Abstract

By accepting the uncertainty principle in an Euclidean plane, the position of point, line, and every geometric shape can be identified by the statistical parameters of mean and standard deviation. A heuristic model of Euclidean Statistical Geometry will be established. An axiomatic approach for undefined terms and their use in geometric objects is presented. The consistency of geometric interpretations of the mean, variance, and correlation coefficient for random vectorial events in the Euclidean plane are discussed.

Key Words: Euclid’s postulates, undefined terms, fuzzy points, fuzzy lines, belongs to, statistical geometry.

1 - Introduction to History of Geometric Logic:

Socrates may be the first who taught a formal logic and developed a dialectical logic. Greek society at his time did not appreciate a political implication of that kind of logic. His student and successor Plato, began to continue to use formal logic in his teaching and it was later developed by Plato’s student, Aristotle at the school of Athens. Aristotelian logic was based on his syllogism which was a foundation of deductive reasoning used in geometry about two thousand years. Aristotle claimed that "Thing either is or it is not" [5].

The first systematic approach to geometric proof using deductive reasoning was presented in Euclid’s book of "Elements", a monumental work written when he was teaching at the school of Alexandria in (300 B.C.). Euclid tried to present his definition of "point", "line", and "belongs to". The fifth postulate of Euclid’s geometry occupied the mind of many genius mathematicians for several centuries. Many branches of Non-Euclidean Geometry denying the fifth postulate were created (see [5]).

David Hilbert proposed in 1898 an axiomatic foundation in his book "Grundlagen der Geometrie" Foundation of Geometry, a rigorous approach to Euclidean Geometry. Hilbert presented "point", "line", and "belongs to" as undefined terms. Similarly, element, set, and belongs to in set theory are undefined terms. In his approach a point either belongs to a set or does not. In all areas of logic, set theory, and geometry, the membership of an element of a set is considered as an undefined term (see [4] and [5]).

The formal Boolean logic based on a binary statement of \{yes, no\}, \{true, false\}, or \{0,1\} was developed, flourished, and applied to computation structures in the twentieth century.
In formal logic, we accept a complete statement as a logical proposition that is either true or false. In 1965, Lotfi A Zadeh of the University of California, Berkeley, published a paper titled Fuzzy Sets in which he described the mathematics of fuzzy set theory. Fuzzy logic is used to define the membership in a fuzzy set (see [7]). R. A. Fisher was actually inspired by the geometry and imagination of pictures but found that it was difficult to prove the formulations using geometry (see [1]). The Geometric approach” by Saville and Wood 1997 is actually the same as the approaches in this paper, but we will emphasize the foundation to construct and unify the principles of geometry and statistics [6].

2 - Elements in Statistical Euclidean Plane:

To establish and explore the foundations of this new geometrical approach, the parallel postulate is temporarily put aside. There is need to introduce new definitions for all of the undefined terms: ”point”, ”line”, ”belongs to”, and ”congruence” which was used by Euclid in his "Elements". Our goal is to construct a new approach that satisfies the postulates of Euclidean Geometry and the axioms of the probability space. This combination of two axiomatic systems may be called Statistical or Fuzzy Geometry. We define point (x,y) to be an Euclidean point which is the Mean of a sample subset of discrete points. A line is a statistical line in which every point belonging to this line is the mean of the cluster points. The Heisenberg uncertainty principle explains that one cannot determine the position and velocity of a particle when it is applied to a physical motion. According to this assumption point A is the mean point, the line D is the mean line. The most challenging part is the term ”belongs to” that can be interpreted as a set membership. In this introductory article, the undefined terms ”betweenness” and ”congruence” will not be tackled. In a practical example when the cloud of uncertainty is attached to "point", "line" and plane, the position of the point will be identified by the mean and variance of the sample data points in the plane.

3 - Review Foundation of Geometric Probability Space:

Harding and Kendall 1974 used a different approach to introduce their ”point process” and ”line process” in stochastic geometry in the Euclidean plane. The foundation in a heuristic approach for a discrete sample space will be presented here.

Sample Space in the Euclidean Plane: Assume random events $E_1 = \{ x_i \in I_1 \subset R: a \leq x_i \leq b, \ i = 1, 2, ..., N \}$ and $E_2 = \{ y_j \in I_2 \subset R: c \leq y_j \leq d, \ j = 1, 2, ..., M \}$. Define $\Omega = E_1 \times E_2$ then all possible subsets of $\Omega$ will be a sample space.

Event: The power set of $\Omega$ is the set which contains all elements $A = \{ u = (x,y) \in \Omega \subseteq R \times R \}$. Every element of the power set $S = P(\Omega)$ is called an event. As described in the following, the event could be a point, line, or any region of $\Omega$ while considering a triplet $(\Omega, S, p)$ as a probability space. Foundation of Fuzzy Geometry: Consider a plane whose elements are ”points” and ”lines”. Assume that the relationship ”belongs to” and ”equal to” are undefined terms. Adopting the uncertainty principle
means that it is not possible to determine exactly the position of the elements of the plane and their relationship.

Fuzzy Point: Given any event A ∈ S = P(Ω), a fuzzy point is an ordered pair \( \mathbf{w} = (x, y) \) such that \( x \) and \( y \) are the mean of some \( x_i \) and \( y_i \) for \( i = 1, 2, ..., n \). For the sake of understanding the notions of fuzzy point, let us assume a brush marking the flat plane like a fuzzy pencil such that among all of these points there is a point which marks the plane differently and represents the mean point.

Fuzzy Line: A line on which every point is a fuzzy point. When we draw a line by a fuzzy pencil then among all lines there will be one line on which each point is a fuzzy point.

Fuzzy Triangle: Suppose three statistical lines \( a, b, \) and \( c \) represent three events \( A, B, \) and \( C \) from the sample space \( \Omega \). The triangle \( \triangle ABC \) is called a statistical triangle. We accept the following four fundamental Euclid type postulates in the plane without further proof.

**Postulates**: Having clarified all undefined terms we now lay out our postulates.

i) There exists exactly one fuzzy straight line passing through two fuzzy points.

ii) Every fuzzy line can be extended endlessly.

iii) Given a fuzzy point, there exists only one fuzzy circle with a given radius.

iv) All fuzzy right angles are congruent.

Random Vector: Assume vector \( \mathbf{x} = (x_1, x_2, ..., x_n) \in \mathbb{R}^n \) with the identity vector \( \mathbf{e}_i \in \mathbb{R}^n \) and a probability function \( p \in (\Omega, S, p) \). We define random vectors \( \mathbf{u} \) and \( \mathbf{v} \) such that

\[
\mathbf{u} = (\mathbf{x}, p(\mathbf{x})) = \sum_{i=1}^{n} x_i p(x_i) \mathbf{e}_i \quad \text{and} \quad \mathbf{v} = (\mathbf{y}, p(\mathbf{y})) = \sum_{i=1}^{n} y_i p(y_i) \mathbf{e}_i
\]

It is important to introduce addition and scalar multiplication for random vectors.

i) \( \mathbf{u} + \mathbf{v} = (\mathbf{x}, p(\mathbf{x})) + (\mathbf{y}, p(\mathbf{y})) = (\mathbf{x} + \mathbf{y}, p(\mathbf{x} + \mathbf{y})) \)

ii) \( a \mathbf{u} = a \cdot (\mathbf{x}, p(\mathbf{x})) \), for any constant real number \( a \in \mathbb{R} \).

Inner Product of Random Vectors: Given two random vectors \( \mathbf{u} \) and \( \mathbf{v} \) we define the inner product by the following

\[
\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{n} x_i p(x_i) v_i^* = \sum_{i=1}^{n} x_i p(x_i) y_i = \mathbf{u}^T \cdot \mathbf{v}
\]

This general definition for random vectors can be used to establish a fuzzy vector space.

Mathematical Expectation: Let us use \( n \) for sample size and \( N \) for population size. If one dimensional random vector \( \mathbf{x} \) represents an observation event \( \{x_1, x_2, ..., x_N\} \) with the probability \( p(x_i) \) in the homogeneous plane (density is uniform) then we define the mean using the expectation of a random variable of \( \mathbf{x} \) as follows

\[
\mu_x = E(\mathbf{x}) = \sum_{i=1}^{n} x_i p(x_i) = \frac{\sum_{i=1}^{n} x_i}{N}
\]

The population mean for random variables \( X \) and \( Y \) is denoted by \( \mu_X \) and \( \mu_Y \).

Sample Variance: Suppose that the mean of a sample is given, the second moment about the mean is called the variance, that is

\[
s^2 = E(\mathbf{x} - \mu)^2 = \sum_{i=1}^{n} (x_i - \mu)^2 p(x_i)
\]

In this formulation \( x_i \) is the set of a finite number of points on the statistical line with homogeneous distribution and \( p(x_i) = 1/n \). Thus relation (2) can be modified to \( s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 \).

The sigma in this definition can be represented by the magnitude of a vector connecting the point \( \mathbf{x} \) from
a fixed point \( \pi \). This relation implies that \( s_x = \frac{1}{n^2} \| \pi^2 - \pi \| . \)

When \( \pi \) is a random vector in a plane then for every selected sample, there is an associated mean \( \pi \) and variance \( s_x \) such that the magnitude \( \| \pi^2 - \pi \| \) is constant. Thus, for every sample there is a circle centered at \( \pi = (a, b) \) and radius \( s_x \). The geometric interpretation of the variance of two random vectors which is distributed in two dimensional plane \((\pi, \pi')\) would be very helpful to proceed further. When a sample is selected, then the associated centers \((\pi, s_u)\) and \((\pi, s_u)\) represent two circles centered at \( \pi = (a, b) \) and \( \pi = (c, d) \). The relations between these two circles have very important applications both in statistical and geometrical points of view.

Properties of Random Vector Space: The following basic relationship involving inner products can be verified (see [2]). Suppose that \( a = (a_1, a_2, ..., a_n) \) and \( b = (b_1, b_2, ..., b_n) \) are vectors and \( X = (X_1, X_2, ..., X_n) \) is a random vector with uncorrelated components of equal variance \( \sigma^2 \). Then

(a) \( E(a, X) = \langle a, E(X) \rangle \),
(b) \( \text{Cov}[<a, X>, <b, X>] = \langle a, b \rangle > \sigma^2 \),
(c) \( \text{Var} <a, X> = \|a\|^2 \sigma^2 \).

When the random vector \( X \) has uncorrelated components of equal variance \( \sigma^2 \) and the vector \( a = (a_1, a_2, ..., a_n) \) is a unit vector the \( \text{Var} <\pi, X> = \sigma^2 \).

The mean of a probabilistic vector \( \mu = (\pi, p(\pi)) = \sum_{i=1}^{n} x_i p(x_i) \), can be expressed in the following form

\[
E(\mu) = \mu = \sum_{i=1}^{n} x_i p(x_i) = <\pi, p(\pi) > \]

where the superscript \( T \) represents the transpose. It can be verified that \( E(\mu + \nu) = E(\mu) + E(\nu) \).

4 - Geometric Interpretation of Sample Correlation Coefficient:

A ”relation” between the elements of two sets \( A \) and \( B \) is interpreted by any subset of the Cartesian product \( A \times B \). When \( A \) and \( B \) are events, then the sample correlation coefficient is a measure which quantifies the degree of (linear or nonlinear) association among two or more random variables. It is important to emphasize that when we assume a linear association and search for a linear model we may miss more complicated associations where the variables are related in a nonlinear fashion. There is always a danger to miss certain aspects of the data when we characterize a scatter plot with just a single summary statistics.

Assume \((\pi, \eta)\) is a point that is a center of sample data. We can draw vertical and horizontal lines through the center point \((\pi, \eta)\). These lines divide the scatter plot into four quadrants. Deviations \((x_i, y_i)\) are important measures but standard deviations \((x_i - \pi)/s_x\) and \((y_i - \eta)/s_y\) which are dimensionless measures are useful for measuring the relationship and association.

The sample correlation coefficient depends on the product \((x_i - \pi)/s_x \) and \((y_i - \eta)/s_y\) and is defined as the average

\[
r = \frac{1}{n - 1} \sum_{i=1}^{n} \left( \frac{x_i - \pi}{s_x} \right) \left( \frac{y_i - \eta}{s_y} \right)
\]

The right hand side can be modified to vector form

\[
r = \frac{\sum_{i=1}^{n} (x_i - \pi)(y_i - \eta)}{\sqrt{\sum_{i=1}^{n} (x_i - \pi)^2} \sqrt{\sum_{i=1}^{n} (y_i - \eta)^2}}
\]

Geometric Interpretation of Correlation: Assume that data are distributed in the Euclidean plane. The correlation coefficient between two vectorial events can be interpreted as a
dot product of unit vectors of two events; or the projection of one vectorial event in the direction of the unit vector of another event emanating from the center of the sample. Covariance: Given two vectorial events \( \vec{x} \) and \( \vec{y} \) from the sample space. From the traditional definition of the covariance of these two random variables comes the mean value of the product of the deviations from their own means. In symbols: 
\[
\text{cov}(\vec{X}, \vec{Y}) = E(\vec{x} - \bar{x})(\vec{y} - \bar{y}).
\]

Using the fact that the joint probability of \((x_i, y_j)\) will be \(p(x_i, y_j) = 1/N^2\), the marginal probability will be \(p(x_i)\) and \(p(y_j)\). Then the relation covariance can be expanded such that 
\[
\text{cov}(X, Y) = \frac{1}{N} \sum_{i,j=1}^{N} (x_i - \bar{x})(y_j - \bar{y})p(x_i, y_j) = \frac{1}{N^2} \sum_{i,j=1}^{N} (x_i - \bar{x})(y_j - \bar{y}) = \frac{1}{N^2} \sum_{i,j=1}^{N} (x_i - \bar{x})(y_j - \bar{y}) = \frac{1}{N^2} (\vec{x} - \bar{x})(\vec{y} - \bar{y})^T.
\]

Notice that symbols \((\vec{x} - \bar{x})/N\) and \((\vec{y} - \bar{y})/N\) have beginning and end points, so they represent vectors. The magnitude of the vectors in the above dot product can be calculated. The dot product of the two vectors in this relation is 
\[
\left\{ \begin{array}{l}
\| (\vec{x} - \bar{x})/N \| = \| \vec{x} - \bar{x} \| = s_x \\
\| (\vec{y} - \bar{y})/N \| = \| \vec{y} - \bar{y} \| = s_y
\end{array} \right. \tag{4}
\]

Thus, the covariance relation can be interpreted in terms of the magnitude and angle between two vectors. That is 
\[
\text{cov}(\vec{x}, \vec{y}) = \| (\vec{x} - \bar{x})/N \| \| (\vec{y} - \bar{y})/N \| \cos(\omega) \tag{5}
\]

Using the relation (4) in (5) we obtain \(\text{cov}(X, Y) = s_xs_y \cos(\omega)\). Thus, the angle between two random vectors is 
\[
\cos(\omega) = \frac{\text{cov}(X, Y)}{s_xs_y}.
\]

Lemma 1: For the entire sample space the variances \(\sigma_u\) and \(\sigma_v\) will satisfy 
\[
\text{cov}(\vec{u}, \vec{v}) = \sigma_u \sigma_v \cos(\omega). \tag{6}
\]

This is a strange result. However the analogy of covariance and dot product has real meaning. This is taking place when we define \(\omega\) as an angle at a common point between two variance circles \((\vec{u}, \sigma_u)\) and \((\vec{v}, \sigma_v)\). The covariance is the dot product between two vectors \((\vec{x} - \bar{x})/N\) and \((\vec{y} - \bar{y})/N\) in the statistical plane in which the data are distributed in a fuzzy geometric plane. This angle can actually explain the mysterious similarity between our standard statistics and fuzzy Euclidean Geometry. The following theorem is the immediate result of our discussions.

**Theorem 1:** The correlation coefficient between two statistical lines \(U\) and \(V\) is 
\[
\rho = \cos(\vec{x}, \vec{y}).
\]

Proof: Since the correlation coefficient is defined by 
\[
\rho = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y},
\]
and we use covariance formula then the result will be obvious. As a result of this theorem we have \(-1 \leq \rho \leq 1\).

**Geometric Interpretation of Covariance:** To explain the geometric interpretation of covariance in the geometric plane we present the following theorem.

**Theorem 2:** Suppose that three fuzzy points \(A, B,\) and \(C\) are given in a fuzzy Euclidean plane. The following relations are equivalent; In a fuzzy triangle \(\triangle ABC\) with segments \(a, b,\) and \(c\) we have
\[
\left\{ \begin{array}{l}
\text{Mean: } a^2 = b^2 + c^2 - 2bc \\
\text{Variance: } \sigma_{a+b+c}^2 = \sigma_a^2 + \sigma_b^2 + \sigma_c^2 + 2\rho \sigma_b \sigma_c
\end{array} \right. \tag{7}
\]

Proof: (i) The position of points \(A, B,\) and \(C\) will be determined by their means and standard deviations. These points will also uniquely define lines \(AB, AC,\) and \(BC\). Three segments \(a, b,\) and \(c\) are well defined by the measurement of the mean and standard deviations of three sides of the triangle \(\triangle ABC\. For every random point \(A, B,\) and \(C\) the cosine relation for the triangle \(\triangle ABC\) by Theorem (1)
produces $\rho = \cos(\omega)$. Substitute this relation for $\cos(\omega)$. The statistical metric relation (7) will be concluded.

(ii) To compute the variance of $b + c$ we apply the definition formula for variance $\sigma_{b+c}^2 = E[(b + c - \mu_{b+c})^2]$. Since $\mu_{b+c} = \mu_b + \mu_c$ the previous relation can be expanded to
\[
\sigma_{b+c}^2 = E[(b + c - \mu_b - \mu_c)^2] = E[(b - \mu_b)^2 + (c - \mu_c)^2 + 2E(b - \mu_b)(c - \mu_c)] = \sigma_b^2 + \sigma_c^2 + 2\text{Cov}(b, c).
\]

Fuzzy Pythagorean Theorem: Two vectors $U$ and $V$ are uncorrelated if $\text{cov}(U, V) = 0$. Therefore, the correlation coefficient for uncorrelated events is equal to zero $\rho = \cos \omega = 0$. From our knowledge of geometry, we conclude that $\omega = \pi/2$. Meaning that the fuzzy lines are orthogonal and from the statistical point of view, implies that the events $U$ and $V$ are uncorrelated. Thus, the orthogonality implies the uncorrelations and vice versa. Particularly from orthogonality $U \perp V$ we will have the following fuzzy Pythagorean Theorem.

Corollary 3.1: Assuming that $a$, $b$ and $c$ are the magnitude of three sides of a fuzzy right triangle then
\[
i) \quad a^2 = b^2 + c^2 \quad \text{and} \quad ii) \quad \sigma_{b+c}^2 = \sigma_b^2 + \sigma_c^2.
\]
For every random vector $\overline{X} = (x_1, x_2, ..., x_n)$ the norm of the inner product $\|\overline{X}\|^2 = \langle \overline{X}, \overline{X} \rangle = (\sum_{i=1}^{n} x_i^2)$.

5 - Conclusion:

By attaching the Heisenberg uncertainty component to points, lines, and planes, we can bridge the gap between two theories; one from twentieth century in probability and the other, geometry with a two thousand year history. Alternatively, we can assume that sample space and events are sets of points in a plane with their probability function. A different logic needs to glue these two spaces that may be called statistical or fuzzy logics. In this attempt, the axioms of the probability space are applied to the Euclidean plane. This means that the foundation of statistical geometry is postulated in the two dimensional Cartesian Coordinate system. This idea can be generalized to higher dimensional Euclidean space. In this article, we started with a simple discrete probability space in a flat Euclidean plane where the events are simply, points and lines. The geometric objects in this plane can be identified by their mean and standard deviations.

References