Infinitesimal Calculus as an Epistemic Mediator: A commentary on the use of Squares in Elementary Statistical Theory

by Andrew Dynneson at UTEP – and Aaron Alvarez

June 21, 2016

Contents

Abstract

This is a commentary on the use of squares in elementary statistics. One sees an ubiquitous use of squares in statistics, and the analogy of "distance in a statistical sense" is teased out. We conjecture that elementary statistical theory has its roots in classical Calculus, and preserves the notion of two senses described in this paper. We claim that the senses of the differentials dx/dy hold between classical and modern infinitesimal Calculus and show how this sense becomes cashed out in both models. We suggest that the commonalities of both accounts can be used to find an informational realist account of Calculus. This account involves infinitesimal Calculus as acting as an epistemic mediator that expands our conceptual space and enables us to create new measures.

1 Introduction

Statistics is something in-between the historical and the contemporary, however we do not possess the same intuition as when the formuli were obtained. Even though we use these same tools and mechanisms within our own intuitional framework; we do not assimilate all of the historical intuitions. There is no difference between the form given to these concepts, however we do not have the same intuitions about those concepts.

The use of squares is ubiquitous in statistics, here we focus on the use of squares in elementary statistical theory. We look at the formuli in modern statistics, and attempt to explain the intuition and representations of it. In particular, we focus on the differentials dx/dy. We conjecture that the classical foundations of these statistical intuitions are rooted in the Leibnizian sense of the Calculus, which proceeded the modern theory. In this model, space has a specific euclidean magnitude. There are two senses with which we think about distance, and these differ in how they are represented. These two senses have their form preserved in modern calculus. They do not have their referent preserved though. In modern Calculus, they no longer correspond to euclidean magnitudes of space. Instead, Robinson's formalization captures the role of well-formed meaning and conventionality in the process of representation. We will build upon those features to make some suggestions of an informational realist account of infinitesimal Calculus, in the context of statistics.

We will provide a description of classic infinitesimal Calculus and provide a starting point for an informational realist account. Our overall account will focus on the the sense of the infinitesimal as expressed in dx/dy. We will show how the dx/dy express two different senses. The two senses are distinct and are understood differently in the classical and modern Calculus. In the classical or Leibnizean sense, the infinitesimal is understood as capturing a geometrical space that is represented as both a discrete and as a continuous space. Others such as the late classical model of A.L. Cauchy, also made use of the sense of the geometrical.

There are also two senses for which we treat the infinitesimal, dx. What that difference consists of also varies between the classical and modern accounts. The ontological unity in the modern account is less dense, with a specific magnitude, (dx), or Δx, which is relative to the level of precision, ε. The Leibnizian account seems to indicate that there is an infinitesimal, a-priori, that there exists a number dx that is smaller than any positive number. The ontological unity of the object (dx), in the modern case, is less dense because the unity does not confer the infinitesimal as being an object itself, but instead as a way to represent objects, in relation to eachother. Δx fixes a finite quantity, and dx is an infinitesimal quantity. The interchange between the two senses of dx and Δx, and how that is resolved, is the difference between premodern and contemporary Calculus.

We propose to resolve this difference by claiming that Calculus is an epistemic mediator by conferring well-formed structure to measurements. We use the concept of epistemic mediator like Lorenzo Magnani [2002]. He claims that epistemic mediation is the creation of prostheses for human minds. They enable us to create implicit knowledge through a mediated tool.

This gives it ontological unity, as information. In our informational realist account, Calculus parses noise by creating the condition for semantically

meaningful representations of a continuum. Carving discrete elements within a continuum. The mathematical sense of capturing the discrete and the continuous holds across both modern and classical Calculus. The difference is what the infinitesimal Calculus maps onto. The classical infinitesimal's referents are related to the representation of space as being both discrete and as part of a continuum. Modern Calculus does not do this, Robinson's formalization captures the fixing, not of a specific magnitude, but as a re-representation of form. What this means is that Calculus enables the re-representing of a continuum into a syntactically meaningful form, of discrete points.

Contrastingly, we claim that a general feature that holds between the classical and modern accounts of Calculus is that it is a measurement of difference, within a continuum. We suggest that this may be a direction to develop a concept of Caluclus that is rooted in informational realism. This difference is the amount of discrete objects contained within a continuum. You measure the space, and in the process of doing so, you discretize it.

2   Standard Deviation

Let $X = \{x_i\}_{i=1}^n$ be a simple collection of data-points. Let $\overline{x} = \frac{1}{n} \sum x_i$ denote the simple average (mean) of these n points. Neglecting Bessel's Correction, the Standard Deviation in statistics of these data is s, and it is measured as the square-root of the variance:

$$s^2 = \frac{\sum (x_i - \overline{x})^2}{n}$$

If the standard-deviation can be regarded as an analogy for "average distance from the mean," i.e. "average deviation from the mean" in a "statistical sense." The question immediately arises as to why the direct distance measure is not used, i.e. $s_{mod} := \frac{1}{n} \sum |x_i - \overline{x}|$, simply, "the average distance of the points from the mean." The reason is tied to the existence of two different senses underlying the relation of the points to the center.

One rightfully imagines the data plotted along a line, as in:

a)

b)
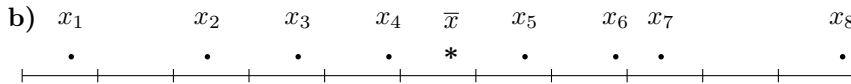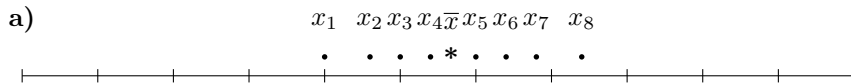
a) Data with "some spread." and b) Data with "more spread" than a).

Why does a statistician not simply measure the distance from a data point to the center of that data: $|x - \overline{x}|$? Then, it is intuitive that one would simply average this distance, and this would be the standard measure of the spread of the data.

The measure captures an Euclidean sense. This can be seen in the classical infinitesimal calculus. For two points $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ in Euclidean Space $\mathbb{R}^n$, the distance measure is denoted d, and is measured as the square-root of the sum of squared component-differences:

$$d^2 = \sum (x_i - y_i)^2$$

Indeed, this is an extension of the Pythagorean Theorem into an arbitrary number of Euclidean dimensions. In particular, $(\mathbb{R}^n, d)$ takes place within Euclidean Space.

The component of comparison is "the center," for each point in the data, as though the square-distance from the point $(x_1, x_2, \ldots, x_n)$ to $(\overline{x}, \overline{x}, \ldots, \overline{x})$ is computed, and averaged. Then, you square root the variance to get the "standard deviation." This is contrasted from our sense of distance, as the points being along the same axis. Here, we have the sample space S and the measure of spread of a simple-sample, s. The process of measurement takes place within the sample-space, (S, s).

When given form, a standard deviation is like a spider, with each point on its own axis:
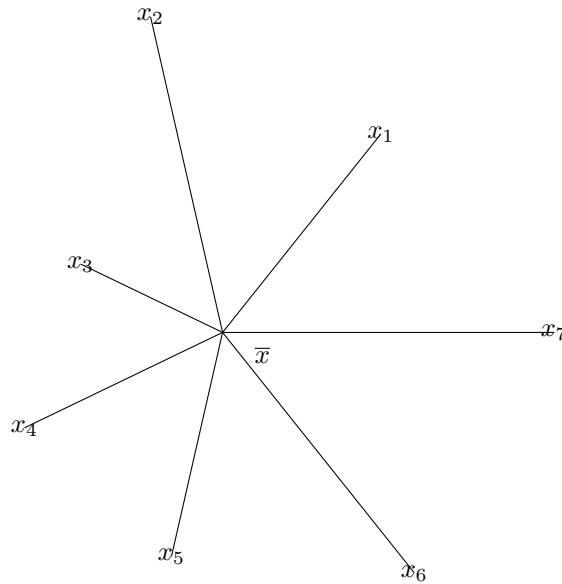
This is the statistical sense: The variance is the average of these distances, of which are similar to that of the Pythagorean theorem. In the extended Pythagorean theorem, one insists on measuring along the legs of the "triangle", rather than simply measuring across the "hypotenuse."

If it is indeed the same object, you should be able to perform that other measurement, s_{mod}. Since we do not practice it, there seems to be a different sense behind the object s, and it is thus a different object than s_{mod}. A catagoricity test may indicate the difference or similarity of objects underlying the senses.

The standard deviation is also a measure of uncertainty (c.f. section on Gaussian).

2.1 Relation of S.D. to Leibniz's Infinitesimals.

We claim that the standard deviation has a different ontological status than Euclidean distance. The average distance of the points from the mean has a different intension than the statistical sense of an infinite continuum of points. In the classical model, the statistical sense captures an infinite density of points within space. There is an infinite amount of space between the points, and hence, a continuous set, of potential measurements, of the points, contained within the space, which is modelled as the sample space. Therefore, the average distance of the points from the mean, as measured through a standard deviation, is a sample space. The statistical sense is restricted by a graphing relationship except when it is transformed via a standard deviation, as a process of measurement. The Euclidean space is broken into a continuum of arbitrary points. These arbitrary points can be grasped with the statistical sense only once the standard deviation is found. These points can be understood as occuring within the interces of a linear continuum. The squaring of these relations is meant to depict the existence and cocurrence with the other non-linear relation of the points from the mean and the points from the center.

Leibniz's intuitions are still built into the modern Calculus [Katz and Sherry: Leibniz's Infinitesimals]. In the classical sense, this infinitesimal Calculus is the epistemic mediator for time and motion within Euclidean space, between points. It epistemically mediates because it enables us to grasp a relation and have epistemic access, through that relation. You are capturing linear time in the classical Calculus through the process. The points do not exist independently seperate from the continuous space. You are analyzing within something, and extrapolating. In the modern Calculus, there is the sense of the neccessity of an epistemic mediator between two senses. In the classical, these two senses have specific magnitudes, which is not the case in modern Calculus.

All versions of infinitesimal Calculus have differentials which capture one quantitive relationship. The difference is what they think the magnitude of the differential is. By changing a differential to zero you change the relationship that the derivative is participating in. By definition, the categoricitiy is different. They measure two different objects, the modern [Robinson's model] is more tractible, and more surveyable than the classical.

When the space is continuous, it is categorically a different object than when it is discrete. In theory, one should be able to form an isomorphism between these categories, in the form of the limit, but there are lingering issues in the language of limits [Kats and Sherry: Laws of Continuity]. The space within which the objects (the points) reside is continuous.

Chopping the space into discrete points via the measurement of the standard deviation makes it the problem of discrete versus continuous variables that come up in modern Calculus. The differential is our way of coping with this disconnect. dx is finite when it imparts conditions for intelligibility (Δx). dx is infinitesimal when it imparts semantic intelligibility (when it yields meaningful information), when it is well-formed, and thus it yields intelligible results. This interchange is the claim that is often assumed. It becomes circular, and "a-priori"-oriented. In integration, Cauchy did not treat ε as a limit-process, but was an evanescent quantity [Beckers 1999 p.10].

The modern infinitesimal is ($\forall \varepsilon : \exists dx : \varepsilon \gneqq dx \gneqq 0$). The pre-modern Leibnizian intuition is that ($\exists dx : \forall \varepsilon : \varepsilon \gneqq dx \gneqq 0$). Pre-modern, there is an a-priori

infinitesimal that neccesarily exists. It is smaller than any positive number. The modern infinitesimal is that dx is relative to which level of precision you are dealing with.

The derivatives are a ratio of two differentials. Two differentials are incomparable (0/0), but are related via finite changes: $\Delta y R \Delta x$. One wonders if the derivative is really a ratio of two differentials, or if it is rather a semi-formal process of limitation.

What people may claim, is that we do not do this type of Calculus any more, we do Robinson's style of Calculus. There was some discussion of the claims made by Leibniz. Berkley [Philosophical Commentaries, p.354][The Analyst, p.116] and Russel [Principles of Mathematics,p.330], disputed them. Schopenhauer [The World as Will and Representation, p.78] disputed non-constructed mathematics itself. This is one place where our contemporary intuition and the pre-Hilbert or earlier intuitions differ, regarding the acceptability of this claim of infinitesimals.

Experts are not in full agreement as to what precisely the Leibnizian intuition is. Robinson himself claimed that he formalized the Leibnizian intuitions, and that the intuitions are the virtually the same (constructed infinitesimal) [Katz and Sherry: Leibniz's Infinitesimals][Robinson 1967].

Ishiguro states explicitly that Leibniz refers to magnitudes in two distinct senses, *entia rationiz* and *entia fictitum.* "Reference to mathematical fictions can be paraphrased into talk about standard mathematical entities." She states that Russel and Leibniz had more in common than Russel realized. [Ishiguro, 1990, p100]

The fiction is produced by how a statement is quantified. For example, Euclidean space is not real in the sense of absolutely real, but is a heuristically useful tool for measuring distances. Another example is imaginary numbers. They are not capable of being eliminated through syncategormatic paraphrase of a structure. [[Did Ishiguro imply this, or are we claiming this, or reference???]].

Kats and Sherry claim that the reading of Leibniz is not syncategormatic, but instead is a way to refer to relations among sets of real values, and are a pure fiction [Leibniz's Infinitesimals, p.13]. Kats claims that Ishiguro's interpretation is distinctly that of a logical fiction [see Bascelli et al: Leibniz vs. Ishiguro 2016]. He points towards how Leibniz states that his infinitesimal violates definition V.4 of Euclid's Elements (the Archimedean property). [1]

Leibniz calls the infinitesimal dx the "terminus" or the "status-transitus." Leibniz literally gives it two different names, for that reason, to indicate that there are two different senses with which we treat dx. Terminus is the "unassignable;" status-transitus is the "assignable," that is why you can assign it status for which to move. Movement implies some assignable status, whereas the terminus is the inassignable position within a continuum. The infinitesimal translates the assignable into the continuous. Leibniz claims that you do not need metaphysical claims to perfom his Calculus [Leibniz, In Cum Prodiisset p.149]. There is a commitment to having two different types of relationships, they should be isomorphic. They have two distinct senses, which given their position in the Calculus, they should be the same, but that they have different mathematical senses.

The differentials are changed to zero because we are not persuing their re-

---

[1] We thank Mikhail Katz for a personal correspondence clarifying this point.

lationships within the non-linear space. A simple sample is a discrete process whereby the continuum is restricted to a discontinuous set, whose average spread is measured by a standard deviation equation. There is a direct correspondence between the senses of dx as terminus and status-transitus, and the statistical senses of a discrete sample and a continuum of the sample space, respectively.

The deviation is a way to render the statistical sense into a spatial-linear sense. This common ground is a reference to both intentions. This is the role of the discreteness itself. One characteristic of the difference between the statistical sense and the geometric sense is that the former refers to a relational property and the latter refers to a non-relational property. The geometric sense is in the form of linearity and the statistical sense is in the form of non-linearity. The standard deviation refers to the system of points within which the sample resides, whereas the geometric distance refers only to the points themselves, without reference to the system.

Kats and Sherry claim that the "*terminus* encompasses the *status transitus*, involving a passage into an assignable entity" [Laws of Continuity, p.6]. Similarly, the standard deviation is a transitory measure, it may be unreal. It is not even really transitory, it is a seperate object, seperate from the objects it is measuring. You can only measure the objects through the standard deviation. You are epistemically boot-strapping access to the object and objects. There may be a disjunct between the points and its measure. The standard deviation has its own ontological identity, seperate from the points. We are thinking that the measure is representing the points, but you would not have an interpretation of the points, in the statistical sense, without the measure.

3 Linear Regression.

Let $X \times Y = \{(x_i, y_i)\}_{i=1}^{n}$ be a collection of data, for which two variables X and Y may be correlated linearly (for example height and hand-size of individuals) [Note that there need not be any linear correlation, but the data may be related, as in the points on a circle]. Then, we want to find the "line of best-fit" through this cloud of points:
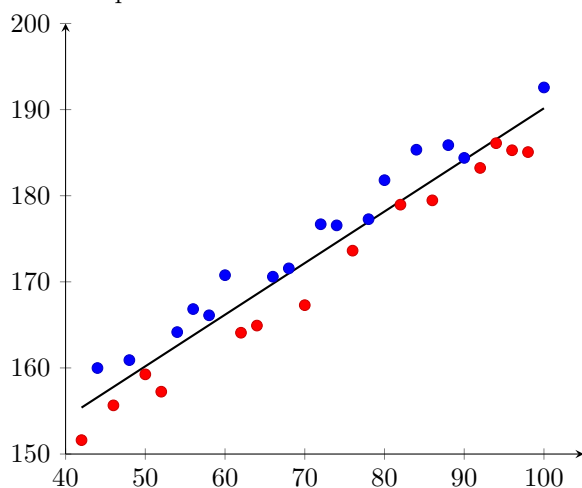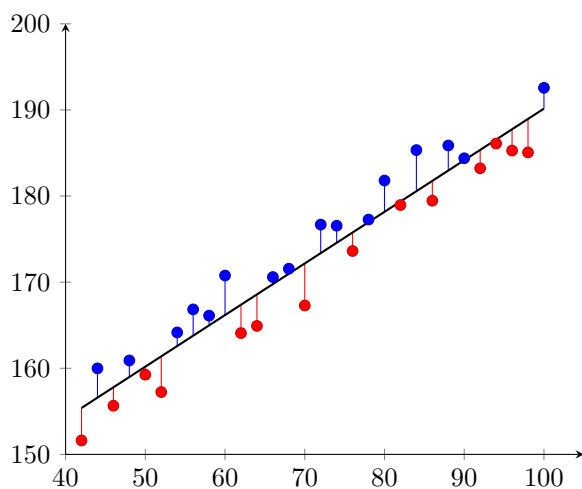
Let $\hat{y} = mx + b$ be this "best line," representing the trend. Even though we have not solved for the "best slope" m and the "best y-intercept" b, we can suppose that they exist. Then, we can find the vertical deviations from each point in the data y_i with that predicted by the trendline $\hat{y}$ at x_i. Once again, it would be intuitive to treat this error measurement in terms of distance;

$$\left[\!\!\left[\begin{array}{c} |y_1 - \hat{y}_1| \\ |y_2 - \hat{y}_2| \\ \vdots \\ |y_n - \hat{y}_n| \end{array} \right]\!\!\right]$$

Then, one would think of the accumulated error $E_{mod} := \sum |y_i - \hat{y}_i|$, and this is the quantity to be minimized. On the contrary, the Error is once again squared:

$$\left[\!\!\left[\begin{array}{c} (y_1 - \hat{y}_1)^2 \\ (y_2 - \hat{y}_2)^2 \\ \vdots \\ (y_n - \hat{y}_n)^2 \end{array} \right]\!\!\right]$$

And the quantity to be minimized is actually the squared error:

$$SSE = \sum (y_i - \hat{y})^2 \tag{1}$$

Indeed, solving for critical points by expanding the square:

$$
\begin{aligned}
SSE &= \sum (y_i - \hat{y}_i)^2 \\
&= \sum (y_i - (mx_i + b))^2 \\
&= \sum y_i^2 - 2y_i(mx_i + b) + (mx_i + b)^2 \\
&= \sum y_i^2 - 2y_i m x_i - 2y_i b + m^2 x_i^2 + 2mx_i b + b^2 \\
&= \sum y_i^2 - 2m \sum x_i y_i - 2b \sum y_i + m^2 \sum x_i^2 + 2mb \sum x_i + \sum b^2 \\
&= n\overline{y^2} - 2mn\overline{xy} - 2bn\overline{y} + m^2 n\overline{x^2} + 2mbn\overline{x} + nb^2
\end{aligned}
$$

Then we differentiate with respect to y-intercepts and slopes respectively and set the derivatives to zero:

$$\left[\frac{\partial SSE}{\partial b} := 0\right] \wedge \left[\frac{\partial SSE}{\partial m} := 0\right] :$$

$$0 =: \frac{\partial SSE}{\partial m} = -2n\overline{xy} + 2mn\overline{x^2} + 2nb\overline{x}$$
$$\rightarrow m\overline{x^2} = \overline{xy} - b\overline{x}$$
$$\rightarrow m = \frac{\overline{xy} - b\overline{x}}{\overline{x^2}}$$

$$0 =: \frac{\partial SSE}{\partial b} = -2n\overline{y} + 2nm\overline{x} + 2nb$$
$$\rightarrow m = \overline{y} - m\overline{x}$$

These yield optimal solutions $b = \overline{y} - m\overline{x}$ and $m = \dfrac{\overline{x \cdot y} - b \cdot \overline{x}}{\overline{x^2}}$.

These solutions are implicit in nature, and contain each-other in their solutions. By holding them both to be optimal, one substitutes b into m, and solves the equation for m:

$$m = \frac{(m\overline{x} - \overline{y})\overline{x} + \overline{xy}}{\overline{x^2}}$$
$$m = \frac{m\overline{x}^2 - \overline{x} \cdot \overline{y} + \overline{xy}}{\overline{x^2}}$$
$$m = m\frac{\overline{x}^2}{\overline{x^2}} + \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2}}$$
$$\rightarrow \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2}} = m\left(1 - \frac{\overline{x}^2}{\overline{x^2}}\right)$$
$$= m\left(\frac{\overline{x^2} - \overline{x}^2}{\overline{x^2}}\right)$$
$$\rightarrow m = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2}$$

Indeed, these are minima, since their second derivatives both have positive parity when evaluated at those critical points:

$$\left[\frac{\partial^2 SSE}{\partial b^2} \equiv +\right] \wedge \left[\frac{\partial^2 SSE}{\partial m^2} \equiv +\right]$$

The solution for b is intuitive, since the point $(\overline{x}, \overline{y})$ can be thought of as "the center of the data," then the result tells us that the "best line" passes through this center, i.e. $(\overline{x}, \overline{y})$ is a solution to the Linear Regression: $\hat{y} = mx + b$, so that $\overline{y} = m\overline{x} + b \rightarrow b = \overline{y} - m\overline{x}$.

The solution for m is not so intuitive.

3.1 SSE Feasibility

One explanation is that the use of squares in statistics is simply utilitarian. For example, distance in its algebraic (absolute-value) form, as $d = |y_i - \hat{y}_i| = \sqrt{(y_i - \hat{y}_i)^2}$, or more vaguely, $\sqrt{f(x_i)}$, where $f(x_i)$ is positive, and let x be some variable of optimization. Then the derivative:

$$\frac{\partial d}{\partial x} = \frac{1}{2\sqrt{f(x_i)}} \cdot \frac{\partial f}{\partial x}$$

.

Summing over the components, and solving for these zeroes is more difficult than optimizing for d^2:

$$\frac{\partial d^2}{\partial x} = \frac{\partial f}{\partial x}$$

Thus, solving for the optimal values of the distance in the squared-geometric sense is inherently easier. Even more vaguely, since square-root is an increasing function, when optimizing $d = \sqrt{\text{Something}}$, it makes sense to optimize the "Something" under the square root, d^2. If you are trying to maximze d, you want to travel as far-right along the curve as you can. On the other hand, if you are trying to minimize d, you want to travel as far left along the curve as you can. If you are trying to minimize |Error|, you can minimize Squared-Error instead.
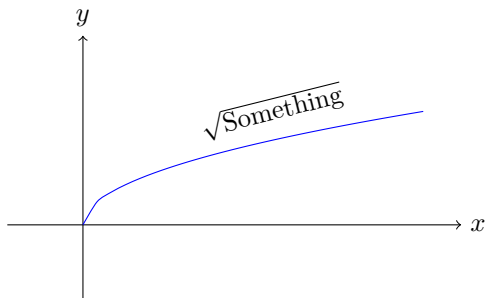
When you perform a regression, you are transforming a series of points into a linear-form. You are rendering it intelligible with the regression, which in theory is neccesarily a part of the continuum, but not neccessarily represented as part of the continuum. Through the process of measuring, you actually transform the cloud of points into a linear form. The regression epistemically mediates the discrete points into a line. You are technically performing a linear regression, however, is the object represented within the continuum of sample-space prior to performing the process? We claim that the calculus acts as a epistemic mediator that creates relations but not objects, because it produces information. Later, we propose reworking Calculus around this question.

You are cutting up the continuum to capture the relevant represented points. The trend : Is the trend represented before performing a regression. The trend already existed, but by cutting it up this way, we enable the representation of the relevant sample. Squaring something expands a fixed point within the continuum, square-rooting something expands something outside the continuum. When you square, you are including something within a single enclosed system (a simple sample), when you square-root you are expanding that object within the larger conceptual framework [Katz and Sherry, Leibniz's Infinitesimals, p.15]. In theory, these are contradictory statements, which is part of the

unspoken conceptual problem. Part of our impetus towards developing a new theory is to avoid this conceptual problem.

Leibniz's Monads are looming in the back-ground. This is the old conceptual view that we are not really interacting with anything in space. Things have an ontology that prevents them from interacting with other things. Objects are ontologically self-enclosed, and can only exhibit their own attributes and properties [Leibniz trans Rescher, Monadology, p.17]. It is very deterministic. It is a base conceptual claim. At the spatial-temporal level we are a single thing or we are not. This is related to classical physics, which is somewhat innaccurate, with the concept of individual particles as indivisible spheres: Individual unities and ontological grounds and their vagueness. In their world, the idea of individual atoms with valences and things moving around does not exist yet.

4 Gaussian Distribution

The Gaussian does not neccessarily correct for the observer effect, however it takes it into account by making the distribution into an inherently non-deterministic system. This differs from classical, strictly deterministic, thinking. This may be a direct response, or this may be unintentional. In this instance, Δx is a measure of uncertainty, it is a standard deviation of x.

4.1 Normalization

The normalization equation is:

$$z(x) = \frac{x - \overline{x}}{\Delta x}$$

This takes into consideration the point itself, as well as its relation with the rest of the data, X. $x - \overline{x}$ measures how far-away from the center the point x is. This can be either positive or negative, depending on if x is more or less than average, respectively. Dividing by the Standard Deviation, Δx, determines how many standard deviations away from the center is x. The measure of a normalized variable within a data-set is thought of as unitless, but is in fact using standard units, or a quantitative measure of number of standard deviations.

4.2 Standard Normal Curve

The standard normal curve is:

$$\frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$$

Here, the coefficient $1/\sqrt{2\pi}$ (which is unfortunately called a normalization coefficient) is because the area under the Gaussian (prior to normalization) is $\sqrt{2\pi}$:

$$\int_{-\infty}^{\infty} e^{\frac{-z^2}{2}} dz = \sqrt{2\pi}$$

Scaling by the total area fixes the total area to be unity, which makes the Gaussian into a probability density distribution.

For many things which are normally distributed, or approximately normally distributed [See Central Limit Theorem], one maps the distribution onto the standard normal curve via $x \mapsto z(x)$, and its probability density distribution is:

$$\frac{1}{\sqrt{2\pi} \cdot \Delta x} \exp\left(-\left(\frac{x - \overline{x}}{\Delta x}\right)^2 / 2\right)$$

4.3   Uncertainty

There is information the moment the measurement is taken. There are measurements recieved, but this is different than the objects with which the means of obtaining the information were thought to measure (position and momenta, with certainty). The moment you do something to obtain information, you create noise. The information does not track truth nor falsity, it tracks the measurement itself. You are checking for consistency. There is change in the world when something is observed, but change implies that there was a position from which the change occured. It is an open non-question as to whether something has a definite position, a-priori. There are no elements within a system that have meaning other than in their relationship with the system. Anything you measure has to be in relation to something else. If an element is independent from everything else, then by definition, it cannot be measured (variant of Info-Computational Naturalism).

This is the difference between the pre-modern and the modern. You must fix the differential dx within a certain tolerance, which is also uncertain, and statistical in nature. You are confirming dx or Δx or Δx, within an existing continuum. dx must be a relation between objects. You cannot have it just be an "infinitesimal." You cannot bootstrap objects out of nothing. You cannot fix the length dx, because it is not represented, with certainty.

Not observing something, it is still expending energy in a thermodynamic sense. It is precisely this reason that the objects are participating in this "relational object", that we can track the actual object, not the representation of it. The entire system is enclosed, but there may not be a way to prove this. This presupposes a realist ontology, that objects exist without being observed. This is an ontological claim. The points are not carved out from the continuum other than points in relation to other points via the standard deviation.

dx must be a finite, Δx. Δx must be strictly greater than some non-zero quantity. There is uncertainty as to what that length (Δx) is. It is statistical. The standard deviation of position is strictly greater than Plank's constant $\hbar$ divided by the uncertainty in momentum, Δp, as a standard deviation. dx must be larger than uncertainty, for Calculus to work properly (or even approximately) at the micro-scales. In the macro-scale, dx can be regarded as nearly zero. In the micro-scale, dx must be regarded as some finite quantity, statistically modelled. At the macro-scale, a reformulation is almost unncesseary, on the contrary, in the quantum regime, a reformulation is neccessary.

$$dx \overset{\alpha}{>} \Delta x \overset{\alpha}{>} \frac{\hbar}{2\Delta P}$$

Δx is the distance with which the concavity of the Gaussian changes, half the length of the interval for which the distribution is concave-down. Outside of $2\Delta x$, whose center is $\overline{x}$, the concavity of the distribution is up.

5 Correlation
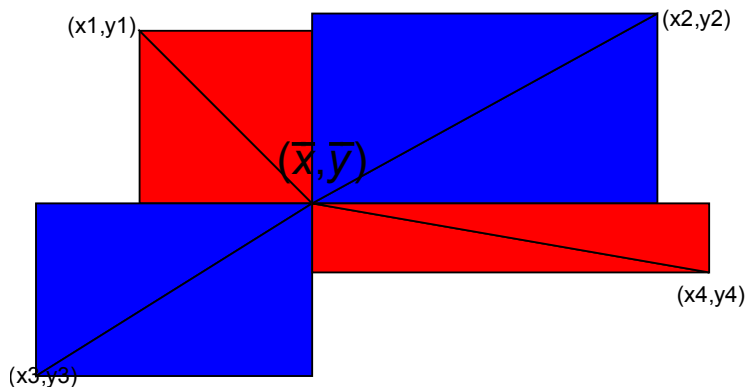
Once one realizes that the data for the set X, Y can be plotted on the $x - y$ axes, then covariance has a wonderful geometric interpretation. One visualizes $(\overline{x}, \overline{y})$ as the center of the cloud of points.

If the point (x, y) is to the right of $(\overline{x}, \overline{y})$, then $x - \overline{x}$ will be a positive distance. If (x, y) is to the left of $(\overline{x}, \overline{y})$, then $x - \overline{x}$ will be a negative distance.
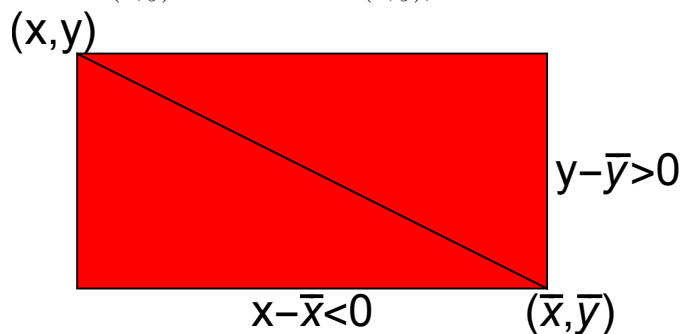
On the other hand, if (x, y) is above $(\overline{x}, \overline{y})$, then $y - \overline{y}$ will be a positive distance. If (x, y) is below $(\overline{x}, \overline{y})$, then $y - \overline{y}$ will be a negative distance.

Drawing a line segment from $(\overline{x}, \overline{y})$ to (x, y), and drawing a rectangle will allow us to visualize the covariance. Suppose that the data is positively accociated. The area of the rectangle is the base times the height, $(x - \overline{x}) \cdot (y - \overline{y})$.

If the point (x, y) is either above and to the right, or below and to the left of $(\overline{x}, \overline{y})$, then this is positive area (i.e., a positive base times a positive height or a negative base times a negative height), and the point adds to the positive accociation of the points, the closer to a square this area is, the more so will the area be maximized.

On the other hand, if the point (x, y) is either above and to the left (negative base times positive height), or below and to the right (positive base times negative height) of $(\overline{x}, \overline{y})$, then this rectangle has negative area, and takes away from the positive accociation of the data.

The covariance then, is sort of the average "linear-ness" of the data:

$$\text{Cov(X,Y)} = \frac{\sum (x_i - \overline{x}) \cdot (y_i - \overline{y})}{n}$$

The covariance captures the proximity of the points to the form of the line, and thus the property of linearity. The variance can also be thought of as the

covariance of X with itself, which only forms squares, recall:

$$\text{Var}(X) = s^2 = \frac{\sum (x_i - \overline{x})^2}{n}$$

Next, the covariance depends on the one hand on how linear the data is, and it also depends on your units of distance. Recall that the data from set X can be "normalized" into a consistant system of measure, across all possible units, by determining precisely how many standard deviations away from the mean a particular point is. For example, using the normalization function from the Gaussian distribution:

$$z = \frac{x - \overline{x}}{s}$$

This allowed us to look-up the values for any variable X by using a single chart, which is for the normalized variable Z. We can use the exact same idea for covariance by normalizing <u>both</u> dimensions X and Y, which will yield:

$$\left(\frac{x - \overline{x}}{s(x)} \right) \cdot \left(\frac{y - \overline{y}}{s(y)} \right)$$

This will require us to find the standard deviations for both the directions X as well as Y:

$$s(x) = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n}}$$

Since we normalized the variables, this now provides a <u>consistent</u> way of measuring linearity, which is homogeneous and unit-less, which we shall call the correlation measure:

$$r = \frac{1}{n} \sum \left(\frac{x_i - \overline{x}}{s(x)} \right) \cdot \left(\frac{y_i - \overline{y}}{s(y)} \right) = \frac{1}{s(x) \cdot s(y)} \frac{\sum (x_i - \overline{x}) \cdot (y_i - \overline{y})}{n} = \frac{\text{Cov}(X, Y)}{s(x) \cdot s(y)}$$

We are enabling the line to form there. r is between -1 and 1, as a measure of strength of the linear relationship. The slope of the regression is positive or negative. In order to render the geometric to the statistical sense, we have to create discrete points within a coordinate space from which we can create a line. Each discrete point is actually part of a continuum, but for utilitarian reasons, we limit it to translating it into a statistical sense.

5.1   Regression Again

We have already shown that $(\overline{x}, \overline{y})$ is a solution to the regression. We have also solved for a form of the optimal slope of the regression, m. Distance and squaring can be done irrespective of direction. Thus we can switch: $(y_i - \hat{y}_i)^2 = (\hat{y}_i - y_i)^2$ within the SSE. Then, with these facts, we use the point-slope form of a line: $\hat{y} - \overline{y} = m(x - \overline{x})$. With an eye towards correlation, we find another form for the slope m, by differentiating *before* expanding the square (via the chain-rule). This yields:

$$m = r \cdot \frac{s(y)}{s(x)}$$

$$SSE = \sum [m(x_i - \overline{x}) + \overline{y} - y_i]^2$$

$$0 =: \frac{\partial SSE}{\partial m} = 2 \sum [m(x_i - \overline{x}) + \overline{y} - y_i](x_i - \overline{x})$$

$$\rightarrow \sum y_i(x_i - \overline{x}) = \sum [m(x_i - \overline{x}) + \overline{y}](x_i - \overline{x})$$

$$= m \sum (x_i - \overline{x})^2 + \sum \overline{y}(x_i - \overline{x})$$

$$\rightarrow m = \frac{\sum y_i(x_i - \overline{x}) - \sum \overline{y}(x_i - \overline{x})}{\sum (x_i - \overline{x})^2}$$

$$= \frac{\sum (y_i - \overline{y})(x_i - \overline{x})}{\sum (x_i - \overline{x})^2}$$

$$= \frac{\sum (y_i - \overline{y})(x_i - \overline{x})}{ns^2(x)}$$

$$= \frac{1}{n} \left[\sum \frac{(x_i - \overline{x})}{s(x)} \cdot \frac{(y_i - \overline{y})}{s(y)} \right] \cdot \frac{s(y)}{s(x)}$$

$$m = r \cdot \frac{s(y)}{s(x)}$$

[Zaiontz]

In theory, these values for the slope should be equal, their distinct forms arise as to which form from the SSE (which you differentiate), and also which form from the line-equation that you begin your derivation.

6 A Proposed Information-Calculus

Infinitesimal Calculus involves the fixing of discrete points within a relevant continuum. The two senses of the infinitesimal dx indicate an informational relation. In contrast the continuum, originally understood in a geometrical sense, is raw data without any context and lacks a well formed structure. It is through the infinitesimal that a semantically well-formed structure is conferred to the data. The Gaussian suggests a reformulation of the sense of the infinitesimal, not as a finite magnitude; which is intuitively, neither big nor small, neither spatial nor physical but rather the condition of having a semantically-well formed structure. The mathematical relationship appears as the fixing of dx. At the macro scale, dx is near zero, and the at the micro, dx is a finite quantity, that is statistically modeled. It is through both conditions that form is conferred upon data. Both relations fix and confer upon an infinite continuum, the status of being capable, of being discrete.

The discrete relationship is a requirement for meaningful-representations of the continuum. Rather than inflating the infinitesimal into an ontological object, we create the conditions for measurement by fixing a convention. This convention allows for representation of the continuum in a way that is meaningful given surveyability conditions.

Robinson's formalization captures the conventionalization that enables Calculus to acquire a meaningful and semantically informed relationship. The se-

mantically well-formed structure is configured in a statistical sense through the standard deviation. This enables us to piece together points in a continuum. This carving out of a relation is important because from it emerges the informational relationship that can then be mapped onto different magnitudes. In this way, Calculus acts as an epistemic mediator, enabling new measurements.

This is the break from Leibniz's dx because there is no ontological magnitude associated with the infintismal. This means it does not directly map onto geometrical relations. Instead the dx is a conventionalized way to create fixed magnitudes in a setting which is informed by context. By fixed we mean that meaning is parsed from the data, in a Statistical Sense.

In the informational view, there is no infinitesimal at the quantum scale, because it is not semantically meaningful to represent the infinitesimal. This is the case, in either sense of both the classical and modern infinitesimal within Calculus. The reason why this is, is that the infinitesimal prevents the Calculus from becoming an endless recursive/bottomless floor of objects. This is for epistemic reasons, you cannot simply create these objects, foundationally, or foundationlessly. The other reason is that the proposed informational dx is relative to the energy level, which is discritized when it is measured.

The Informational Calculus is what can be derived safely from both versions, without inflating or reducing the infinitesimal.

7 Conclusion

Infinitesimal Calculus intersects with the issue of realism and anti-realism in mathematics. Statistics and Calculus develop from the same intuitions, however they represent them differently. In theory we are utilizing an actual transformation as an object. There may be a disconnect in how we intuit this object. In classical infinitesimal Calculus, we understand this object as having ontological weight. In modern infinitesimal Calculus we understand this as being a convention. To develop calculus further, we will have to think of this object in terms of a convention, and the meaning it gives to what is measured. We need to think of dx as a relation of epistemic mediation. This relation becomes an object when it imparts a semantically meaningful structure on what is measured. This avoids inflating or reducing dx.

This relation is a process that is not meant to be countable but instead as sampling of phenomena.Statistics is a part of the sceintific method in virtue of its ability to act as an emprical measure, through its imparting of structure. There is no general relation that determines what to induct here. There is no current principle or axiom forcing Euclidean transformations in this way. This theory which is supposed to be grounded in the geometric sense is not really grounded in the geometric sense.

Sampling is its own mathematical category. What remains to be tested, is if that category maintains isomorphism with the objects it represents. We recommend a test of categoricity.

We suggest that an informational realist account of Calculus should be used as a straddling point between the classical and modern accounts of infinitesimal Calculus. The reason why is because it avoids any of the vices of deflation or inflation. This informationally informed account of Calculus works with Robinson's formalization, but with conceptual relations informed by informational

realism. It sees the advantage of calculus in its imparting of sense to what is measured.The two senses of dx/dy in both the classical and modern account are preserved. It takes the conventionalization and formalism of Robinson's account to be important. In it, the two senses expressed in dx/dy are expressions of the coherency of something to be understood as discrete or as a continuum.

Infinitesimal Calculus understood in the modern and informational realist sense is tool and not beholden solely to geometrical magnitudes. It is tool that acts as a epistemic mediator as a representative tool. It enables representation of various magnitudes and enables them to be semantically meaningful. [Magnani and Dossena 2005]

8   Sources Cited

1. Beckers. Definitely Infinitesimal: Foundations in the Calculus in the Netherlands, 1840-1870. *Taylors and Francis: Annals of Science* (1999).

2. Berkeley. Philosophical Commentaries: Generally Called the Commonplace Book. *T. Nelson and Sons limited* (1944).

3. Berkeley. The Analyst. *Springer Science and Business* p.116 (2012).

4. Bascelli et al. Leibniz vs. Ishiguro: Closing a Quarter-Century of Syncategoremania. *HOPOS: The Journal of the International Society for the History of Philosophy of Science.* The University of Chicago Press (2016).

5. Chiribella, Giulio and Scandolo, Carlo Maria. Conservation of information and the foundations of quantum mechanics. *EDP Sciences* (2016).

6. Ehrlich, Philip. The Rise of non-Archimedean Mathematics and the Roots of a Misconception I: The Emergence of non-Archimedean Systems of Magnitudes. Arch. Hist. Exact Sci. 60 (2006) 1–121.

7. Goldin, Owen. Aristotle, the Pythagoreans, and Structural Realism. *Review of Metaphysics.*

8. Ishiguro, Hide'. Leibniz's Philosophy of Logic and Language. *Cambridge University Press.* 2nd edition (1990).

9. Katz, Mikhail and Sherry, David. Leibniz's Infinitesimals: Their Fictionality, their Modern Implementations, and their Foes from Berkeley to Russel and Beyond *Springer Netherlands: Erkenntnis.* Vol.78, No.3 pp.571-625 (2013).

10. Katz, Mikhail and Sherry, David. Leibniz's Laws of Continuity and Homogeneity*Notices of the American Mathematical Society.* Vol. 59, No. 11 pp.1550-1558 (2012).

11. Leibniz, Cum Prodiisset. mss "Cum prodiisset atque in- crebuisset Analysis mea infinitesimalis . . . " in Gerhardt 1846, pp. 39–50. `https://books.google.co.il/books?id=UOM3AAAAMAAJ&source=gbs_navlinks_s&hl=iw`

12. Leibniz, Gottfried Wilhelm, and Nicholas Rescher. G.W. Leibniz's Monadology: An Edition for Students. Pittsburgh, PA: U of Pittsburgh, 1991.

13. Magnani, Lorenzo. Epistemic Mediators and Model-Based Discovery in Science. *Model-Based Reasoning* pp.305-329 (2002).

14. Magnani, Lorenzo and Dossena, Riccardo. Perceiving the Infinite and the Infinitesimal World: Unveiling and Optical Diagrams in the Construction of Mathematical Concepts Vol.10 Isssue 1 pp. 7-23 *Foundations of Science* (March 2005).

15. Robinson. The Metaphysics of the Calculus. *Studies in Logic and the Foundations of Mathematics* 47:28–46 (1967).

16. Russell, Principles of Mathematics Vol.1 *University Press* p.330 (1903).

17. Schopenhauer, The World as Will and Representation Vol 1. Translated by Janaway, *Cambridge University Press* p.78. (Oct. 28, 2010).

18. Zaiontz, Charles. Real Statistics Using Excel. (2013-2016) `http://www.real-statistics.com/regression/least-squares-method/least-squares-method-detailed/`