Correlation depends on series pairing so order matters. The assertion that two series with the same mean and different uncertainty will have a correlation of one is empirically false. For example, the series (1,2,3) and (4,1,1) have the same mean, different uncertainty (standard deviation) and a correlation of 3.464 which is not 1. The procedure to figure correlation of two series is as follows: first calculate the mean of each series. Then calculate the deviation of each point in a series from that series mean by subtracting the mean from each point. Then square the deviations of each series and sum each which results in two values: a2 and b2. Then multiple the deviations of each series and find the sum which results in one value: ab. Multiply a2 and b2 and take the square root. Finally, divide ab by this value to get correlation.

Perhaps the best way to answer this question is to delve into the reality of coin flipping. Consider the flipping a biased and unbiased coin one hundred million times each (or some really, really big number of times), keeping track of the prior flips and calculating likelihood of heads or tails after each flip. Knowing nothing about these coins, suppose you flip a heads with both coins on the first flip, what is the uncertainty of the outcome of the next flip? Is it not the same with both coins? Lets say after 500,000 flips, the mean of the prior flips is 50% heads on one coin and 60% heads on the other. What are your expectations for the next flip? What is the uncertainty? These question are meant to suggest that what comes next is related to expectations based on what was prior. The fact that the coin is biased does not change the uncertainty of the next flip.

There are two statements in this question and the later depends on the former. The first statement is that a quadratic equation results in wavy line having many peaks and valleys. Anyone who recalls plotting a quadratic on their graphing calculators in high school may remember the result is not a wavy line, but rather a parabola. This is only one global optima and that is the vertex. So the first statement is false. Likely one could stop answering the question at this point, but let us move on and consider the second statement which says the problem is intractable. This is a big word which-in the computer science world-means cannot be calculated in polynomial time and hence is not solvable. Is solving for an extrema of a parabola intractable? No.

A determinant is of a matrix is a long and tedious calculation. A positive matrix is symmetrical (same number of rows and columns) and all values in the matrix are positive. When I think of eigenvectors, I think of a lamda. So, this question totally confused me.

In order for a distance metric to be valid, it must obey the Triangular Inequity Theorem (TIT) which states the sum of any two sides of a triangle must exceed the length of the other side, in Euclidean space. Put another way the shortest distance between two points is a line. This is related to the Shortest Line Theorem which states that the shortest distance from a point to a line is the perpendicular distance. Consider the classic problem of the travelling salesman. This is an optimization problem to figure the shortest path of a salesman who wants to visit all cities and end up back where he started. If you apply TIT to this problem, then the salesman resalizes that travelling from one city to another directly is always shorter than travelling to a city via another city. Intuitively, going through 1 city to get to another is longer than just going directly to the city in a straight line. The metric violates TIT, it cannot be a valid distance metric. KL violates this theorem.

According to the central limit theorem, the sample mean and standard deviation will approximately equal the population mean and standard deviation. If the population is normally distributed, then the sample is normally distributed. But even if the sample is not normally distributed, the sample mean and standard deviation will be approximately normally distributed also.

The sigmoid function-which just means S like. It is a differentiable function unlike a thresholding function. The sigmoid function is s(a) = 1 / (1 + e^-a). The derivative of this function is D s(a) = s(a) \* (1 – s(a)). One can see how as “a” gets very positive it goes to 1… 1 / (1 + e^-1000) and as “a” gets very negative, it goes to 0.

False. Kernels do something called projection which is a descriptive term illustrating how kernels raise and lift dimensions so they can be sliced.

False. Mr Wikipedia says that covariance distance does not adhere to the triangular inequality theorem, so it cannot be a valid distance metric. Please see question above for more details.