There are four kinds of filter specifications that can be converted into IIR coefficients: Butterworth, Chebyshev (type 1, and type 2) and elliptic

see also *FIR*

```
Out[0]=Sample[0];
Out[1]=Sample[1];
for (int Idx = 2; Idx < N; Idx++)
{
  Out[Idx] =
      B0 * Sample[Idx]
    + B1 * Sample[Idx-1]
    + B2 * Sample[Idx-2]
    // … more like this …
    // Next, the feed back
    - A1 * Out[Idx-1]
    - A2 * Out[Idx-2]
    // … more like this …
```

```
    ;
}
```

incidence matrix

An incidence matrix can be used to represent

- The links in a graph or network,
- The transition between states in a state machine
- Who-knows-how in society or organizations

The relative frequency of each 'item' (state, person or node) is called the *stationary distribution* and is the eigenvector of the matrix. There are issues with roles and cliques. Multiplying a matrix by itself you can determine:

- The number of 2 and 3-step paths from I to j
- The identity of the central figure in the group
- Cliques
- Dominance relations in the group

see also *graph, matrix (stochastic), network, transition matrix*

normal form

The normal form of an incidence matrix has the source node as columns, and the destination nodes as rows.

induction
first principle

P_1 (the *basis*) is true (e.g. proven). P_2 thru P_n can be deduced from their (respective) predecessor – e.g. P_2 from P_1, P_3 from P_2, etc.

second principle

P_1 (the *basis*) is true (e.g. proven). P_2 thru P_n can each be deduced from *all* of their (respective) predecessors – e.g. P_2 from P_1, P_3 from P_2, etc.

integer approximation

integral

Path Integral, Surface Integral, Volume integral

$$\oint \mathbf{A}_1 \cdot d\mathbf{l} = \oint \mathbf{A}_2 \cdot d\mathbf{s}$$

d**l** is tanget to the path L. Generally goes counterclockwise around the path L . Be careful in choosing the sense (+/-) as this choice will affect the sign of $\oint \mathbf{A}_1 \cdot d\mathbf{l}$.

d**s** is normal to the area S. Generally out of the page. S is an open surface. Be careful in choose the sense (+/-) as this choice will affect the sign of $\oint \mathbf{A}_2 \cdot d\mathbf{s}$

integration methods

Romberg, Adaptive (when functions change rapidly within interval), other. Trapezoidal, Monte Carlo method (useful for integrating over for or more dimensions).

Romberg method

This iterates of the following process (n is the iteration number)

1. Fit a polynomial of order n to the passed function. First a line fit, then a parabolic fit, etc. The fit is usually done by sampling only *n+1* points, the end points, and equidistant interior points.
2. Integrate the polynomial.
3. Stop when the absolute difference between the current result and the previous result is less than a specified error amount.

interpolation

Aitkens method

inverse

Finding the inverse of a function can be difficult – they are typically not well posed. Using observed data, determine values of one or more model parameters.

with empirical data

Using observed data, attempts to determine the relationship with the specified independent variable (reversing the measurement order). This problem is seldom well-posed. Most

techniques seek to determine values of one or more model parameters. These techniques include:

- Singular value decomposition (the application of a change in orthogonal co-ordinates)
- Principal component analysis (the simplification to major co-ordinates, or reduction in the number of dimensions)
- Independent component analysis
- Empirical orthogonal functions
- Canonical correlation analysis

Jacobian matrix

Relates small changes in endpoint position to small changes in input configuration. Ji,I = dXi,dtheta_j I end point position coordinate; j is the input

Kalman filter

Used to merge two or more noisy signals together to estimate a proper signal. see *Kalman filter (engineering)*

kernel machine
kernel method

The kernel method works by transforming the space a problem is represented in. (The technique may come from operator theory). Takes an algorithm that works on vectors or lists, formulated entirely in terms of dot products. Pull out the dot product operations and replace with a *kernel function*. The function has to meet certain criterion (in the Mercer's theorem) for this replacement to work; namely it must be semi-definite. Examples of kernel machines includes:

- Support vector machine
- Relevance vector machine
- Kernel recursive least squares
- Adaptive sparseness
- Smooth relevance vector machine

see also *function minimization, Fourier transform, Laplace transform, partitions*

uses

Regression/prediction

Classification

Feature selection –which are important

Lagrangian relaxation

Solves integer programming problems by treating variables as continuous.

Lanczos filter

Used to resample a signal for an FFT, with less ringing and issues than other windowing methods.

$$L_k(x) = \begin{cases} \dfrac{k\sin(\pi x)\sin\left(\dfrac{\pi}{k}x\right)}{\pi^2 x^2} & -k < x < k, x \neq 0 \\ 1 & x = 0 \\ 0 & otherwise \end{cases}$$

$$L_k(x,y) = L_k(x) \times L_k(y)$$

where k is the number of lobes (e.g. 2,3,or 4)

see also *fourier analysis*

LAPACK
LINPACK
BLAS

Successor to LINPACK, built on BLAS. Broken into sections that are operations of vectors on vectors, matrix on vector, matrix on matrix, and eigenvalues of matrices. Layered on top of platform-specific small matrix block operations.

BLAS

Level 1: vector operations

Level 2: matrix-vector operations

Level 3: matrix-matrix operations

Laplace's equation

$$\nabla^2 u = 0$$

$$u = \sum_{n=1}^{\infty} b_n \begin{cases} e^{ky} \sin kx \\ e^{-ky} \sin kx \\ e^{ky} \cos kx \\ e^{-ky} \cos kx \end{cases}$$

$$b_n = \frac{2}{\rho} \int dx \; f(x) \sin \frac{n\pi x}{\rho}$$

$$u = (1-i) \sum_{n=1}^{\infty} c_n e^{-k(ix+y)} \left(e^{2ky} + 1\right)\left(e^{2ikx} - i\right)$$

$$c_n = \frac{1}{2l} \int_{-l}^{l} dx \; f(x) e^{-ikx}$$

f(x) = boundary condition

Laplacian operator

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \ldots$$

large numbers law of

$$\lim_{N \to \infty} \left[P(E) - RF_N(E) \right] = 0$$

where

N is the number of samples

E is an event being observed

$RF_N(E)$ is the relative frequency of event E in N samples

P(E) is the probability of E

latent semantic analysis prep work

A: each row corresponds to a term; each column a document (or category). The cell is the weight the term has within each

Using SVD break A into

$$\mathbf{A} = \mathbf{USV}^T$$

Reduce the rank to k. Keep the first k columns of **U**,**V**, and **S**. Keep the first k rows of **S**.

Compute the inverse of S_k

Compute $\mathbf{V}'_k = \mathbf{V}_k$ w

where w is a vector, such that $w_i = |d_i|$. d_i is the vector for row i of $\mathbf{V}_k$.

query

q: each row corresponds to a term; The cell is the weight the term has within each

$$q' = U_k S_k^{-1} q$$

$$tmp = \mathbf{V}'_k q'^T / |q'|$$

The best document (category) corresponds to the row in tmp with the maximum value

weight

The weight is composed of global weight for the term, and a local weight (specific to the document)

$$A_{i,j} = l_{i,j} \mathbf{g}_i$$

$f_{i,j}$ is the count of occurrences of term i in document j

gf_i # times term i appears total

df_i # documents in which term i appears

n number of documents

Type	*formulae*
augnorm	$l_{i,j} = 0.5 \; (1 + f_{i,j} / \max_i(f_{i,j}))$

Table 11: *Local weighting functions*

binary	$l_{i,j} = 1$
log	$l_{i,j} = \log (1 + f_{i,j})$
term frequency	$l_{i,j} = f_{i,j}$

Table 12: *global weighting functions*

Type	*formulae*
binary	$g_i = 1$
entropy	$g_i = 1 + \log n \sum_j p_{ij} \log p_{ij}$ $p_{ij} = f_{ij} / gf_i$
gfldf	g_i = # times term i appears total / df_i
ldf	$l_{i,j} = \log_2 (n / (1 + df_i))$
normal	$l_{i,j} = f_{i,j}$

least squares method

A goodness of fit measure between a function and an approximation. The quality of the approximation is estimated by taking the magnitude of the vector distance between the two. To do this, at each point there is a value for both functions, compute the difference between the values. Square this value. Sum all of these values together; this sum is the score – the candidate with the smallest sum being the best.

$$\mathbf{c} = \mathbf{Q}^{-1}\mathbf{U}$$

see also *linear regression, sampling section*

linear regression

$$f(x) = c_0 + c_1 x$$

$$\mathbf{c} = \begin{bmatrix} c_0 \\ c_1 \end{bmatrix}$$

$$\mathbf{Q} = \begin{pmatrix} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix}$$

$$\mathbf{U} = \begin{bmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{bmatrix}$$

polynomial regression

$$f(x) = c_0 + c_1 x + \cdots + c_k x^k$$

$$\mathbf{c} = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_k \end{bmatrix}$$

$$\mathbf{Q} = \begin{pmatrix} N & \sum_i x_i & \cdots & \sum_i x_i^k \\ \sum_i x_i & \sum_i x_i^2 & \cdots & \sum_i x_i^{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_i x_i^k & \sum_i x_i^{k+1} & \cdots & \sum_i x_i^{2k} \end{pmatrix}$$

$$\mathbf{U} = \begin{bmatrix} \sum_i y_i \\ \sum_i x_i y_i \\ \vdots \\ \sum_i x_i^k y_i \end{bmatrix}$$

Legendre polynomials

Useful for spherical symmetry problems. Spherical charge distribution

likelihood estimator

Used to estimate the probability of an event from sample data. If the data does not have all of the symbols and their variations well represented (it is "sparse"), a technique like the *expected likelihood estimator* should be employed.

see also *estimator, probability, sampling, stoplist*

estimating

Estimating independent probability from sample data: get data, does data have everything well represented (does it reflect the full variation)? If so use a Maximum Likelihood Estimator; otherwise use an Expected Likelihood Estimator.

The probability for some events may be specified; for other is may be calculated from structural relationship data. Some symbols are so common their probability will be significantly underestimated; a "stop list" is used to ignore them.

expected likelihood estimator

This technique is used when the sample data is too small to use other methods. This handles those cases where x can occur, but it was not represented in the samples. There are two different, but similar, techniques. The first adds a pseudo-count to each possible sample:

$$P(x) = \frac{\varepsilon + \#\,Times\ x\ occurred}{\#\,Total\ Samples + \varepsilon * \#\,Symbols}$$

$\varepsilon \approx \frac{1}{2}$

This can greatly underestimate the probability for symbols and events that did occur, but in small numbers, if the data set is missing many symbols overall. The second case only adds the pseudo-count to those symbols that weren't represented in the data set:

$$P(x) = \begin{cases} \dfrac{\#\,Times\ x\ occurred}{\#\,Total\ Samples + \varepsilon * \#\,Symbols\ without\ Count} & x\ in\ dataset \\ \varepsilon & otherwise \end{cases}$$

$\varepsilon \approx 10^{-3}$

probabilitity of a sequence $P(e_n | e_0 \ldots e_{n-1})$

$= \dfrac{C(e_0 \ldots e_n)}{C(e_0 \ldots e_{n-1})}$ Many methods of translating from observed frequency to probability over-estimate the probability some events, under-estimate the probability of very-command and very-rare events. See also smoothed stuff.

Table 13: *Methods of computing a sequence's probability givens its frequency count*

Form	*Description*
$\frac{1}{N} C(e_0 \ldots e_n)$	Maximum likelihood estimator
$\frac{1}{N+B} C(e_0 \ldots e_n) + \frac{1}{N+B}$	Laplaces Law to smooth. B is the number of bins the training instances are divided into
$\frac{1}{N+B\lambda} C(e_0 \ldots e_n) + \frac{\lambda}{N+B\lambda}$	Lidstone & Keffret-Perks

N is the total number of samples

line length estimating

see *queue (estimating)*

linear program

Used when:

- The objective is to minimize (or maximize) a quantity (or expression)
- Operating under certainty
- There are constraints and
  - Alternative options

specification

The specification has is notated with the form

$$\begin{aligned} &\text{minimize } cx \\ &\text{subject to } Ax = b \\ &\qquad\qquad x \geq 0 \end{aligned}$$

where

x = the vector to be solved for
A= matrix of known coefficients
c,b= Vectors of known coefficients

The first statement is call the goal, the subject is the constraints

solver

Two kinds of iterative solvers:

- Interior point algorithms; these have polynomial complexity
- Simplex techniques, worst case is exponential number of iterations, but that is rare. Iteration is easier to implement.

iterative solver

produce a series of approximate solutions. interior-point methods: polynomial complexity. simplex methods: worst case is exponential, command case is fast / easy.

direct solvers

compute exact solutions; Gaussian elimination. Drawback: memory intensive.

uses

Typical business uses:

- Add a new product or activity (each would be a variable)
- Remove a product or activity
- Adjustment of technological coefficient. Changes in profit/unit or cost/unit of the objective function. Changes in resource requirements or units of products or activities in constraints.
- Constraints (requirements): add or remove a constraint, changes in available resources or capabilities of production / service centers or limits on demands.

linear regression
stepwise

Alternates between forward and backward steps to develop a best-fit model. The forward steps add variables, so long as the goodness of fit *improves*; the goal is to find the variables with best goodness of fit. The backward steps removes less important variables, so long as the goodness-of fit doesn't get too bad.

linear system
equivalent systems

Both have n unknowns, both have the same set of solutions

Logit regression analysis

Considers how to estimate a bivalent variable's conditional probability from empirical factors:

1. First, transform of the probability (p and 1-p, respectively) into a continuous (and unbounded) range: $\log \frac{p}{1-p}$
2. Next, apply a regression between events and the transformed probability
3. Finally, inverse the transform to estimate the probability based on these factors ($X_1..X_n$):
   $\tilde{P} = \frac{e^{\alpha+\beta_1 x_1+\ldots}}{1+e^{\alpha+\beta_1 x_1+\ldots}}$

See also *classification*

longitudinal study

Observes one group of subjects repeatedly over time

Mahalanobis distance

A distance of features using statistics. Measures the correlation of a variable against a known sample.

Markov model
linear birth process

The linear Markov birth process model that assumes that the probability of an arrival is dependent of the size of the population and the time interval. The probability of having *m* arrivals in time *t* is:

$$P(m,t) = \left(1-e^{-\lambda t}\right)^{m-1} e^{-\lambda t}$$

where λ is the arrival rate (number of units/sec)

see also *Bayesian model, matrix (simulation), queue distribution, stochastics, trends (predicting)*

channel matrix

Describes probability transitions

emission matrix

E_{ij} = probability of emitting symbol s_j given we are in state i

transition matrix

T_{ij} = probability of transition from state i to state j

matrix characteristic polynomial

$$f(t) = \det(tI - A)$$

Example:

$$A = \begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix}$$

1. $f(t) = \det(tI - A)$
2. $f(t) = (t-2)t - 1(-1)$
3. $f(t) = t^2 - 2t + 1$ *normalized form*

computation

Computing an efficient multiplication order for ABC….

Computing the inverse of a lower-triangular matrix

Converting a matrix to lower triangular, with…

APSP conversion of weighted incidence matrix.

Carrying out a graph traversal by forming a path (APSP vs A* vs Dijkstra) and executing the path.

decomposition

Rearranges a matrix – or the matrices in a multiplication – into a special pair of matrices. When the two are multiplied, with result equivalent to (but much faster to obtain) multiplying the original matrices. The two matrices are usually triangular, or orthogonal.

Form	*Method*	*Special properties*
$A=LL^T$	Cholesky decomposition	L is a triangular matrix
$A=QR$	QR decomposition	Q is an orthogonal, R is an upper triangular matrix.
$A=UsV^T$	Singular value decomposition.	U, V are orthogonal matrices; s is a diagonal matrix. A's size must be NxM with N>=M.
$P^TA=LU$	LU decomposition	L, U are lower and upper triangular matrices. P is a permutation matrix. P and A must be square matrices of the same size

Table 14: *Decomposition methods*

Form	*Matrix form*	*Complexity*
A^{-1}	General	$O(n^3)$[2]
	Diagonal	O(n)
AB	General	O(#rows in A * #columns in A * #columns in B)
	A or B is diagonal	O(#rows in A * #columns in B)
	A and B is diagonal	O(Max(#columns in A, #columns in B))
A^k	A is square	$O(2n^3\log_2 k)$[3]
		O(???) for the Cayley-Hamilton method
	A diagonal	L, U are lower and upper triangular matrices. P and A must be square matrices of the same size

Table 15: *Matrix operation methods*

[2] There is the Strassen method, but with its emphasis on addition, it is not faster on modern machines (who have multiplication at rates approaching addition operation)

[3] I don't know how fast it is for the Jordan canonical form method

exponentiation

Efficiently raising a matrix to a power. Pad approximation. Golub and Vin loan. The binary method is the most efficient for integer powers; This tracks the results of raising **A** to a power of two, and multiplies those to form $\mathbf{A}^k$

exponentiating Jordan normal form method.

Raising a matrix to a power sometimes may be more efficiently accomplished by diagonalizing the matrix first (using Jordan normal form), and raising the diagonal to a power:

$$\mathbf{A}^k = \mathbf{P}\mathbf{D}^k\mathbf{P}^{-1}$$

Cayley-Hamilton method

This finds the characteristic polynomial (for square matrices) from

$$p(\lambda) = \det(\lambda\mathbf{I} - \mathbf{A})$$

The polynomial is symbolically re-arranged, with appropriate substitutions (often recursively), until a form is a suitable form is found, often appearing as:

$$\mathbf{A}^k = c_1\mathbf{A} - c_0\mathbf{I}$$

see also *characteristic polynomial*

forms

Echelon, row echelon, Sparse, triangular (lower and upper), along diagonal, etc.

inverse

A matrix can be inverted into another matrix, $B = A^{-1}$. In some cases, it is more convenient to create a pair of matrices that serve as the inverse

To invert a matrix A into a single inverse matrix,

1. Form an n x n matrix [A | I]
2. Use elementary row operations to transform [A | I] to the form [I | B]

See *left inverse* and *right inverse*

left inverse

A left inverse matrix is one where $A_l^{-1}A = I$. (The matrix A must not have a determinant of 0). The inverse is in the form $\frac{B}{C}$:

D,E,F, … are the rows of matrix A;

$$B = \begin{bmatrix} \dots & E \times F & \dots \\ \dots & F \times D & \dots \\ \dots & D \times E & \dots \end{bmatrix}$$

$$C = [D, E, F]$$

note that the left hand side of cross-product is D,E,F,.. rolled left by one; and the right hand side of the product is D,E,F… rolled right by one.

See also *Affine Transform (Cramer's rule), right inverse*

right inverse

A right inverse matrix is one where $AA_r^{-1} = I$. (The matrix A must not have a determinant of 0). The inverse is in the form $\frac{B}{C}$:

D,E,F, … are the rows of matrix A;

$$B = \begin{bmatrix} \vdots & \vdots & \vdots \\ E \times F & F \times D & D \times E \\ \vdots & \vdots & \vdots \end{bmatrix}$$

$$C = [D, E, F]$$

note that the left hand side of cross-product is D,E,F,.. rolled left by one; and the right hand side of the product is D,E,F… rolled right by one.

See also *Affine Transform (Cramer's rule)*

norms

Based on the order of the matrix's values, and used by various linear algebra problems. There are four common types of matrix norms:

- The norm in L1 space
- The norm in L2 space
- Euclidean norm
- The max-norm or infinity norm

see also *condition number*

permutation matrix — A matrix where each row and column contains only a single cell with a non-zero value, a one. Identity matrices are permutation, and permutation matrices are a type of stochastic matrix.

solving system of equations — the constraints – linear equations – and items to be found (variables) can be solved by matrix methods

stochastic matrix — A type of transition matrix, where columns represent the 'from' state and rows represent the 'to' states. The cell represents the probability of this transition; the column represents the probability distribution. Each cell must be ≥ 0 and ≤ 1, and each column must add to 1. The eigenvalue of such a matrix is 1.

Avoid models with hundred states. Track the history of each individual; probability, cost, and utility may depend on the history. Consider sparse matrices and subsets.

Note: the above form is called a *left stochastic matrix;* many texts describing Markov chains use a *right stochastic matrix,* a form that is equivalent to the transposition of the above.

See also *probability (conditional)*

McCarthy91 — A recursive function that evaluates to 91 for each value from 1 to 101:

$$M(x)=\begin{cases} M(M(x+11)) & x\le 100 \\ x-10 & x>100 \end{cases}$$

mean — The simple mean of a finite, discrete set is typically expressed as:

$$\langle x\rangle=\frac{1}{n}\sum x_i$$

see *average, distribution, expected value, geometric mean, normal distribution (estimating the distribution), probability density function*

geometric mean — $\log\langle x\rangle=\frac{1}{n}\sum\log x_i$

harmonic mean — $\frac{1}{\langle x\rangle}=\frac{1}{n}\sum\frac{1}{x_i}$

The harmonic mean is always less than or equal to the arithmetic mean, penalizing uneven performance more than the other.

of probability density function — Using a probability density function:

$$\left\langle f(x)\right\rangle_{x0}^{x1}=\frac{1}{x1-x0}\int_{x0}^{x1}dx\,f(x)$$

mode — The most likely outcome

model — System of equations including differential equations to describe behaviour.

modeling

identifying the most significant variables / methods

- Linear regression
- Logistic Regression
- Discriminant Analysis
- Naïve Bayesian models
- Predictive models
  - Additive regressions
  - Decisions trees

- Neural networks
- Support vector machines
- Bayesian networks

- Combining models
- Types of methods
  - Classification
  - Clustering
  - Analytical models
  - Probabilistic or statistical modeling
  - Prediction and estimation
  - Dependency analysis
  - Search
  - Optimization

business modeling

Forecasting:

Regression: The relationship should be true, even if different data is input.

Evaluation of the model: changes, minimally fit, where are the errors? Different value error models.

Data, data variables.

Cost driver: the factor that determines the annual amount spent; can be hard to identify.

Seasonality: many things are coupled (with the / via) time of your budget cycles, shipping.

Theoretical terms:

Costs, depreciation, GDP, inflation

Valuation

Some values are purchased, received from others, or specified by corporate view.

Inference of properties, market segmentation, marketing campaigns.

modulo

in the domain of

Monte Carlo methods

Used when Bayesian networks are too complicated (they 'are' $O(n^2)$). Integration is useful when integrating over four or more dimensions. Circle and square example.

Models. Combining models often diverge from empirical reality. The reasons for the divergence are informative.

multiples

When computing multiples – using a counted set S – you must bear in mind if you are concerned about different kinds of items. For instance, do you only care about *cards* – 52 in a deck – or do you care about *Aces, Kings, etc.* – 13 kinds, with 4 each.

$${}^{n}M_{r} = \begin{cases} n_1 \cdot \ldots n_m & \text{When there is } m \text{ kinds of items} \\ n^x & \text{When there is only kind of item in the set} \end{cases}$$

N-body problem

Efficiently compute systems of the form:

$$\forall_{i=1\ldots n} f(x_i) = \sum_{j}^{n} k(x_i, u_j) s(y_i)$$

For example, gravity (f is acceleration)

$$k(\vec{p}_i, \vec{p}_j) = \frac{G}{\left|\vec{p}_i - \vec{p}_j\right|^2}$$
$$s(m_j) = mass\ of\ \ j$$

see also *fast multipole method.*

negative frequency

The wave pattern is reversed from the equivalent positive frequency pattern

network Span, tour, shortest path, APSP matrix

non-parametric methods When working with populations too small to qualify for a standard distribution, or working with categorical data.

- Chi-squared test
- Verifying the median: The number of items below the median should be the same as the number of items above the median. (Ignore the items that are the same as the median)
- Wilcoxon test (similar to above)
- Testing randomness: Runs test. Look at a sequence *n* items long, and count the number of runs of the same item. The number of runs of any given length should be $\frac{n}{2}+1$
- Equality of medians from two populations: The Mann-Whitney test
- Equality of medians from more than two populations: the Kruskal-Wallis test

see all *scalar values (types of)*

normal distribution A probability density function that approximates the distribution of many random variables (as the proportion of outcomes of a particular sort in a large number of independent repetitions of an experiment in which the probabilities remain constant from trial to trial). Typically needs at least 30 items. It has the form of

$$est\, f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-x^2}{2\sigma^2}}$$

Subtract *m* (the mean) to x when the mean is not centered on 0. The probability that *m* units will happen given *N* steps.

$$P(m) = \sqrt{2\pi Npq} e^{-\frac{[m-N(p-q)]^2}{8Npq}}$$

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}}$$

see *distribution*

estimating the distribution When working with a continuous sampling of data known to follow a normal distribution (a Gaussian distribution), only the minimum and maximum values need to be known in order to calculate the current distribution:

$$\langle m \rangle = \tfrac{1}{2}(m_{\max} + m_{\min})$$

$$\sigma = \tfrac{1}{2 \cdot 2.58}(m_{\max} + m_{\min})$$

Random walk

$$W(n1) = \sqrt{\frac{|B_2|}{2\pi}} e^{-\frac{1}{2}|B_2|(n_1 - Np)^2}$$

N = number of steps total

p = Number of steps left

q = Number of steps right

$$B_2 = -(Npq)^{-1}$$

normal orthogonal function Un(x) over the interval $-\pi < x \leq \pi$

$$u_n(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} & n = 0 \\ \frac{\cos nx}{\sqrt{\pi}} & n = 1,2,3... \\ \frac{\sin|n|x}{\sqrt{\pi}} & n = -1,-2,-3,.. \end{cases}$$

$$\int_{-\pi}^{\pi} dx\, u_n(x) u_m(x) = \delta_{nm}$$

Can approximate any function with Fourier series as:

$$f_n(x) = \sum_{n=-N}^{N} \alpha_n u_n(x) \approx f(x) = \sum_{n=-\infty}^{\infty} \alpha_n u_n(x)$$

Where in the former case (With some finite number of terms) then the error can expressed as mean-squared error:

$$E_N = \int_{-\pi}^{\pi} dx \left[f(x) - f_N(x) \right]^2$$

numerical algorithms

stable low error; may use permutation on matrix to keep basic algorithm stable.

odds

Prob/(1-prob)

log-odds

$$= \log\left(\frac{p}{1-p}\right)$$

$$quantile(x) = \alpha\left(\frac{p}{1-p}\right)^{1/\beta}$$

$$CDF(p) = \left[1 + \left(\frac{x}{\alpha}\right)^{-\beta}\right]^{-1}$$

ordinary differential equations (ODE)

ODE's are differential equations in terms of single variable. The solvers vary in their speed and accuracy.

- Only one boundary of the interval has a specified value. Runge-Kutta (order 4) is generically the most useful. Bulirsh-Stoer. For stiff problems Burlirsh-Store or Rosenbrock
- Method of weight residuals obtains an approximate solution, using a "trial function'
- With both boundaries specified, methods vary. The simplest is the shooting method.

see also *derivative, differential equation, partial differential equations*

first order

Typically first ordinary differential equations are in a *Cauchy* or *Standard form:*

$$\frac{d}{dx} y = f(y(x), x)$$

(Approximate solution via method of residuals) This function on the right is a closed formula whose only independent variables are y and x. The equation is rewritten:

$$\frac{d}{dx} y + pg = g$$

$$p = 0 \quad y = \int dx\, g \qquad \alpha = \int dx\, p$$

$$g = 0 \quad y = \frac{k}{\mu} or\, 0 \qquad \mu = e^{\alpha}$$

$$g \neq 0 \quad y = \tfrac{1}{\mu} \int dx\, \mu g \quad \mu = e^{\int dx p}$$

second order

$$\frac{d^2}{dx^2}ay + \frac{d}{dx}by + cy = 0$$

$$\left(\frac{d}{dx} - r_1\right)\left(\frac{d}{dx} - r_2\right)y = 0$$

$$r_1, r_2 = \frac{-b \pm \sqrt{b^2 - 4ca}}{2a}$$

$b^2 - 4ac > 0 \quad y_c = Ae^{r_1 x} + Be^{r_2 x}$

$b^2 - 4ac = 0 \quad y_c = (A + Bx)e^{-bx/2a}$

$b^2 - 4ac < 0 \quad y_c = e^{\alpha x}\left(Ae^{\beta i x} + Be^{-\beta i x}\right)$

$y_c = e^{-bx/2a}\left(A\cos i\beta x + B\sin i\beta x\right)$

$$\alpha = -\tfrac{1}{2}b$$

$$\beta i = \frac{\sqrt{4ac - b^2}}{2a}$$

$$\left(\frac{d}{dx} - r_1\right)\left(\frac{d}{dx} - r_2\right)y = g$$

$$let\left(\frac{d}{dx} - r_1\right)\left(\frac{d}{dx} - r_2\right)y_c = 0$$

$$y_c = AU(x) + BV(x)$$

$$y = U_1 U + V_1 V$$

$$U_1 = -\int dx\, Vg \Big/ \left(U\frac{dV}{dx} - V\frac{dU}{dx}\right)$$

$$V_1 = -\int dx\, Ug \Big/ \left(U\frac{dV}{dx} - V\frac{dU}{dx}\right)$$

optimization two types of search: random search, and deterministic

random search direct, Monte Carlo, and Simplex

order
partial A binary relation ≤ on set S is a partial order if (and only if) S is reflex, S is anti-symmetrical, and S is transitive.

strong A binary relation that is asymmetric, irreflexive, and transitive.

weak A binary relation that is transitive, reflexive and anti-symmetric

orthogonal functions trigonometric, fourier, legendre, hermite

oscillator see *engineering glossary*

over determined system A system with more constraints than variables – each equation is a constraint.

parenthesis ordering Affects the cost of multiplication polynomials (e.g. use Horner's method) and matrices. Even more valuable with non-commutative algebra's since they cannot be reordered.

partial differential equations methods include the grid method and method of residues.

see also *ordinary differential equation*

Pearson Product Moment Correlation Correlation is a measure of how two variables affect each other on a standardized scale. It is calculated as a scaled (normalized) version of covariance

$$Cov(x,y) \Big/ \sqrt{\text{var}(x)\,\text{var}(y)}$$

"The standard measure of correlation is called Pearson's product moment correlation coefficient, or, for short, simply the correlation coefficient, symbolized as r. The correlation coefficient ranges from +1 for perfect positive correlation, to 0 for no correlation, to –1 for negative correlation.

Stephen Gould, The Mismeasure of Man*, p241-242*

"In rough terms, r measures the shape of the ellipse of plotted points. Very skinny ellipses represent high correlations – the skinniest of all, a straight line, reflects and r of 1.0. Fat ellipses represent lower correlations, and the fattest of all, a circle, reflects zero correlation (increase in one measure permits not prediction about whether the other will increase, decrease, or remain the same).

"The correlation coefficient, though easily calculated, has been plagued by errors of interpretation…

"Simplification. I began with two dimensions, which I have now, effectively, reduced to one. Since the correlation is so strong, we may say that the line itself (a single dimension) represents nearly all the information originally supplied two dimensions. Secondly, I can… make a reasonable inference about the *cause* of this reduction to one dimension... tightly correlated because they are both partial measures of an underlying phenomenon."

percentile score Percentage of test takers who score below the given score

permutation Different orderings of the elements. The number of permutations are:

$$P_r^n = \frac{n!}{(n-r)!}$$

Where
r is the number of objects used in the permutations
n is the set of possible objects that might be selected for use in the permutations

see also *combination, counting*

perturbation analysis A technique that approximates a system to some similar system whose behaviour we can predict.

Given an insoluble set of equations. Treat this as a version of a well-understood situation with a small "perturbation" that complicates the system.

Φ
golden mean

$$\phi = \tfrac{1}{2}\left(1+\sqrt{5}\right)$$

planar embedding Puts a graph on a 2D plane without overlap of lines (or nearly so).

Display layout uses different methods: wiring for PCB or IC's is one use, where the wires often have constraints and there are costs for each 2-D plane and transitions between them. Things that look like circles and arrows are aesthetically laid out using a spring embedder.

Poisson distribution

$$P(x) = \frac{\lambda^x}{x!}e^{-\lambda}$$

$$P(servicetime > y) = e^{-\mu\lambda}$$

where λ is the expected number of occurrences (or events) in the observation period. P(x) is the probability of the event at x (?)

see also *Bayesian model, Markov model, trends (predicting)*

Poisson's equation

$$\frac{\partial^2}{\partial x^2}T(x,y)+\frac{\partial^2}{\partial y^2}T(x,y) = -f(x,y)$$

$$\nabla^2 U = f(x,y,z)$$

see also *Helmhotz equation*

Poisson model The Poisson model is a Markov model that assumes that the probability of an arrival is independent of the size of the population. It is, effectively, a variation of the Poisson distribution The probability of having *m* arrivals in time *t* is:

$$P(m,t) = P(m)_{\mu=\lambda t} = \frac{(\lambda t)^m}{m!}e^{-\lambda t}$$

where λ is the arrival rate (number of units/time unit),

u is the service rate

Probability of a delay in arrivals between time a and time b: $e^{-\lambda a} - e^{-\lambda b}$

Number of items in all queues $= \frac{\lambda}{\mu - \lambda} = \frac{utilization}{1 - utilization}$

utilization per server with many servers = system utilization/num servers

Per server Queue length $= \frac{U^2}{n^2} \frac{n}{n-U}$ (U = system utilization). This approximation is not quite correct, but close

see also *Bayesian model, Markov model, Poisson distribution, trends (predicting)*

polygon area

The cross-product method.[4] The vertices are ordered along one path. The signed are is computed as 1/2 the sum of the cross-products (determinants in two dimensions):

$$= \frac{1}{2} \sum_{i}^{n-1} \begin{vmatrix} x_i & x_{i+1} \\ y_i & y_{i+1} \end{vmatrix} + \frac{1}{2} \begin{vmatrix} x_n & x_1 \\ y_n & y_1 \end{vmatrix}$$

$$= \frac{1}{2} \sum_{i}^{n-1} (x_i y_{i+1} - x_{i+1} y_i) + \frac{1}{2} (x_n y_1 - x_1 y_n)$$

If the area is negative, the traversal was clockwise; otherwise the traversal was counterclockwise

polynomial evaluation

The polynomial evaluation depends on the form of the polynomial. In some cases this is an estimation method. Refer to the table below:

Table 16: **Polynomial evaluation**

Form	*Evaluation method*
$c_n x^n + c_{n-1} x^{n-1} + \ldots + c_0$	Horner method. Evaluate for v and -v
Horners form $c_0 + x(c_1 + x(c_2 \cdots))$	Evaluation for v and 0v
Chebyshev polynomial	Chebyshev algorithm
Newton polynomial	Newton's method

monic

$x^n + c_{n-1} x^{n-1} + \ldots + c_0$

minima, maxima

The minima and maxim occur when d/dx P(x) = 0, and possibly at boundaries.

roots

The roots occur when P(x)=0. See the table below for possible techniques based on form.

Table 17: *Polynomial roots*

Form	*Root solving method*
$ax^2 + bx + c$	Quadratic method: $x_{1,2} = 1/2a \left(-b \pm \sqrt{b^2 - 4ac} \right)$

[4] Flint, Abel 1808. *A System of Geometry and Trigonometry: together with a treatise on surveying: teaching various ways of taking the survey of a field; also to protract the same and find the area. Likewise, Rectangular Surveying; Or, an accurate method of calculating the area of any field arithmetically, without the necessity of plotting it.* Second edition. Hartford: Oliver D. Cooke. (See pp. 59–94

Bart Braden, 1986. "The surveyor's area formula" *The college Mathematics Journal* 17(4):326-337. Proved the technique

> degrees	use Laguerre's method to find all of the complex roots.
...	try a generic root finder

positional notation
1941

A system of expressing numbers in which the digits are arranged in succession; the position of each digit has a placed value and the number is equal to the sum of the products of each digit by its place value.

power
raising to

See *exponentiation*

prediction

see *coefficient of determination, forecasting, trends (predicting)*.

present value
net

Attempts to identify the minimum amount of money we should have now (v_0) to be better off than having a given amount, v_t, in the future.

$$v_0 = v_t(1+r)^{-t}$$

where
v_0 = net present value
v_t = value at time t
r = discount rate

primes

Babbage primes, Mersenne primes, Twin primes.

principal component analysis

Used to find a set of components that are uncorrelated.

PCA transforms a set of variables into uncorrelated variables.

Original: $x_1, x_2, \ldots, x_n$

Result: $z_1, z_2, \ldots, z_n$

$$z_i = \sum_{j=1}^{p} a_{ij} x_j$$

properties of the transform

Most information is in Z_1 least is in Z_n, high variance = high information.

$$\text{var}[z_1] > \ldots > \text{var}[z_n]$$

No correlation between any variables:

$$Cov[z_i, z_j] = 0, i \neq j$$

Total variance is consumed:

$$\sum \text{var}[x_i] = \sum \text{var}[z_i]$$

probability

Phenomenon are random when taken in isolation (perhaps thru partial knowledge), but can make predictions based on aggregate observations. Probability estimation into a number of estimation steps.

1. Probabilities inferred from measurement. Simple regression to counts. Include shaped by a distribution known by the phenomenon, theory, system, etc. filter out variance and systematic bias. To determine the probability from the empirical data see *likelihood estimator*. A probability distribution is used to model likely *but unobserved* events. Such a distribution has three forms of functions: a probability density function, a probability distribution function, and a function whose values are randomly distribution according to the distribution. The type of distribution is selected based upon the structure of the problem. The distributions parameters are estimated or calibrated based upon the *knowns* of the problem context.
2. Probability (or structure) known a-priori by the nature of the expression

see also *distribution, likelihood estimation*

a priori theory

Probability is a measure of how rational it is to believe that event e will happen, relative to what is known. No trials need be run before the probability is assigned. TO compute probability of an events occurring in given circumstances is to divide the number of ways in which it can occur by the total number of possible outcomes of those circumstances, provided there is no reason to

believe that anyone of those possible outcome is more likely than any other. No event has any *intrinsic* probability. It can be assigned a probability only on the basis of the evidence available to the person making the assignment.

relative frequency theory

Probability is a measure of how often members of an observed group have an observed feature.

sample space

A set of all possible mutually exclusive outcomes. Given any sample space (uniform or not), and the probabilities of the associated with the points, the probability of an event is found by adding the probabilities associated with all sample points favorable to the event.

method of counting

Fundamental principle: If one thing can be done N1 ways, and after that a second thing can be done in N2 ways, the two things can be done in succession in that order in N1*N2 ways. This can be extended to doing any number of things one after the other, so that the product is N1*N2*N3…

models

Probability models *emulate* – they do not simulate the innards. Prior probabilities arise from prior knowledge or inference. Conditional probability reflects mechanisms and how they work.

of alternative non exclusive events

Try to reformate as exclusive events, or calculate the probability of failure, then subtract it from one.

conditional probability

Conditional probability provides the key definitions for chaining inference. Forward chaining is accomplished with the Chapman-Kolmorogorov theorem; Backward-chaining uses Bayes' theorem. This probability is notated as:

$P(A|s)=$ Probability of A, if s is true, i.e. $P(s \to A)$

This is computed using some *prior* probabilities.

$P(s|A)=$ Probability of s, if A is true, i.e. $P(A \to s)$

$P(A)=$ Probability of A occurring

$P(s)=$ Probability of S occurring

Scoring methods can be used if the scores are chosen carefully – e.g. the log of the probability. These behave the same as mutually-exclusive conditional probability networks.

Table 18: **Forms of conditional probability chains**

Discrete domain	*Continuous domain*
$P(B) = \mathbf{P}^{\#variables\text{-}1} P_{priors}$	$P(A\|B) = \int_M dM \int_\theta d\theta\, P(A\|\theta) P(\theta\|M)$
	$P(A\|B) = \int_\theta d\theta\, P(A\|\theta) P(\theta\|B)$

Chapman-Kolmogorov theorem

A *matrix* can be using (easing work); each state corresponds to a variable binding:

State 1: var1 = val1
State 2: var1 = val2
…

The subsets of the matrix, the rows corresponding to single variables, form a *stochastic matrix.* The normal form of a conditional probability matrix has the source node as columns, and the destination nodes as rows. A variable as multiple columns – one for each possible value. With the given variables,

$$P(Var_1 = val|x) = \begin{cases} 1 & x : Var_1 = val \\ 0 & otherwise \end{cases}$$

In the conditional area,

$$P(Var_i = val|Var_i = val) = 0$$

The conditional probability for every variable value can be found by using a matrix of conditional probabilities (P) that acts as transform:

$$P(\text{var} = value) = \mathbf{P}^{\#variables\text{-}1} P_{priors}$$

$$\mathbf{P}^n \equiv \mathbf{P}^{\#variables\text{-}1} \quad n \geq \#variables - 1$$

see also *Bayes probability, matrix (stochastic)*

Table 19: *Normal form of a conditional probability matrix (with var_1 as the given)*

			Given: Var_1		Given: Var_n	
			false	true	val_1	… val_m
To: Var_1	false		1	0	0	0
	true		0	1	0	0
…						
	val_1		$P(\neg Var_1 \mid \neg Var_n)$		0	0
Var_1	val_1				0	0

of a sequence

$P(e_0 \ldots e_n)$ does not need the whole sequence; usually only window into the last few items (eg 3) is need to get a reasonable estimate. Bayes can be used where a given is a sequence (a substring, or a set of possibly unordered characters in the string), a set (i.e. orderless) of items in the sequence.

See *likelihood estimation*

converting from rate

$$p = 1 - e^{-rt}$$

where
r = Rate = events / population time

see also *transition probability*

Table 20: *Various functions described by probability distributions*

Form	*0*	*Description*
$E(x)$		Expected value
$f(x)$	Mean	This is a probability density function. The formulation of probabilities is in terms of density functions; seemingly counter-intuitive itis useful.
$F(x) = \int_{-\infty}^{x} du f(u)$	Mean	The distribution function, the cumulative probability
$P(x) = f(x)$ $P(x \le b) = F(b)$ $P(a < x \le b) = F(b) - F(a)$ $= \int_a^b dx f(x)$	Mean	An estimate of the probability of x, or of x being true. In the unqualified form, it becomes the probability density function. If the mean of x is not at zero, use x-m.
$W(x)$	Mean	This is a probability density function, except that it has been shift so that the mean is x=0.
$Z(z)$	Mean	Z is the probability function, except that z is units of standard-deviations, and z=0 is the mean. (Z(0) is the probability of the mean, Z(1) is the probability x is 1 stddev to the right, Z(-2) is the probability when x is 2 stddevs to the left of the mean.

Table 21: *Probability equations*

Given units	*Evaluation method*
arrival, time	Markov linear birth process

	Poisson model Queue model
events, time	Probability Completion time (see *Beta distribution)*
items	Permutation
probability	Bayes theorem
steps, time	Gaussian distribution
transitions, time	Transition probability

Table 22: *Probability estimators*

Form	*Estimation method*
P(a)	Independent probability
P(a|b)	Markov model
$P(a\|c_1..c_n)$	Smoothed or Viterbi

Table 23: *Probability distribution for a context*[5]

Context	*Distribution*
film's total gross	? **Calibration:** how much it has earned so far. (How long it has been out is not necessary)
length of stem people's height eye spacing	Normal distribution **Calibration:** ???
length of a ruler's reign (congressman's term length, pharoah's lifespan)	Erlang distribution **Calibration:** how long it has been in the oven so far.
length of time to bake a cake	Complex and irregular distribution **Calibration:** how long it has been in the oven so far.
length of a telephone-queue	Power-law (it was thought to be a Poisson distribution)
length of a word	Poisson distribution
Number of unique words in a text	Hank's law
Popularity index (frequency of use)	Zipf's law

Table 24: *Evaluation of First-Order Logic*

Form	*Dependent*	*Independent*	*Mutually Exclusive*
$P(A \vee B)$	$P(A)+P(B)-P(A \wedge B)$	$P(A)+P(B)-P(A)P(B)$	$P(A)+P(B)$
$P(A \wedge B)$	$P(A \mid B)P(B)$	$P(A)P(B)$[6]	0
$P(A \mid B)$	$P(A \wedge B)P(B)$	$P(B)$	0

problem Some problems can be assigned a number (on what scale?) indicating how amendable it is to

[5] Tom Griffiths (Brown University), Joshua Tenenbaum (Mit) in *Psychological Science* (2006), asked how bayesian people are

[6] One source suggested: $\frac{P(A)P(B)}{P(A)P(B)+P(\neg A)P(\neg B)}$

ill-conditioned, condition number — digital computation.

well-posed — Implies that the problem can be translated into a computable, stable algorithm; problems that are not well-posed need to be reformulated into such a form. Well-formed problems have single solution, and it depends continuously on the input.

programming mathematical programming — Finding a best from a set of alternatives. Each is given an objective function, to find which is best, and constraints

Technique	*Out*	*In*
Linear program	vector	Domain of variables
Integration solver	Vector sequence	Differential equations
search	Vector	Domain of values
Simulation	Sequence	

Table 25: *Mathematical programming techniques*

approximations — Linear approximation, multi-variable linear approximation, root-finders, function minimizers, SAT, MaxSAT, LazySAT.

Computer algebra: Discrete calculus, recurrent, term rewriting, generating functions

Automated Reasoning: paramodulation, binary hyper-resolution, equality reduction, conflict analysis, modal logic.

process Bernoulli — Two states, and the transition matrix is not dependent on time or sequence

deterministic — The processes current state is determined by its properties (e.g. constants and model formulae), previous states (including differential equations), and its inputs.

see also *ordinary differential equation, partial differential equation.*

stochastic — Similar to a deterministic process, except there is a random element as well – possibly errors in the model, possibly other (unknown) inputs).

see also *Markov model, Poisson model*

proof constructive mathematics — Holds that to prove the existence of a mathematical object it is necessary to show how to construct it, atleast in principle. The procedure must end after a finite number of steps. At the end of each step, it must be evident how to proceed to the next step.

inductive method — See *induction*

pure-existence proof — Demonstration of existence without construction. Must exist, otherwise there is a contradiction.

quantile — Points at regular intervals of a cumulative distribution function.

Quantile(.9) == the value of X that is 90% of the range

quaternion — Very similar to a vector, although it did not become the preferred formalism in physics.

queue — Queues are noted by the following parameters

λ Arrival rate, items/time
μ Service rate, items/time

$\frac{\lambda}{\mu}$ utilization rate

$\frac{\lambda}{\mu}$ probability that a service point is in use

Lq Average length per server, number of items in the queue

Lq average length for whole system

W Average wait time per item

see *Markov model, Poisson model*, and *trends (estimating)*

controllable — A single channel queue is controllable iff $\lambda < \mu$

estimating — To estimate the amount of space required to hold a queue, use two Poisson models: one to estimate the queuing rate, and another to estimate the dequeuing rate.

see *Markov model, Poisson model*

phalanx — Where each 'server' has its own separate queue, rather than several servers serving a single shared queue.

radix complement — A method of subtracting two numbers by converting one to its complement (akin to its negative) and adding it to the other. The complement can be formed by subtracting each digit from one less than its radix, then adding one to the least significant digit, executing carries as required.

notation — A system for representing numbers using a positional representation, each digit has a significance based on its position. Usually this value is (Value of digit)*Radix**(Digits to left of radixpoint-1) Each digit may range in value from zero to one less than the radix.

point — The point that separates the digits in the integer part of a number from the digits in the fractional part. This may be an implied point; it may be a special character such as a period or a comma.

Ramsey theory — Branch of combinatorics that studies how big a set has to be before a rule is guaranteed to be true. How many people must be at a party to ensure that atleast 5 people know each other, or atleast 5 are strangers?

random number — Linear congrugation method of generation: X[n+1] = k*x[n]+ c%n
Logistic formula method of generation: X[n+1] = r*X[n]*(1-X[n]) ; y[n]=(X[n]-a)/(b-a)

knuth's statistical tests for randomness — chi2 test for correlation channels

recurrence — How fast recurrence functions call themselves. If you know the frequency, you can set the buffer size.

recursive function primitive — The parameters to the call to itself can be the original parameters, possibly multiplied by a constant, and possibly added to a constant.

regression — Takes careful measurements of the independent variable and dependent variables. Uses this samples to create a function that estimates the dependent variables from a given independent variable.

see also *exponential regression, function minimization, linear regression, logarithmic regression, power regression*

Type	*Estimator function*	*estimate inverse*
exponential	$f(x) = be^{mx}$	$x = \frac{1}{m}(\ln y - \ln b)$
linear	$f(x) = mx + b$	$x = \frac{y-b}{m}$
logarithmic	$f(x) = b + m\ln x$	$x = e^{(y-b)/m}$
power	$f(x) = bx^m$	$x = e^{(\ln y - \ln b)/m}$

Table 26: *Regression functions*

exponential regression

Approximates variable data as a constant raised by the independent variable(s):

method: define $y' = \ln y$
feed x,y' into linear regression, calling the results m and b' and translate them as:
$b = e^{b'}$

linear regression

Approximates variable data as a scaled form of the independent variable or a function of the independent variable(s).

method: $m = \text{cov}(y,x)/\text{var}(x)$
$b = avg(y) - m \cdot Avg(y)$

significance: $r^2 = correlation(y,x)^2$
r^2 measures how much of the change in y comes from x;
multiply r^2 by 100 to get the percentage change in y from x.

To compute y in terms of a function f(x), substitute "f(x)" for x in the above relationships

logarithmic regression

Approximates variable data as a logarithm of the independent variable(s):

method: define $x' = \ln x$
feed x',y into linear regression

power regression

Approximates variable data as an exponent of the independent variable(s):

method: define $x' = \ln x$
feed x',y into linear regression

relation

Table 27: *Relation*

anti-symmetric	$\forall_{a \in domain} \langle a,b \rangle \in R \wedge \langle b,a \rangle \in R \rightarrow a = b$
equivalence	R is reflexive, symmetric, and transitive
partial order of domain	R is reflexive, anti-symmetric and transitive
reflexive	$\forall_{a \in domain} \langle a,a \rangle \in R$
symmetric	$\forall_{a \in domain} \langle a,b \rangle \in R \rightarrow \langle b,a \rangle \in R$
total order	A partial ordering, where a<b,a=b,or a>b for any 2 elements in domain
transitive	$\forall_{a \in domain} \langle a,b \rangle \in R \wedge \langle b,c \rangle \in R \rightarrow \langle a,c \rangle \in R$

relationship strength

Coefficient of correlation
Coefficient of determination
Standard error

residual

What is left – unexplained – by analysis; typically the difference between the predicted and actual value. After a complete analysis, the residual should me minute and random. If there is a pattern remaining in the residual, the analysis is not complete (no further analysis is possible). A graph is a good way to check the residual.

See also *regression*

ring lattice

a completely structured regular graph, parameterized by degree K. Each node is connected to K neighbors (K/2 ona side)

risk

P probability an individual will experience an event. Do not confuse with rate, which is the number of events per population time. Portfolio approach to classifying risk:

- A broad risk associated with the population – market risk of owning a stock, for

example, which is measured by looking at the distribution description for the population.

- A specific risk associated with the specific individual. The stock's volatility, for example, as measured by its variability relative to the population as a whole
- Investment risk

Characterize against a benchmark – R^2, Beta against market indices.

Consequences of risk (loss). Evaluating likelihoods of events, and the severity out outcome; combining the two into a level of concern.

see *safety index, Weibull power law*

root finder (f) A root finder finds a value to pass to *f* that will be:

$$f(x) = 0$$

A root finder can, of course, find a value *x* to pass to such that it returns any value (c) you specify:

$$f(x) - c = 0$$

This is useful for calibrating scientific instruments. It can also be used to emulate functions that we do not know how to compute (as easily):

- $f(x)$ is inverse of the function we are trying to emulate
- c is the input value to the object function (the one we are trying to emulate)
- x is the computed result

For a cube root as an example:

$$func(input) = \sqrt[3]{input}$$
$$f(x) = x^3$$
$$f(x) - input = 0$$

Bairstow method An adaptation of Newton's method.

Laguerre's method A technique for finding the roots of a polynomial, including any complex roots.

Speed: Faster than Newton's method. Like other methods, it employs an initial guess for a root, and the time to compute the roots depends upon the quality of the guess.

Requires: First and second derivative of the polynomial

Newton's method Newton's method is a simple gradient root finder; it converges quickly if converges it all. To illustrate the method, we'll use cube root as an example: $func(input) = \sqrt[3]{input}$.

Newton's method needs four elements:

- x_0: An initial guess of where the root (the solution) might be. This might be the input value / 3.
- Symbolically rearrange func() into its inverse; $f(x) = x^3$
- An error function $Err(x) = f(x) - input$. In our example this is: $Err(x) = x^3 - input$
- y(x): An optional derivative function of f(x); for cube root this is $y(x) = 3x^2$
- A simple Taylor expansion to approximate $f(x)$: $^n g(x) = f(x_n) + (x - x_n)y_n$ This approximation forms a straight line; the overall method is used to find $g(x) = 0$,

For a function of a single variable, Newton's method then employs the following steps (with n initially zero):

1. Stop iterating if $|Err(x_n)| < \varepsilon$
2. If a derivative function was supplied, compute the value at x_n; otherwise calculate the derivative at x_n using a finite different method. Call this value y_n.
3. The next guess for x is:

$$x_{n+1} = x_n - \frac{Err(x_n)}{y_n}$$

4. Continue iterating with n++

To find the root of several simultaneous equations, the test step one is modified to test all of the functions (each must be less than a tolerance threshold), and the generation of a new x_{n+1} in step 3 takes the smallest non-zero x:

$$|Err_1(x_n)| < \varepsilon$$
$$|Err_2(x_n)| < \varepsilon$$
$$:$$

With multiple independent variables, steps 2,3,4 are modified:

1. Stop iterating if $|Err(x_n)| < \varepsilon$
2. Calculate the derivatives of Err(x,y) at x_n,y_n using a finite different method. Call the finite difference d/dx c_n, and d/dy d_n.
3. Construction a simple Taylor expansion to approximate $f(x,y)$:

$$^n g_1(x,y) = f_1(x_n,y_n) + (x - x_n)c_n + (y - y_n)d_n$$
$$^n g_2(x,y) = f_2(x_n,y_n) + (x - x_n)c_n + (y - y_n)d_n$$

Powell's method — Modifies Newton's method – it continues to step 4 if x_{n+1} is better than x_n in some way.

stochastic approximation

S-plus — Modes: Logical (Boolean), numeric, complex, character

Data Types:

- Atomic: Vector, Matrix, Array, Time Series, Factor, Ordered Factor
- Recursive: List, data frame, model formula, model design, model terms, ANOVA Model, ANOVE table, Linear model, Generalized Linear Model, Generalized Additive Model, Minimum-Sum non-linear model, Least Squares non-linear model, Seasonal Time-Series Decomposition, Tree-based model, Logical Regression.

sampling
reservoir — Can find N random examples form a set of unknown size.

representative — One that has been chosen in such a way that the personal and social characteristics of the sample included are like those in the entire population in all important respects relevant to the study.

sequential random

small sample sets — With small samples – 30 items or less – use non-parametric methods of Student's t distribution or a Binomial distribution (if the items are bivalent)

techniques — The following techniques are used to sample a population faster and with less work – but with less accuracy than other techniques:

- Cluster sampling
- Multistage sampling
- Quote sampling
- Stratified sampling

satisfaction procedure

A method that solve a system of equations by finding values for variables. Root finders and symbolic techniques can provide exact answers. Approximate with linear regression, simplex method, function minimizers, and so forth.

ODE's can be solved by Runge-Kutta methods – see solver

A,theta in sin, cose, e-I can be solved with an FFT

S,theta in Affine Transforms can be solved with FFT

Certain algebra with EigenVectors and backsolving.

scalar value
types of

Independent of dimensions, units:

- Nominal or categorical values – e.g., male or female, red green or blue, etc. With nominal values, one can only find the most popular (the mode), not a median or mean value.
- Ordinal or ranked values – for example, poor, good, or great. With ranked values, one can find the mode or median, but not a mean
- Interval values – these are useful when the difference between two values is meaningful, although the absolute value isn't guaranteed to be. Temperature and Year are examples. The difference can have a mode, median and mean. The absolute value has a mode, a median, and occasionally a mean. You can not say one temperature is twice as hot as another – the ratio changes with units; nor can you say one year is three times another year.
- Ratio scaled data is like interval values (it has a mean, mode, and a median), except that it has a meaning when compared with others. For example, a bowling ball can be twice the mass of another ball.

sequential probability ratio test

Turing's problem: were two strings encoded by the same rotor settings? E.g. U(x,y) and V(a,b) – we know x and a, are U and V the same? He used two scores for each character:

If xi==ai, then returned c0 (usually positive)

If xi!=ai then return c1 (usually negative)

Would then go thru string summing each character's score. If the total score was > threshold, then it would conclude it was the same rotor settings. As Bayes scores, these are:

$c_0 = \log\left[\frac{1}{13} / \left(\frac{1}{26}\right)\right]$ the log ratio of probability they are the same if the same rotor, to the probability that they would be same with different rotors.

$c_1 = \log\left[\frac{12}{13} / \left(\frac{25}{26}\right)\right]$ ratio of probabilities that they are different.

genets

A similar problem for Gene alignment, but solved using Levenshein distances and finding the candidate alignment points.

Find alignment point of several chromosomes

A few nucleotieds will not match

Some have '-'s inserted

There can be lots of "junk" before the alignment point

Karlin-Altschul theory – count # of matches in a region.

Significance $= 1 - e^{-Kmne^{-ks}}$ k, lambda,K-A parameters

M,n length of the two sequences.

series
geometric

$x^0, x^1, \ldots x^n$

Sums to: $\frac{1-x^n}{1-x}$

accelerator — The Aitken method of speeding up the rate of convergence of a series.

sets
cardinality — The count of items in a set. If the set is not finite, it is the aleph of the set.

equal — all the same members in both sets A and B.

equivalent — The sets have the same number of members, and there is a 1:! correspondence between the elements.

equivalence relation — a relation on the members of a set, if it is reflexive, symmetric, and transitive.

shortest path — See also *A*, APSP, Min Spanning Tree, Topological sort*

Performance in worst case is $O(n^2)$ or worse.

- Graph is densely connected
- Next() is *slow* (usually) – e.g. O(n) or worse
- Satisfaction may or may not be slow

Performance can be much better when:

- Graph is very sparsely connected
- Next() is implemented to be O(1)
- Satisfaction is fast.

for purchase scheduling — Shortest path method for purchase scheduling.

Add to the graph, for all possibilities DateBought-> DateSold, annotated with TCO.

Find the shortest path. This will indicate when to purchase and to sell

Weakness, not able to handle overlapping dates.

Various value & meaning, convert to cost of ownership form: yearn-1 -> yearn, annotated with cost of ownership.

SIFT
scale invariant feature transform

simplex method — Nonlinear (Nelder-Mead). A non-derivative method. Employs fewer evaluations than other function minimization methods. Developed by Danzig. See also L.G Kachian's methods.

The steps in Nelder-Mead simplex are: reflect, expand, contract outside, contract inside, shrink.

JA Nelder & R Mead, "A Simplex Method for Function Minimization," 1965 Computer Journal 7, 308-313 .
Jeffrey C. Lagarias, James A Reeds, Margaret H Wright, Paul E Wright, "Convergence Properties of the Nelder-Mead Simplex in Low dimensions" May 1, 1997 SIAM Journal of Optimization

simplification — see also *factor analysis*, and the use of parameterized *distributions*

simulation

- Model,
- Initial conditions,
- Element set and their states,
- Time sequence,
- Output states

standard capabilities — A simulation tool tends to have the following:

1. "Structured data input.
2. "A predetermined time-flow mechanism," classified either as "uniform or

Operations Research: Principles and Practice *2ed, A Ravindram, Don Phillips, Jams Solberg, 1987 John Wiley & Sons,*

variable time flow."

3. "Echo checks and/or error checks for program input and logic structure.
4. "Random-number generation routine(s)…
5. "A variety (capability) of random deviate generators.
6. Most essentially "a *clock routine* that automatically stores, sequences, and chronologically selects simulated events through time and maintains model equilibrium.
7. An event mechanism, primarily classified as either "events discrete in time continuous through time."
8. "Automatic statistical collection (generation) functions.
9. "Standard simulation output of relevant data and simulation statistic..
10. "Proper documentation and instructions" of how the system works and how to employ it properly.

simulink process

MATLAB's Simulink process:

1. Initialization. Sets the number and dimensions of inputs and outputs. Sets the sample time, etc.
2. Calculates when the module should be called next, based on the variable time step.
3. Calculate the outputs in the major time steps; when done, the outputs are valid.
4. Update discrete steps, perform once-per-timestep actions
5. Integration. Continuous models have their output and derivatives called at minor time steps to allow computing the continuous steate:
    a. Calculate derivatives
    b. Calculate outputs
    c. Calculate derivatives
    d. Locate zero crossings
6. repeat from 2

A module has a name, current time, state vector, and inputs. Output can be the integrated values, or a function defined on the integrated value. The value integrated is stored elsewhere and passed as a parameter. Parameters passed include:

- Current time
- State vector being integrated
- Input vectors passed at start – to represent instance variables.

Table 28: Suggestions for improving simulation credibility

Table is quoted from Todd R Andle, Alec Yasinac "On the Credibility of Manet Simulations" IEEE Computer, *July 2006, p48-54*

Problem	*Solution*
Lack of independent repeatability	Properly document all settings. Publication venues have limited space, so typically include only major settings. Provide all settings as external references to research web pages, which should include freely available code / models and applicable data sets.
Lack of statistical validity	Determine the number of required independent runs. Address sources of randomness (such as pseudorandom number generators) to ensure simulation run independence. Collect data only after deleting transient values or eliminating it be preloading .. tables and traffic queues.
Improper / non-	Validate the complete simulation (developed protocol, traffic, radio model, and scenario) against a real-world

existent validation	implementation. When this isn't possible (such as during early concept development), validate the simulation against analytical models or protocol specifications. The latter will be less precise, but you can further refine it as implementations are realized.
Unrealistic application traffic	.. base traffic generation on intended applications.
Improper precision	Use [this type of] simulations to provide proof of concept and general performance characteristics, not to directly compare multiple protocols against one another.
Lack of sensitivity analysis	Sensitivity analysis can identify a chosen factor's significance (parameter settings that change in a study). ..

singular value decomposition

Used to find an approximate eigenvector for a non-square matrix.

See also *PCA*

skeletal set

essential geometry of scene (points and connection)

small-world model
Watts and Strogatz

N: number of nodes
K: degree of each node at the first step
p: probability of rewiring edge

1. Start with N nodes, connect every node to K neighbors (k/2) on a side.
2. Rewire each edge with probability p. Duplications and links to self are not allowed.

Scale free model
Barabasi & Albert

m0: number of nodes at step 0
m: number of edges at each step.
N: final number of nodes in the graph

1. start with m0 nodes.
2. add a new with m edges (max m0) to the others
3. choose m nodes to connect to the new node, weighted for the nodes with the larger degrees.

smoothing

dampens fast variation (e.g. noise) and corrects certain sample biases.

solver
discrete solver

Model has no state or has a discrete state. A variety of brute force or logical state space searches can be employed.

fixed-step, continuous solver

User specifies the state derivatives. The solver computes the model's state by integration. The integration formulae include:

- Dormand-Prince formula
- RK4: a fourth order Runge-Kutta
- Bogacki-Shampire formula
- Heun's method, aka improved Euler formula
- Euler's method

variable-step, continuous solver

User specifies the state derivatives. The solver computes the model's state by integration, but the time-steps are variable. The integration formulae include:

- Runge-Kutta(4,5) formula, using the Dormand-Prince pair (considered a good default)
- Runge-Kutta(2,3) pair of Bogacki & Shampine
- Adams-Bashforth-Moulton PECE solver

- Numerical differentiation formulas (NDFs)
- Rosenbruck formula, order 2
- Trapezoidal
- TR-BDF2, 2-stage trapezoidal, Runge-Kutta

sort

To segregate items into groups according to some definite rules

spherical symmetry

Many 3D problems have spherical symmetry. In general, differential equations are hard to solve. However, some spherical-symmetry can be formed as special ordinary differential equations with known solution techniques. associated Legendre equation, and associated Legendre polynomials; there are Legendre polynomials for said (and theyre simpler), and Rodrigues formula

spline
b-spline
Bézier spline

Pierre Bézier 1970.

A spline is a smooth curve connecting two points; another two points are used to control the curve. (x1,y1,x4,y4) are the two end points, (x2,y2,x3,y3) are the two control points:

$$x = ax_1 + bx_2 + cx_3 + dx_4$$
$$y = ay_1 + by_2 + cy_3 + dy_4$$

a,b,c,d vary with *t*, which ranges from 0..1. For B-splines, they are defined as:

$$a = -\tfrac{1}{6}t^3 + \tfrac{1}{2}t^2 - \tfrac{1}{2}t + \tfrac{1}{6}$$
$$b = \tfrac{1}{2}t^3 - t^2 + \tfrac{2}{3}$$
$$c = -\tfrac{1}{2}t^3 + \tfrac{1}{2}t^2 + \tfrac{1}{2}t + \tfrac{1}{6}$$
$$d = \tfrac{1}{6}t^3$$

For Bézier splines, they are defined as:

$$a = -t^3 + 3t^2 - 3t + 1$$
$$b = 3t^3 - 6t^2 + 3t$$
$$c = -3t^3 + 3t^2$$
$$d = t^3$$

cubic spline

A cubic spline is a smooth curve connecting two points.

spectrum

Strengths of various frequencies that make up a function

Strengths of various variables that make up a function

stability analysis

Since most methods in control theory can't 'design' a control-loops parameters that likely to be stable, a set stability-checks are need to further tune them. Techniques include:

- Bode representation,
- frequency-response,
- Laplace method,
- Nyquist criterion,
- root-locus method

standard deviation

Standard deviation measures how much a function or set of numbers varies from the mean value. Normalized (to a Gaussian distribution) with a Z-score. There are two forms. When all data from a finite population is considered use the form:

$$\sigma_n = \sqrt{\operatorname{var}(x)}$$

When a subset of data from the population is considered, use the form:

$$\sigma_{n-1} = \sqrt{\tfrac{n}{n-1}\operatorname{var}(x)}$$

standard error The standard deviation of a sample set (since it is an approximate measure of a population and its standard deviation).

stationary point A point at which

$$\frac{d}{dx} f(x) = 0$$

statistics
descriptive Descriptive is needed for inferential

inferential Inferential statistics infers about some unknown on the basis of statistics. This includes Bayesian & frequency probability, fiducial probability, eclectic probability. For a process, estimation of point value, interval; test a hypothesis, predict or for cast.

An application of logic using numerical methods.

One & two-sample problems
Distribution shapes, location, scales
For counts & proportions
Tests for goodness of fit

statistical model Developing a statistical model:

pe S-Plus 2000 Guide to Statistics V1 MathSoft

1. "Determine the variables to observe…
2. "Collect and record the data observations
3. "Study graphics and summaries of the data to:
   i. "Discover and remove mistakes
   ii. "To reveal low-dimensional relations between variables
4. "Choose a model describing the important relationships seen or hypothesized in the data
5. "Fit the model using an appropriate modeling technique
6. "Examine the fit using model summaries and diagnostic plots
7. "Repeat 4-6

statistical tests
for a single hypothesis Tests how likely the event was to happen by change; if likelihood is below a threshold (ie, is unlikely) then the effect is considered to linked to the hypothetical cause. Note: avoid applying this technique to find the 'cause' if you have many hypothesis to test. It will over-estimate the linkage with the hypothesis; one out of 20 hypothesis will be accepted as a fluke).

See also *regression analysis*

multiple hypothesis Testing which, if any, of the candidate hypothesis best describe.

Table 29: **Statistics functions**

Type	*function*
Fermi-Dirac	$f(\varepsilon) = \left(e^{\alpha} e^{\varepsilon / kT} + 1\right)^{-1}$
Bose-Einstein	$f(\varepsilon) = \left(e^{\alpha} e^{\varepsilon / kT} - 1\right)^{-1}$
Maxwell-Boltzmann	$f(\varepsilon) = \left(a e^{\varepsilon / kT}\right)^{-1}$

$\alpha = -\varepsilon_f / kT$

ε_f = fermi energy

Steiner trees Solve wiring and combinatorial problems (approximate)

stereo geometry 3D scene projected onto two different 2D planes

see also *fundamental matrix*

Stirling approximation

See *factorial*

Stochastic matrix

See also *probability (conditional)*

Student's t-distribution

Used with small populations – e.g. 30 or less. (Larger populations are usually best modeled with a normal distribution.

$$t = \frac{\bar{x} - \mu_0}{\sigma_{n-1}/\sqrt{n}}$$

n=sample size
σ_{n-1}= sample stddev; computed with a n-1 not a n
μ_0= required population mean,

x= sample mean computed with a n-1 not a n

N=population size

Correction factor on any computed value is $\sqrt{\frac{N-n}{N-1}}$

support vector machine

See also *kernel method*

system

Overview. State space and notation. Representation in state space; special representations. Special aspects of state. Stables states (eg electron in orbit). Bound states. Stationary States. Constraints imposed by uncertainty. Description of the state of the system. Ordering (and partial orderings) of states.

State dynamics. Evolution: how the system changes with time. Unstable states, lifetime of unstable states. Coupling of states; weak coupling and its effects; vibration & vibration models, vibration states.

entities & measurement

Order of magnitude with entity. Description of physical quantities. Measurement of physical quantities. Observables; scalar observables, vector observables. Observability. Quantitative rules. Measurement process. Mean value. Simple & hard; non-degenerate (simple). Degenerate – several final states can be associated with the same measurement value.

analysis process

Decomposition techniques. Separation of variables (eg along coordinates), spectral decomposition, as a sum of functions.

Parameterization (|k,l,m> ⇔ Psi k,l,m (R)).

Classify problem – then use right process to solve. Split problem into a set of sub problems that have been classified.

Matching conditions. When a problem is split into different forms for each section, they still must join or match values at the split points akin to boundary value.

Identify dynamic variables of system

Generating functions

Recurrent relations

Operations & useful properties. Procedure to complement, even and odd operators, representation as a vector or matrix.

system

A set of variables related with a theoretical model, or a set of hypothesis

closed-system

A system in which the state at time t can be predicted perfectly from knowledge of the state at any previous time t'

open-system

The state of the system at time t cannot be predicted perfectly. Some systems appear closed for most cases. An open system is generally believed to be a subset of a larger, closed system. Only the subset is being observed, so we do not know all the variables present, so the ones we see may change in unpredictable ways.

functional explanation for a phenomenon

An explanation that

1. Defines a system
2. Relates the elements of the system to each other in terms of what each element does, rather than in terms of the physical mechanism that performs the action.

Tarski decision procedure

Given a finite system of equations and inequalities, and any number of unknowns, can decide if it has a solution in real numbers.

Taylor's method

Taylor's method for emulating g(x), use a polynomial, and compute each of the coefficients, using the original formula:

$$c_0 =, c_1 = \frac{d}{dt} g(0), \ldots, c_n = \frac{d^n}{dt^n} g(0)$$

Theorem prover

Examines the various hypothesis of the current subsystem. Compares with the logic of the program. Raises exception when the hypothesis is violated (contradicted), violation usually indicates a bug.

time

Sequence index,

see also *distribution (time)*

transform

Integral: $f' = Tf = \int dt\, K(t,u) f(t)$

$$f = T' f' = \int du K'(u,t) f'(u)$$

Matrix: $Tx = \mathbf{A}x$

transition matrix

type of: *incidence matrix* for each state-to-state transition. Can be weighted. The normal form of a transition matrix has the source node as columns, and the destination nodes as rows.

The relative frequency of each state, overall, is called the *stationary distribution* and is the eigenvector of the matrix.

The relative frequency of initial states is called the *initial distribution.*

The relative of frequency of states a given step is called the *marginal distribution.*

Role and cliques. By squaring the matrix and finding its eigenvector, one can find the cliques and roles: items with positive eigenvalues are *members of a clique*, while items with negative values are *links* between cliques (although not in cliques themselves).

see also *eigenvector, graph, incidence matrix, network, stationary distribution*

probability

$$tp_1 = 1 - (1 - tp_t)^{\frac{1}{t}}$$

where

tp_1 = transition probability for 1 year

tp_t = transition probability for t years

see also *probability*

transitive closure

A transitive closure is a graph that answers reachability questions. Can I reach point B from point A? Most often, it is in the form of an incidence matrix that serves as a binary relation: there is a 0 for if you can't reach B from A, otherwise you can.

For any given graph, there is only one transitive closure graph.

size

Transitive Closure Size is a metric used in guiding Database query optimizations (most recursive ones, usually)

transitive reduction

Similar to a Transitive Closure, but the graph is different. It minimizes the number of edges needed to answer the reachability question. You can have several transitive reduction graphs that are equal, but not identical in arrangement

trends decomposition

Linear trend, self (auto correlation),

See also *FFT*

predicting

Predicting trends is often well suited to the Poisson model. Consider a meeting that started at 6:40pm with twelve people being necessary for a quorum. At 6:55pm, 15 minutes after it started, 9 people have arrived (and possibly left). The average rate of arrival, *v,* is 9/15, or 0.6; the total time, *t,* is 15 minutes. Should you wait another ten minutes for another 3 people? The probability of a total of 3 (m) people arrive in 10 (t) minutes is:

Example taken from Pepper Whites book "The Invention Machine."

$$P(3people, 10min) = \left(\frac{9people}{15min}10min\right)^3 \frac{1}{3!} e^{-\frac{9people}{15min}10min} = 0.0892$$

So, no, you should not wait. The Poisson model can also be employed to predict the number of people arriving the next few minutes:

$$\langle m \rangle = \sum_i iP(i,t)$$

And the Poisson model can be used to predict how long to wait for a certain number of people arrive at a large meeting or event:

$$\langle t \rangle = \sum_i iP(m,i)$$

see *arrival (average), Poisson model.*

linear extrapolation

Based on near flat trends:

$$F_{t+1} = F_t + a(A_t - F_t)$$

$$S = \sqrt{\tfrac{1}{T}\sum_i^T (A_i - F_i)^2}$$

a = smoothing constant ~0.2 .. 0.3
= 1-damping factor
A= actual value
S = Standard error of estimate
T = # of observations

uncertainty

$$\varepsilon_{x'} \le 2x - x' < \varepsilon'_{x'}$$

where

x is the real value that x' represents

x' is the representation of $x = \begin{cases} \varepsilon_x & x < \dfrac{\varepsilon_x + \varepsilon'_x}{2} \\ \varepsilon'_x & x \ge \dfrac{\varepsilon_x + \varepsilon'_x}{2} \end{cases}$

width of the x bin is $\dfrac{\varepsilon_x - \varepsilon'_x}{2}$

utility function
exponential

$$U(x) = 1 - e^{-x/\tau}$$

where
x = wealth (in $), or change in wealth
τ = Risk tolerance
= Inverse of risk aversion coefficient

valuation

Estimation, expected value forecasting, net present value, prediction

Vandermonde Matrix form

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ a^2 & b^2 & & z^2 \\ & \vdots & \ddots & \\ a^k & b^k & & z^k \end{bmatrix}$$

$$\det(AB) = \det(A)\det(B)$$

$$\det(A/kA) = k\det(A/2)$$

$$\text{Det} = \begin{matrix} (a-b)\cdot & (a-c)\cdot & \cdots & (a-z)\cdot \\ & (b-c)\cdot & \cdots & (b-z)\cdot \\ & & \cdots & (y-z) \end{matrix}$$

Can be compressed to:

$$[a\, b \cdots z]$$

variance

$$\text{var}(x) = \text{cov}(x,x) = 1\!\!/_n \sum_i^n x_i^2 - \langle x \rangle^2$$

$$\sigma^2 = \int_{-\infty}^{\infty} dx\,(x-\mu)^2 f(x)$$

SVM
support vector machine

see kernel machine

They can't handle truly big data and so can't be trained in a reasonable amount of time. Some of the solvers bombed out before completing it. They are also overly sensitive to outliers.

volume

$$V + dV = (x + dx)(y + dy)(z + dz)$$

WalkSAT

```
for (I=1; I < Max Tries; I++)
{
  Solution = random truth assignment
```

```
    for (J=1; J < Max Flips; J++)
    {
      if all clauses satisfied then return solution
       c = random unsatisfied clause
      with probability p
        flip a random variable in c.  (This prevents us from getting stuck in a local optima)
      else
       flip a variable in c that maximizes the number of satisfied clauses.  This is a hill-climbing,
  greedy-step
    }
  }
  return failure
```

Max WalkSAT

```
for (I=1; I < Max Tries; I++)
{
  Solution = random truth assignment
  for (J=1; J < Max Flips; J++)
  {
    m =sum of weights (sat clauses)
    if (m > threshold) return solution
     c = random unsatisfied clause
    with probability p
      flip a random variable in c.  (This prevents us from getting stuck in a local optima)
    else
     flip a variable in c that maximizes m.  This is a hill-climbing, greedy-step
  }
}
return failure with best solution found
```

wave equation

$$\nabla^2 u = \frac{1}{v^2}\frac{\partial^2 u}{\partial t^2}$$

u = displacement from equilibrium
v = speed of propagation of the waves
f = frequency
λ= wavelength
v=f λ

ω= angular frequency

k=wave number

$$\omega = 2\pi f = \kappa v$$

$$\kappa = \frac{2\pi}{\lambda} = \frac{\omega}{v}$$

$$u' = \begin{cases} \sin\kappa x \sin\omega t \\ \sin\kappa x \cos\omega t \\ \cos\kappa x \sin\omega t \\ \cos\kappa x s \cos\omega t \end{cases} = \begin{cases} \sin\frac{\eta\pi x}{l}\sin\frac{\eta\pi v t}{l} \\ \sin\frac{\eta\pi x}{l}\cos\frac{\eta\pi v t}{l} \end{cases}$$

$$u = \sum_{n=1}^{\infty} b n_u'$$

$$u = \frac{1-i}{y}\sum_{n=1}^{\infty} c_n\left(e^{ik(x+vt)} + e^{-ik(k+vt)} - (1+i)e^{ik(vt-x)}\right)$$

$$c_n = \frac{1}{2l}\int_l^{-l} dx\ g(x)e^{-ikx}$$

weights

Initial. Adjusting. If the adjustment process is a good one, the initial weights can be any value and lose the their bias.

z-score

$$z_i = \frac{AIi - MAI}{\sigma}$$

i = individual

AIi = individual's key characteristics

MAI = median population key characteristics

z < -2 problem, underperforming group

z < -3 sever problem

Above and below normal

Selecting key characteristics ("key descriptor variables")

Zipf's law

$$frequency = \frac{k}{rank}$$

BB Mandelbrot's revision

$$frequency = \frac{k}{(C + rank)^{\theta}}$$

Where K is such that the frequencies sum to n

C is 1… 100

Table 30: **Time-influenced distributions (normalized)**

Distribution	*Density function* $f(m,t)$	*Special properties*
Bessel		
Beta		
Markov, linear birth	$(1-e^{-\lambda t})^{m-1} e^{-\lambda t}$	
Poisson	$\frac{(\lambda t)^m}{m!} e^{-\lambda t}$	
Present value	$x(1+r)^{-t}$	
Transition probability	$tp_t = 1-(tp_1 - 1)^t$	
Weibull power law	$Ge^{-\lambda t}$ (rate)	

Math forms on left

$(f * k)(x) = \int_{-w/2}^{w/2} d\delta\, f(x) k(x-\delta)$ See *convolution*

$(f \circ g)(x) = f(g(x))$ composition

A^{-1} Inverse of A

A^{+} Pseudo-inverse of A

A**x**=b To solve for **x**, use left (back) division. To solve for b, use simple matrix multiplication. Back division is tricky to implement overall efficiently. It is hard to construct the matrix and vector.

see also *condition number*

A**x**=s**x** To solve for **x**, use eigenvalue or eigenvector solver.

$A/B = AB^{-1}$
$A/B \approx AB^{+}$

$A\backslash B = A^{-1}B$

b**x** Left division is, for performance, done differently with triangular matrices, symmetric (Hermitian) matrices, square non-Hermitian matrices, and non-square matrices.

$c = \sum \Delta x f(x)$ to calculate c, use trapz

$c = \int_0^b dx\, e^{-x^2}$ to calculate, use erf(x), with proper scaling factors

$c = \int_a^\infty dx\, e^{-x^2}$

to calculate, use `1-erf(x)` with proper scaling factors

$c = \int_{-\infty}^\infty dx\, e^{-x^2}$

To calculate convert to another representation:

$$c = \sqrt{\int_{-\infty}^\infty dx \int_{-\infty}^\infty dy\, e^{-(x^2+y^2)}}$$

$$c = \sqrt{\int_0^\infty dr\, r \int_0^{2\pi} d\theta e^{-r^2}}$$

$c = \int_a^b dx\, f(x)$

to calculate c, use `quad()`, `quadl()`

$c = \iint dx dy\, f(x,y)$

to calculate c, use `quad()`, `quadl()`

$c = \int_{-\infty}^\infty dx \int_{-\infty}^\infty dy f(x^2+y^2)$

To calculate, change coordinates so that the function is in one variable:

$$c = \int_0^{2\pi} d\theta \int_0^\infty dr\, r f(r^2)$$

${}^nC_r$

see *combination*

$c = (x+a)^n$

$$(a+b)^N = \sum_{n=0}^N {}^NC_n a^n b^{N-n}$$

$$(a+1)^N = \sum_{n=0}^\infty {}^NC_n a^n; N<0, |a|<1$$

See *binomial expansion*

e^x

$$e^x \approx \frac{1.0000725 + 0.50863618x + 0.08582937x^2}{1.0 - 0.49109193x + 0.07770847x^2}$$

$$\max error(r) = 0.86899 \times 10^{-4}$$

$f(x) = \sum \Delta x g(x)$

To calculate f(x), use cumtrapz, cumsum

$f(x) \approx g(x)$

To convert a measured g(x), or more complicated formula g(x) into a simpler approximation, f(x), consider linear regression, root finder, or function minimization.

$\Gamma(n)$

$$\Gamma(n) = \int_0^\infty dx\, x^{n-1} e^{-x}$$

$$\Gamma(n) = c^n \int_0^\infty dx\, x^{n-1} e^{-cx}$$

$$\Gamma(n+1) = n\Gamma(n)$$

$$\Gamma(n)\Gamma(1-n) = \frac{\pi}{\sin \pi n}$$

$$\frac{d}{dn}\Gamma(n) = \int_0^\infty dx\, x^{n-1} e^{-x} \ln x$$

use the gamma() function

$\ln\Gamma(n)$ use the lgamma() function

γ

Euler's constant (used in gamma cacluations)

λ

The arrival rate in queuing models.

${}^nM_r$

see *multiples*

${}^nP_r$

see *permutation*

$P(A|s)$

Probability of A being true, given that s is true, i.e. $P(s \to A)$. Also: that event s will lead to event A.

$$P(D \mid M) = \int \theta d\theta P(D \mid \theta) P(\theta \mid M)$$

see also *probability*

$P(x \le b)$

The probability, in a Gaussian (or normal) distribution, that a value *x* is less than or equal to *b.*

$$P(x \le b) = \tfrac{1}{2} + \tfrac{1}{2} erf\left(\frac{b-m}{\sqrt{2}\sigma}\right)$$

$P(a < x)$

The probability, in a Gaussian (or normal) distribution, that a value *x* is greater than *a.*

$$P(a < x) = 1 - P(x \le a)$$

$P(a < x \le b)$

The probability that a value *x* is between *a* and *b.* This is typically calculated from a probability density function:

$$P(a < x \le b) = \int_a^b dx\, f(x)$$

In a Gaussian (or normal) distribution this is:

$$P(a < x \le b) = P(x \le b) - P(x \le a)$$

$$P(a < x \le b) = \tfrac{1}{2} erf\left(\tfrac{b-m}{\sqrt{2}\sigma}\right) - \tfrac{1}{2} erf\left(\tfrac{a-m}{\sqrt{2}\sigma}\right)$$

$P(m,t)$

The probability of having *m* events in time *t*

See *Markov model (linear birth process), Poisson Distribution*

$R_{P,N}$

Thermal noise power at receiver input

$$R_{P,N} = 10\log\left(kT_N B_N 10^3\right) dBm$$

k = Boltzmann's constant
T_N = Temperature of the noise, 2140 Kelvin
B_N = Bandwidth of the noise 36 10^6 Hz

See also *bit energy, message energy*

$Z(z)$

The probability density function, centered on the mean of the distribution, and z is in units of standard deviations.

See also *probability*