**November 27,2019:**

**Schema theorem Genetic Algorithm (Holland)🡪** The Schema Theorem says that short, low-order schemata with above-average fitness increase exponentially in frequency in successive generations. The theorem was proposed by John Holland in the 1970s. It was initially widely taken to be the foundation for explanations of the power of genetic algorithms.

**Building block hypothesis🡪** A genetic algorithm seeks optimal performance through the juxtaposition of short, low-order, high performance schemata, called the building blocks.

**Exact Schema Theorem🡪** Holland's schema theorem has three parts, one for selection, one for crossover, and one for mutation. The selection part is exact, whereas the crossover and mutation parts are approximations.

**Genetic variety :** Genetic variation refers to diversity in gene frequencies. Genetic variation can refer to differences between individuals or to differences between populations. Mutation is the ultimate source of genetic variation, but mechanisms such as reproduction and genetic drift contribute to it as well.

Variance= (std dev)2

V= σ2

**Dispersion measure:** In statistics, dispersion (also called variability, scatter, or spread) is the extent to which a distribution is stretched or squeezed. Common examples of measures of statistical dispersion are the variance, standard deviation, and interquartile range.

Set-(unordered) collection of elements.

Cardinality: uniqueness of elemsents.

For example:

Orders set 🡪 |{1,2,3,4}|= 4

|{1,2,3,4,4}|=4

Membership E:

Set has no structure to track no of elements m in it only tell that u are member or not.

Position= Ordered.

**Ordinal VS Cardinal 🡪**

Cardinal numbers: one, two, three, four, five, … count things; ordinal numbers: first, second, third, fourth, fifth, … specify the order of things in a sequence. In that sense they are different. But of course order is determined by counting, e.g. the number of runners reaching the finish line in the order of arrival.

Carl Friedrich Gauss. Carl Friedrich Gauss was a prominent figure in the nineteenth century Germany for his accomplishments in the discipline of mathematics. He is known for his monumental contribution to statistics, algebra, differential geometry, mechanics, astronomy and number theory among other fields.

In statistics, a **Poisson** distribution is a statistical distribution that shows how many times an event is likely to occur within a specified period of time.

Mean:

In math, the mean is the average of a set of numbers. To find the mean of a data set, add up all of the numbers in the set, and then divide that total by the number of numbers in the set.

Median: The "median" is the "middle" value in the list of numbers.

Mode: Number/value that appears the most.

Qualitative vs quantitative:

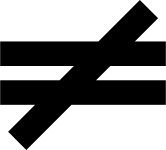
Quantitative data is information about quantities, and therefore numbers, and qualitative data is descriptive, and regards phenomenon which can be observed but not measured.

Deviation : distance from

1. **( i=n Σn** (xi- x̄) )/ n = i=n**Σ**n (xi) / n

= n x̄ / n

= x̄ - x̄ = 0

Data = (**Σ (**xi- x̄) )/ n  0

2. ( **Σ** |xi-x|) / 2(n).

3. ( **Σ** (xi- x̄)2)/n

V = variance

σ standard deviation

Under simplification assumptions of central tendency describe data set by

1. N size
2. x̄ average
3. σ

1, 2, 3, 4, 5

N = 5, x̄= 3

(**Σ** xi)/n= x̄ = nx̄ = j**Σn** xi= 15 vs 13

15-13=2

Y =2x 🡪 independent

Where y = dependent

d.o.f. = degree of freedom = 1

**Statistics** is a collection of Quantitative data; a characteristic of a sample. It is used to estimate the value of a population parameter.

**Variance** is a measure of how dispersed or spread out the set is. A variance of zero indicates that all of the data values are identical. All non-zero variances are positive. It is also the expectation of the squared deviation of a random variable from its mean.

**Fundamental of Statistics:**

**Statistics**: Statistics is the science concerned with developing and studying methods for collecting, analyzing, interpreting and presenting empirical data. Recalling that all data can classified as either, categorical or quantitative.

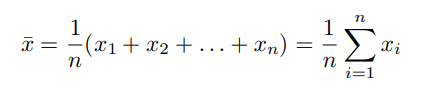
**Measures of Central Tendency:**

A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data. As such, measures of central tendency are sometimes called measures of central location. They are also classed as summary statistics. The mean (often called the average) is most likely the measure of central tendency that you are most familiar with, but there are others, such as the median and the mode.

The mean, median and mode are all valid measures of central tendency, but under different conditions, some measures of central tendency become more appropriate to use than others.

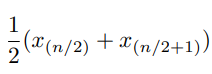
**Few Fundamental Properties of Statistics:**

**Means:** For a collection of numeric data, x1, x2, . . ., xn, the sample mean is the numerical average

****

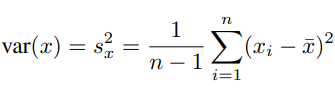
**Medians:** The median take the middle value for x1, x2, . . ., xn after the data has been sorted from smallest to largest, x(1), x(2), . . . , x(n) .

If n is odd, then this is just the value of the middle observation 

If ns is even, then two values closest two the center are averaged. 

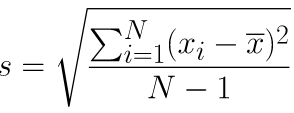
**Mode:** The mode of a set of data values is the value that appears most often. For example:

In set {1, 2, 3, 4, 4, 3, 4, 3, 4, 5, 6, 4, 5} the mode is 4, as it occurs most often.



**Standard Deviation:** Standard Deviation is the square root of the sample variance. Standard Deviation measures how spread out the values in a data set are around the mean. More precisely, it is a measure of the average distance between the values of these data in the set and the mean.

Let population consist n elements, {x1, x2, x3. . .. xn}, with the mean of x ̅



**Independent Variables:** An independent variable is a variable whose variation does not depend on another variable. The independent variable is the variable that is manipulated by the experimenter to determine its relationship to an observed phenomenon, called the **dependent variable.**

**Dependent Variables:** Dependent variable is a variable whose variation is depend on another variable.

For example, let’s assume we have following equation. Where,

Now if you we pick the value for *r* then we can’t pick the value for *A,* in this case *A* is dependent variable and *r* is independent variable.

**Degrees of Freedom:** Degrees of freedom in a statistical calculation represent how many values involved in a calculation have the freedom to vary. The degrees of freedom ca be calculated to help ensure the statistical validity of chi-square tests**,** t-tests and more. For example, given average of four numbers, which is 3.75, and there are only 3 numbers are given, out of four. We need to find the number that missing, lets assume missing number is *x*

(5 + 5 + x + 2) = 3.75 \* 4

(5 + 5 + x + 2) = 15s

X = 15 – 5 – 5 – 2 s

X = 3

[where x has no freedom, its values are based on other variable values].

**Keywords:** schema theorem, building block hypothesis, exact schema theorem, genetic variety, dispersion measure, ordinal, cardinal, Poisson, degrees of freedom, dependent variables, independent variables, standard deviation, mode, medians, means, measures of central tendency, statistics, variance

**Notes12/02/2019**

**Genetic Algorithm**

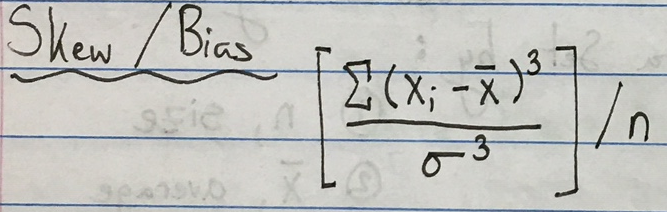
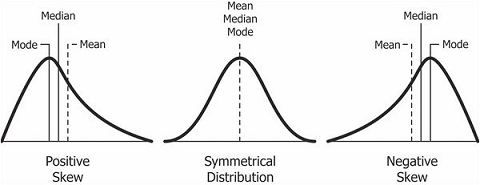
Statistics about the initial population (and/or subsequent) about the genetic algorithm(“evolutionary”) process.

**T-test:**

A t-test is a type of inferential statistic used to determine if there is a significant difference between the means of two groups, which may be related to certain features.

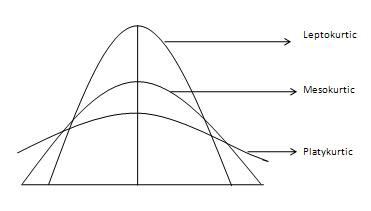
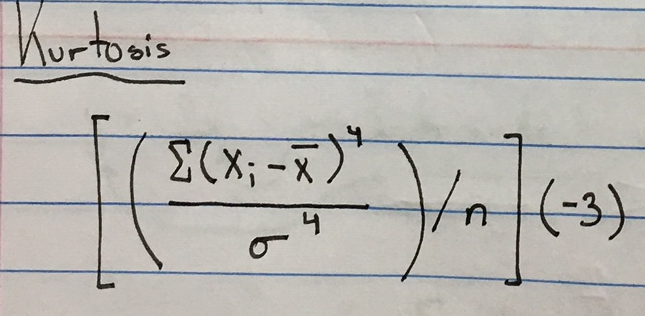
**Skew/Bias:**

Skewness refers to distortion or asymmetry in a symmetrical bell curve, or normal distribution, in a set of data. If the curve is shifted to the left or to the right, it is said to be skewed.

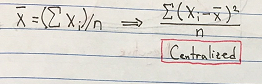
**Kurtosis:**

Kurtosis is a statistical measure that is used to describe the distribution. Whereas skewness differentiates extreme values in one versus the other tail, kurtosis measures extreme values in either tail.



**Moment:**

Class of functions based on the powers of variable.



**Z-score:**

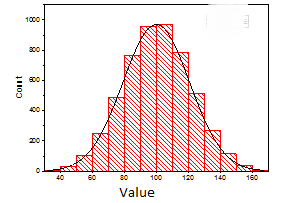
A Z-score is a numerical measurement used in statistics of a value's relationship to the mean (average) of a group of values, measured in terms of standard deviations from the mean.

Deviation – Distance from the mean, z-score connect mean of different data sheet.



**Distribution histogram:**

Used for display frequency, a frequency distribution that demonstrates how repeatedly each unlike value in a data set occurs.



Given any data set,

67% will be in 1st Sigma

95% will be in 2nd Sigma

97% will be in 3rd Sigma

**IQR**: Inter Quartile Range🡪 which means 50% of the data counted. Rest doesn’t show the other performance.

Statistics in GA: We can use Statistics in termination of Genetic algorithm. Now we have termination based on max generation or no better fitness will produced in next generation.

So, we can use statistics variance to stop GA. We have to make sure that variance large enough to simulated annealing.

**Simulated Annealing** a process that refers to at some point we cannot check for global max and min, so we chuck the point which could be local max. so after throwing it away in different ways, if it comes back at same position, then that could be global max. We cannot be sure but at that point, we can assume that is the global max.

**Sigma scaling:**

Ranked based selection: rank is based on z-score.

In genotype approach, we have 0,1,\*.

A schema patter, H = 10\*1\* 🡪 there is four choice:

1. 10010
2. 10011
3. 10110
4. 10111

Since \* can produce either 0 or 1.

Order(H) -> fixed number of 0 and 1’s in H.

Defining length of schema, H 🡪

Delta(H) = 22-19 = 3

H’ = 0\*\*\*\*

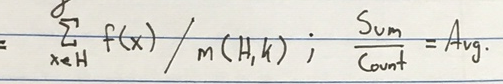
Delta(H’) = null

Fitness function, F(x)

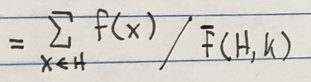
Generation of K, so

F(H,K) = range of values

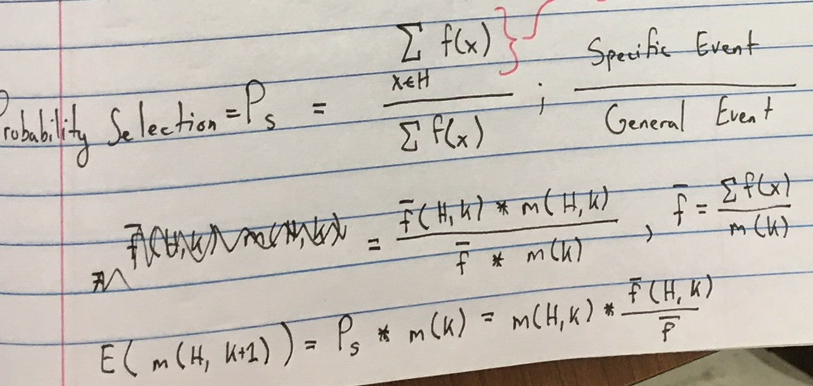
Average((H,K)) = average fitness of individual chromosomes schema H have,



M(H,K) = number of chromosomes that schema H have



Probability of Selection :



**Keywords:** t-test, skew/bias, kurtosis, moment, z-score, distribution histogram, IQR, simulated annealing, sigma scaling

**December 4th, 2019**

**Schema Theorem**

* + Proposed by John Holland (1970s)
  + Fundamental Theorem of Genetic Algorithms
  + Schema with with above-average fitness, short defining length, and lower order is more likely to survive
  + Schemata – a template that consists of a subset of strings with similarities at certain positions in the string
* The fitness of a schema is the average fitness of all the strings matching in the schema.

**Genotype Level (Assumption #1)**

f(x) = fitness of chromosome x

G – population

k – time

(G) = average fitness of all chromosomes in G

N = |G| = m(k)

(H) = ∑ f (x) / |H|

x ∈ H

Average (mean) = sum / count

Membership: m (H, k) = |H| at iteration k

Schema H =

1 0 \* 1 \*

1 0 **0** 1 **0**

1 0 **0** 1 **1**

1 0 **1** 1 **0**

1 0 **1** 1 **1**

\* indicates that the value can be 0 or 1

**m(k) = |generation| at iteration k**

**assumption #2 – fixed for notational purposes**

Average fitness at time k

(Gk) = ∑ f (xi) m (k) = count

xi ∈ Gk

m(k) = ∑ f (xi) ∑ f (xi) = m(k) \* (Gk)

xi ∈ Gk ; xi ∈ Gk

\_\_\_\_\_\_\_\_\_\_\_\_

(Gk)

Average fitness for members in H (at time k):

(Hk) = ∑ f (xi)

xj ∈ Hk

m (H, k) = count

m (H, k) = ∑ f (xj) ∑ f (xj) = m (H, k) + (Hk)

xj ∈ Hk.  ; xj ∈ Hk

\_\_\_\_\_\_\_\_\_

(Hk)

**Defining Length:**

𝛿 (H) = position (last fixed bit) – position (first fixed bit)

𝛿 (H) = 4 – 1 = 3

(actual length of H = 5)

Probabilistically, average is viewed as the **Expected Value.**

Ordered set S = {1, 2, 2, 3, 3, 3, 4, 4, 4, 4}

|So| = 10

Unordered set S = {1, 2, 3, 4}

|Su| = 4

Probability = frequency (events) / frequency (all events)

Example: Eye Color

**B**rown, **bl**ue

Random Sample Sa = 1000 people

p (x = 1) = 1/10 -> BB

p (x = 2) = 2/10 -> Bb

p (x = 3) = 3/10 -> bB

p (x = 4) = 4/10 -> bb

Out of 1000 people, 75% will most likely have brown eyes and 25% will most likely have blue eyes.

E (x ∈ Soa) = ∑ p (xj) \* xj

xj ∈ Su

(1/10) \* 1 + (2/10) \* 2 + (3/10) \* 3 + (4/10) \* 4 = 30/10 = 3 // weighted average

Expected value (blue) = probability (blue) \* |Sa| = 25% \* 1000 = 250

Schema H

m (H, k)

m (H, k H)

GA – refers to selection/crossover/mutation

E (m (H, k H)) = GA (m (H, k))

**Assumption #3 – Roulette Wheel Selection**

Ps (x) = f (x)

\_\_\_\_\_

N

∑ f (xi)

i = 1

Ps (H) = ∑ f (x)

x ∈ H = ∑ f (xi)

\_\_\_\_\_\_\_\_ x ∈ H

N \_\_\_\_\_\_

∑ f (xi) ∑ f (xj)

i = 1 xj ∈ G

Expected Value:

E (m (H, k H)) = Ps (H) \* m (k)

= ∑ f (xi) \* m (k) = ∑ f (xi) \* m (k) = m (H, k) \* (Hk)

xi ∈ Hk  xi ∈ Hk \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_ \_\_\_\_\_\_\_\_ (Gk)

∑ m (k) + f (Gk)

xj ∈ Gk

Psurvival = 1 – Pdisruption

**Defining Length**: number of crossover points that could affect the schema

Given l bits

* Length of chromosome: l
* Split: l – 1 (number of places you could possibly break things up)
* Probability of crossover points affecting your schema?

Pc (H) = 𝛿 (H) / (l – 1)

Probability of parameter affecting your schema:

Pc (G) – will it occur

Pc (H) – if so, where will it occur

Pc (G) \* Pc (H)

**Probability of Survival:**

l – Pc (G) 𝛿 (H)

\_\_\_\_

l – 1

Pm (G) – GA parameter

Your schema is affected in o(H) places

(l – Pm (G)) – will it occur

o(H) – where will it occur

Psurvival mutation = (l – Pm (G)) o(H)

**Keywords:** length, expected value, roulette wheel selection, schema theorem, genotype level