Title
The Gauss-Pearson Decomposition and the Link Between Classical Philosophy and Modern Statistics

Author
Ekström, Joakim

Publication Date
2011-12-01

THE GAUSS-PEARSON DECOMPOSITION AND THE LINK BETWEEN CLASSICAL PHILOSOPHY AND MODERN STATISTICS

JOAKIM EKSTRÖM

Abstract. This article discusses the mathematical and philosophical premises of Gauss' and Pearson's methods for advancing science through empirical observation. At the core is the notion that an observation can be decomposed into an ideal part and a random part, and that the relation between the observation and its two parts is known. The premise, named the Gauss-Pearson decomposition, is formalized in an algebraic framework that is as general as possible. Examples illustrate how the decomposition fits naturally into the body of statistical methodology and how it can facilitate understanding of statistical methods.

1. Introduction

The present article discusses the notion that an observation can be decomposed into two parts: one part that represents the true nature of the phenomenon and one part that represents the observational error. The decomposition, in the article referred to as the *Gauss-Pearson decomposition*, is discussed in terms of its history and origins, and it is put into a modern algebraic framework.

An early illustrative example that many are familiar with is the regression analysis model, where the observation y is presumed to be a sum of a function value $f(x)$ and a random variable u, i.e.

$$y = f(x) + u.$$

The observation, y, is thus decomposed into a true value, $f(x)$, and a random error, u. Indeed, this decomposition is presumed in most statistical analyses, often without mention. The present article discusses reasons, both scientific and pedagogical, why the decomposition, rather than being suppressed, should be embraced and elevated to the position of a pillar of statistics.

This article is a contribution to the rigor of statistics and the clarity of its premises for the benefit of the methods and all who use them. The article discusses the history and origins of the Gauss-Pearson decomposition, treats its algebra in detail, and illustrates through examples how the decomposition fits naturally into the body of statistical methodology.

2. Origins and early history

The notion of a phenomenon having a true nature can be traced back at least to Plato's theory of forms. In fact, Taylor (1936), Grube (1935) and Irwin (1999) all argue that the notion is even older and was used by Socrates and possibly as early as the Milesian school of philosophy in the 6^{th} century B.C. However, as White (1976) points out, history has attributed the theory of forms to Plato, and speculation on earlier origins is difficult due to lack of primary sources.

Plato was a Greek scientist, principally a mathematician, who was born 428-27 B.C. into a distinguished family in the Athenian city-state (Taylor, 1936). He was a prolific writer and the first author of classical antiquity whose works have survived in their entirety. Besides the mathematical work, Plato was active in the political arena and wrote extensively on topics such as public policy, education, and morality. According to Taylor (1936), Plato expressed the feeling that his most valuable contribution was the founding of the original Academy, the embryo of modern day academia, circa 387 B.C. Further, Plato bears the distinction of having been continuously cited by members of the scientific community, among them notable authors such as Aristotle, Galilei (1632), Arnauld & Nicole (1662), Bernoulli (1713), Newton (1726) and Pearson (1911). For these and many other reasons, Plato is a natural candidate for the honorary title 'Father of the Scientific Community'.

Plato did not present the theory of forms as a complete exposition. Rather, it emerges piece by piece throughout his works. In fact, the name theory of forms was created for discussion of Plato's works by subsequent scholars (Ross, 1951). A condensed and simplified, though for the present discussion sufficient, version is given below.

Postulate (Plato's theory of forms)**.**

(i) Each and every phenomenon has an underlying true nature, a Platonic ideal form.
(ii) The senses are inherently unreliable and observations inevitably subject to flux.

As a result, Plato argued, observations cannot, in any reliable or meaningful way, be used for expanding knowledge and advancing science. In an understatement, one can say that Plato would not be particularly fond of the modern statistical science. Rather, Plato favored logical deduction within an axiomatic framework as the only manner in which knowledge can be reliably gained. *Euclid's Elements* is a magnificent early example of deductive reasoning. Besides Plato's own field of mathematics, the deductive method would be successfully employed in physics, see, e.g., Archimedes (c. 287-212 B.C.), Galilei (1638), Newton (1726) and Einstein (1905).

In the historical context, it is worth noting that the measurement instruments available during Plato's lifetime were quite rudimentary. Consider, for example, time: in the early antiquity it would have been difficult to measure the physical quantity with an accuracy of one second, while the best modern day atomic clocks have an accuracy greater than 10^{-15} second. Other physical quantities that can now be measured with extraordinary accuracy are, e.g., weight, length, radiation and electric charge. It is possible that Plato, who at heart was a practical realist (White, 1976), would have been persuaded about the usefulness of these measurement instruments and the value that they lend to their observations. In the present discussion, however, Plato serves as the indispensable antagonist.

Galilei (1632), a work on astronomy with many references to Plato, accepts the theory of forms but makes one crucial addition. Galilei argues that the following, which in this article is formalized into a postulate, should hold.

Postulate (Galilei's observational error postulate)**.** *In addition to Plato's theory of forms, if the observers are clever men, intelligent and dextrous in handling their instruments, then positive and negative observational errors are equally probable, and small errors are more probable than large errors.*

Galilei's postulate is remarkable in that it suggests that the observational errors have a *probability distribution* and that some knowledge about the distribution exists. Although the details are not entirely clear, Galilei (1632) applies the postulated knowledge for the purpose of mitigating the flux and lack of reliability which is inevitable under Plato's theory of forms. The fact that Galilei's work was published before Pascal studied the arithmetic triangle (1665), commonly considered the historic starting point of probability theory (see, e.g., Hald, 1990), makes the postulate even more remarkable.

The first instance in which it was formally suggested that probability should be used as a vehicle to bridge deductive and inductive methods was Jacob Bernoulli's *Ars Conjectandi* (1713). If a deductively falsified statement is represented by the number zero and the deductively proven statement represented by the number one, Bernoulli argues, then the continuum of probability interpolates between these two, linking deductive and inductive methods. To determine the statement's probability, which in itself is an ideal form, Bernoulli suggests that it should be empirically determined through *repeated observation*. In arguing the merits of this proposal, Bernoulli proves his result which has become known as the law of large numbers. The law of large numbers carries a profound impact on the present discussion, and a modern version is stated below.

Theorem 1. *Suppose that μ is a real number and, for $i = 1, 2, \ldots$, that $x_i = \mu + u_i$ are noisy observations of the same, where the u_i's are independent random variables with mean zero and bounded variance. Then, the arithmetic mean converges in probability to μ as the sample size goes to infinity.*

Proof. An application of Chebyshev's inequality yields the result. □

Consequently, in a large number of observations the flux which the observations inevitably are subject to will on average *cancel out*, leaving the ideal form of the observed phenomenon in bare sight. Hence the law of large numbers, which is a deductively derived result, yields a path to arbitrarily accurate approximation of the ideal form though empirical observation. The result is that the second part of Plato's theory of forms can be mitigated entirely. No matter how inaccurate the measurement instrument is, the approximation can made arbitrarily accurate by simply repeating the observation sufficiently many times. It is crucial, though, that the measurement instrument is well-calibrated, which in retrospect explains the importance of the conditional premise of Galilei's observational error postulate. Notably, the last paragraph of Bernoulli (1713), a work which Bernoulli died finishing, reaches out to Plato and discusses how he would have reacted had he been aware of Bernoulli's results.

3. The decomposition

In the second book of *Theoria Motus* (1809), Gauss set out to derive the most probable Kepler orbit given observations of a heavenly body. For this purpose, two statistical pieces were needed; the density criterion first proposed by Daniel Bernoulli (1778), and the decomposition that is discussed in the present article.

The decomposition, which is also discussed at length in Gauss (1821b) and Gauss (1821a), is based on the premise that each observation consists of two parts: one corresponding to the Platonic ideal form and one corresponding to the observational error. The premise is formalized as follows.

Postulate (Gauss' observation postulate)**.** *Each observation consists of two parts: the true value of the phenomenon that has been observed, and an accumulation of random errors that is unavoidable and cannot be eliminated.*

Note that Gauss' observation postulate presumes the existence of such parts, i.e. it presumes the theory of forms. Gauss does not refer to Plato by name, but uses theory of forms terminology when explaining his postulate, for example 'No observation in the world of senses can have absolute accuracy' (Gauss, 1821a). The sentence is in essence the second part of Plato's theory of forms as given in Section 2, and the term world of senses is typical of the theory of forms terminology.

Pearson (1900), which introduces the statistical hypothesis test, uses the same notions albeit with a slightly different terminology. Dealing with frequencies in particular, Pearson interchangeably uses the terms theoretical frequency, population frequency, and theoretical population frequency in place of Gauss' true frequency. With respect to the other part, both Gauss and Pearson use the term random error. In the present article, the two parts are referred to as the ideal part and the random part, respectively.

Both Gauss (1809) and Pearson (1900), moreover, contain an explicit expression for the observation as a function of the ideal and random parts, detailing how the three are related to each other. The expression is the sum

$$x = \mu + u,$$

where x is the observation, μ the ideal part, and u the random part. Both Gauss and Pearson further assume that the random part is normally distributed with mean zero. Using this decomposition and the density criterion, Gauss derives the most probable Kepler orbit and, of historical note, also shows that it is given by the method of least squares. Pearson derives the statistical hypothesis test and shows that it can be defined through what Pearson calls the chi-distance.

With respect to Plato's theory of forms, the ideal part corresponds to the world of forms, the observation corresponds to the world of senses, and the random part constitutes the discrepancy between those two. Mathematically, the core property of the decomposition is that the ideal and random parts can be algebraically solved for. This property, in combination with an assumed probability distribution of the random part, allows for the most probable ideal part, given observations, to be derived.

The decomposition, named the Gauss-Pearson decomposition, is the detailed premise of Gauss' and Pearson's methods. Their two methods for generating and testing hypotheses through observations, the method of least squares and the chi-square test, respectively, as well as further-developed variants, are currently some of the most commonly used methods in science. Since the methods fundamentally rest of Plato's theory of forms, they also represent what is possibly the most concrete link between the father of the scientific community and its modern day members.

4. Algebraic framework

The present section aims to develop the algebraic details of the Gauss-Pearson decomposition. To achieve the greatest level of versatility, the decomposition is dealt with in the most general setting possible. For this reason, suppose that the observation, x, is an element of an arbitrary abstract set, denoted $\mathbb{X}$. Because the observation is an approximation of the ideal part, using Gauss' wording, it is reasonable that the ideal part, μ, is similar to the observation in nature and consequently an element of the same set, $\mathbb{X}$.

Fundamental to the Gauss-Pearson decomposition is the property that random parts of repeated observations can cancel out, thus yielding a better approximation of the ideal part (cf. the law of large numbers). Algebraically, the random part, u, must therefore be an element of a group. Consequently, the mapping $(\mu, u) \mapsto x$ is a right group action, acting on $\mathbb{X}$ by permutation. However, in order to keep the algebra within reasonable limits it is sensible to simply equip the set $\mathbb{X}$ with a group structure and let the random part be an element of the group $\mathbb{X}$. This condition conforms with the tradition of Gauss and Pearson, and is seldom restrictive. The group $\mathbb{X}$ then acts on itself, and the mapping $(\mu, u) \mapsto x$ is a binary operation. If $\mathbb{X}$ is not closed under the binary operation, it is natural to simply let the group be generated by, and then quickly renamed to, $\mathbb{X}$.

Continuing in the tradition of Gauss and Pearson, the binary operation is denoted additively, i.e. $x = \mu + u$, despite the fact that the binary operation need not be commutative. Furthermore, for the random part to be a random variable in Kolmogorov's sense, $\mathbb{X}$ must have a topology. In combination with Gauss' observation postulate, the reasoning yields the following definition.

Definition (Gauss-Pearson decomposition)**.** *Let $x \in \mathbb{X}$ be an observation and suppose that Gauss' observation postulate holds. Suppose further that $\mathbb{X}$ is a topological space, Ω a probability space, and that $\mu \in \mathbb{X}$ and $u : \Omega \to \mathbb{X}$ are the ideal and random parts, respectively, of x. The Gauss-Pearson decomposition is the identity*

$$x = \mu + u,$$

where $+$ is a binary operation such that $(\mathbb{X}, +)$ is a group.

Remark 1. The fact that $\mathbb{X}$ has an identity element makes the decomposition most unrestrictive. For example, if the random part attains the identity element with probability one, then the observation equals the ideal part with probability one. If, on the other hand, the ideal part equals the identity element, then the observation equals an accumulation of random errors. The latter could possibly be the case if the observation is, say, a registration of pure noise from outer space with no interpretable meaning.

Remark 2. Since $\mathbb{X}$ is a group, there are unique algebraical solutions for the ideal and random parts: $\mu = x - u$ and $u = -\mu + x$ respectively. Solving for the random part is essential to many statistical methods, not least the methods of Gauss (1809) and Pearson (1900).

Remark 3. Applying the expectation operator to the Gauss-Pearson decomposition yields the expression $E(x) = \mu + E(u)$. Gauss (1821b) contains a comprehensive discussion on whether the expectation of the random part necessarily must equal the identity element. Specifically, Gauss' assumptions are such that the expectation, median, and mode are all equal. Gauss argues that if the expectation of the random part were not the identity element, then it should have been eliminated already. Like Galilei (1632), Gauss (1821b) reasons that the observers must be presumed competent and able to calibrate their measurement instruments properly. The expectation of the random part is often referred to by the term *bias*.

5. Examples

With the intent of illustrating how the Gauss-Pearson decomposition fits naturally into the body of statistical methodology and how the decomposition can facilitate understanding of the methods' purposes, some examples are provided.

Example 1 (Method of the arithmetic mean). The arithmetic mean is a statistical method which in its simplicity well-illustrates the fundamentals of the advancement of science through empirical observation.

Suppose $x_1, x_2, \ldots$ are real-valued observations of some phenomenon with Gauss-Pearson decomposition $x_i = \mu_i + u_i$, $i = 1, 2 \ldots$, where $\mathbb{X} = \mathbb{R}$ is equipped with its standard addition and topology. If the observations' ideal parts $\mu_1, \mu_2, \ldots$ are all equal and the random parts $u_1, u_2, \ldots$ are independent and identically distributed, then it holds by the strong law of large numbers that for the limit of the arithmetic mean

$$\frac{1}{n}\sum_{i=1}^{n} x_i \to \mu + E(u) \quad \textit{with prob. one} \text{ as } n \to \infty,$$

and thus, presuming that $E(u)$ exists and is known, the result in theory provides a way to uncover the ideal part of the observed phenomenon. In particular, if $E(u) = 0$ then with probability one $\mu = \lim_n n^{-1} \sum_{i=1}^{n} x_i$, i.e. the arithmetic mean approximates the ideal part arbitrarily well if only the sample size, n, is sufficiently large. Note also that the Gauss-Pearson decomposition is a premise of Theorem 1.

Figuratively, this statistical method provides scientists with a metaphorical pair of binoculars with which they can look into Plato's world of forms and find the answers to mysteries such as those of cancer, Alzheimer's disease, obesity, and so on. In essence, the researcher only has to gather the observations. This remarkable method, while simple, is at present widely employed in scientific research worldwide, not least in the health sciences.

Example 2 (DNA-bases). For an example where observations are not real-valued, let the symbols A, G, C and T denote the four DNA-bases, and let $\mathbb{X} = \{A, G, C, T\}$ be equipped with the discrete topology. An element of the product space $\mathbb{X}^m$ is a sequence of bases of length m; in particular the human genome can be represented by an element of such a product space with m equalling circa three billion. A suitable binary operation may for example be one such that $\mathbb{X}$ is isomorphic to $\mathbb{Z}_4$, i.e. $\mathbb{X}$ is the cyclic group of order 4. For example, the binary operation $+$ may be defined such that its table is as follows.

$\mathbb{X}$	A	G	C	T
A	A	G	C	T
G	G	C	T	A
C	C	T	A	G
T	T	A	G	C

Thus an observation $x \in \mathbb{X}$ has a Gauss-Pearson decomposition $x = \mu + u$, where the ideal part, μ, is the element representing the true DNA-base, and the random part, u, equals $-\mu + x$. The discrete distance, $d(x, y) = 0$ if $x = y$ and $d(x, y) = 1$ if $x \neq y$, may well be suitable for estimation and inference through method of least distance.

Elements of $\mathbb{X}^m$, i.e. DNA sequences of length m, can be added by component-wise addition. A similar approach can be used for most so-called categorical random variables.

Example 3 (Pearson's statistical hypothesis test). Pearson (1900) defined a series of concepts which in combination constitute the statistical hypothesis test. At the core is the Gauss-Pearson decomposition. Suppose $x_1, \ldots, x_n$ are real-valued observations of the phenomenon of interest, with Gauss-Pearson decomposition $x_i = \mu_i + u_i$, $i = 1, \ldots, n$, where $\mathbb{X} = \mathbb{R}$ is equipped with its standard addition and topology. For convenience let the arrow accent denote the sequence of length n, e.g. $\vec{x} = (x_1, \ldots, x_n)$, and note that the Gauss-Pearson decomposition can be written $\vec{x} = \vec{\mu} + \vec{u}$.

Next, consider the hypothesis that the ideal part of the observed phenomenon equals $\vec{v}$, for some given sequence $\vec{v} \in \mathbb{X}^n$. The hypothesis yields a second representation of the observation, $\vec{x} = \vec{v} + \vec{e}$, where $\vec{e}$ is the representation residual. Thus there are two expressions for the observation,

$$\begin{cases} \vec{x} = \vec{\mu} + \vec{u}, & \text{(the Gauss-Pearson decomposition)} \\ \vec{x} = \vec{v} + \vec{e}. & \text{(the hypothesis representation)} \end{cases}$$

Since $\mathbb{X}^n$ is a group under component-wise addition, algebraic manipulations of the expressions yield

$$\vec{\mu} = \vec{v} \ \textit{implies} \ \vec{e} \sim \mathcal{L}(\vec{u}), \quad \text{and} \quad \vec{e} \nsim \mathcal{L}(\vec{u}) \ \textit{implies} \ \vec{\mu} \neq \vec{v},$$

where $\mathcal{L}(\vec{u})$ denotes the probability distribution of the random variable $\vec{u}$. Consequently, the hypothesis can be falsified through falsification of the proposition $\vec{e} \sim \mathcal{L}(\vec{u})$. Pearson (1900) proposed a certain probabilistic form of falsification which has been termed Pearson-falsification. Evaluation of the probability of the proposition $\vec{e} \sim \mathcal{L}(\vec{u})$ is done under Pearson's distance criterion, which uses Pearson's chi-distance, a now obsolete special case of the Mahalanobis distance.

Under the assumption that the random parts are independent and normally distributed with mean zero and variances $\sigma_1^2 \ldots, \sigma_n^2$, the chi-distance satisfies

$$d(\vec{e}, 0)^2 = \sum_{i=1}^{n} \left(\frac{e_i}{\sigma_i} \right)^2 = \sum_{i=1}^{n} \left(\frac{x_i - \nu_i}{\sigma_i} \right)^2,$$

which is familiar to many as the so-called chi-square statistic. Pearson (1900) is discussed at length in Ekström (2011).

Example 4 (Gauss' most probable Kepler orbit). As has been mentioned in Section 3, Gauss (1809) set out to derive the most probable Kepler orbit given observations of a heavenly body. Let $\{f_\theta\}_{\theta \in \Theta}$ be the set of Kepler orbits, each of which map a point in time t to a pair of angles (y_1, y_2) on the celestial sphere. By letting $x = (t, y_1, y_2)$, Gauss' Gauss-Pearson decomposition can be written

$$x = \mu + u,$$

where μ is the ideal part, i.e. the true time and the true angles for the observation, and u the random part, an accumulation of random errors from various sources.

Gauss assumed that the measurement error in the time variable was negligible, while the measurement errors in the angles were statistically independent and normally distributed with mean zero. It follows that the best projection, in terms of the density criterion (Bernoulli, 1778), onto the graph of f_θ, $\mathcal{G}(f_\theta)$, is unique and has the simple expression

$$P_{\mathcal{G}(f_\theta)}(x) = (t, f_\theta(t)).$$

The projection is often referred to as the fitted value and denoted $\hat{x}$. This yields a second representation of the observation, which in combination with the Gauss-Pearson decomposition constitutes the system

$$\begin{cases} x = \mu + u, & \text{(the Gauss-Pearson decomposition)} \\ x = \hat{x} + e. & \text{(the projection representation)} \end{cases}$$

If $\mu \in \mathcal{G}(f_\theta)$ then under the present assumptions it holds with probability one that $\mu = \hat{x}$, and the system then yields $e \sim \mathcal{L}(u)$, where e is the projection representation residual. For a sample of n observations, in the arrow accent notation (see Example 3), it follows that if $\vec{\mu} \in \mathcal{G}(f_\theta)^n$ then with probability one $\vec{e} \sim \mathcal{L}(\vec{u})$. The Gauss-Pearson decomposition thus yields a duality between these two propositions, and the most probable Kepler orbit f_θ corresponds to the most probable projection representation residual $\vec{e}$ relative to the probability distribution $\mathcal{L}(\vec{u})$, as evaluated under the density criterion.

Therefore, the most probable Kepler orbit is the one which maximizes the density of its corresponding projection representation residual $\vec{e} = -P_{\mathcal{G}(f_\theta)^n}(\vec{x}) + \vec{x}$. It is a simple exercise to show that under the normal distribution assumption maximizing the density reduces to minimizing a sum of weighted squares. In particular, if all variances are equal, then maximizing the density reduces to minimizing the sum of squares $||\vec{e}||^2$, where $|| \cdot ||$ denotes Euclidean distance.

Example 5 (Inference about distributions). Plato was adamant that hypotheses be formulated with respect to the ideal forms of phenomena (White, 1976). Indeed, how could knowledge be inferred with respect to something which is wholly unpredictable and subject to perpetual flux? However, by definition the random part of a Gauss-Pearson decomposition is a random variable, and as such it has a probability distribution which by construction is an ideal form. Hence hypotheses can be formulated with respect to the probability distribution of a random part. Actually, Pearson (1900) constructed the statistical hypothesis test for the purpose of falsifying the then-prevalent notion that all variables in nature are normally distributed.

Suppose $x_1, \dots, x_n$ are real-valued observations, $\mathbb{X} = \mathbb{R}$ with its standard addition and topology, and that the ideal parts are ex ante known to equal zero. Then the observations equal their random parts, which in this example are further assumed to be statistically independent and identically distributed with some distribution function denoted F. The empirical distribution is an observation of a probability distribution, and therefore let $y_1, \dots, y_n$ be the empirical distribution function values at $x_1, \dots, x_n$, i.e. for $i = 1, \dots, n$,

$$y_i = \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{(-\infty, x_i]}(x_j),$$

where $\mathbb{1}_A$ denotes the indicator function of the set A. The empirical distribution function is conventionally denoted $\hat{F}_n$, and in the present notation it satisfies $\hat{F}_n(x_i) = y_i$.

The ideal parts of the empirical distribution observations, $y_1, \dots, y_n$, are the true distribution function values $F(x_1), \dots, F(x_n)$, and consequently the empirical distribution observations have Gauss-Pearson decomposition $y_i = F(x_i) + u_i$ for $i = 1, \dots, n$. Under a hypothesized distribution function $G(x_i)$, the random parts $(u_1, \dots, u_n) = \vec{u}$ have a particular joint distribution $\mathcal{L}(\vec{u})$. It is also common that the distribution $\mathcal{L}(\vec{u})$ is asymptotically approximated through Donsker's theorem.

Under the present empirical distribution framework, hypotheses about the ideal part of the empirical distribution observations, i.e. the probability distribution of the random parts, can be generated and tested through, for example, the methods of Gauss (1809) and Pearson (1900). Note, though, that the construction requires that the ideal parts of the original observations are ex ante known. A common practical application is experimental study of the errors of measurement instruments.

Example 6 (The leverage principle). Examples 3 and 4 illustrate how inference can be made with respect to the ideal part of a phenomenon if the joint probability distribution of the random parts are ex ante known, and Example 5 illustrates how inference can be made with respect to the joint probability distribution of the random parts if the ideal parts are ex ante known.

The Gauss-Pearson decomposition has three elements: the observation, the ideal part, and the random part. In the aforementioned examples, two of the three are ex ante known, and knowledge is then inferred with respect to the unknown third element. The case in which the ideal and the random parts are ex ante known and knowledge is inferred with respect to the observation has not been treated in any of the present article's examples, but is in the statistical literature called prediction. Through algebraic manipulation of the Gauss-Pearson decomposition, which by definition has group structure, any element can be solved for and its expression given by the other two. Thus, if two elements are given, then those two form an expression for the third element.

The cases when all three or none of the elements are ex ante known can immediately be dismissed as practically uninteresting. The remaining case is that in which exactly one of the three elements is known, in practice most commonly the observation. For example, it is not uncommon that historical records contain observations of some phenomenon, but no knowledge about either the ideal parts or the joint distribution of the random parts of the historically observed phenomenon exist. In this case, solving the Gauss-Pearson decomposition for either the ideal or the random part yields a function of an unknown, and consequently nothing can be concluded with respect to either of the two parts. In an analogy, the observations are just floating around in the emptiness of outer space and there is nothing against which they can be leveraged.

An understanding of this fact, named the leverage principle, is of profound importance for all who use statistical methods. If neither the ideal parts nor the joint distribution of the random parts are ex ante known, then there is nothing to leverage the observations against. This is a fact which cannot be circumvented. Of course, the random part can be assumed to be, for example, jointly standard normally distributed, but then any inference made with respect to the ideal part hinges entirely upon this assumption, and thus the reasoning is in many ways equivalent to directly assuming that the ideal part equals, for example, the identity element. The latter is of course an absurdity.

Statistical methods can be used in many situations and for many purposes. The leverage principle explains one situation in which they cannot be used.

Example 7 (Interpreting probability). It is commonly noted that Kolmogorov's axioms are such that they lend the mathematical concept of probability properties that are similar to those of relative frequency. A probability is also linked to its corresponding relative frequency in the limit through the law of large numbers. This example discusses how the present framework allows for the relation between probability and relative frequency to be more precisely worded and elaborated on in the context of Plato's theory of forms.

In mathematical detail, let $x_1, x_2, \dots \in \mathbb{X}$ be independent and identically distributed random variables with probability measure $\mathbb{P}$, and A a subset of $\mathbb{X}$. Further, let $\mathbb{1}_A$ denote the indicator function of the set A and let $\hat{P}_n(A) \in \mathbb{R}$ denote the arithmetic mean of $\mathbb{1}_A(x_1), \dots, \mathbb{1}_A(x_n)$, i.e.

$$\hat{P}_n(A) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_A(x_i),$$

which is often referred to as the relative frequency, or empirical measure, of A. For each $n = 1, 2, \dots$, the relative frequency has Gauss-Pearson decomposition $\hat{P}_n(A) = \mu + u$

where the ideal part μ equals the probability $\mathbb{P}(A)$ and the random part u is a normalized binomially distributed random variable. Since the random part has expectation zero, the relative frequency, $\hat{P}_n(A)$, is an unbiased observation of its ideal part, the probability $\mathbb{P}(A)$. Furthermore, it follows by the strong law of large numbers that the relative frequency, $\hat{P}_n(A)$, converges with probability one to its ideal part, $\mathbb{P}(A)$, as n, the sample size, goes to infinity (cf. Example 1).

Consequently, the interpretation of probability that the present framework yields is that probability is the ideal part of the corresponding relative frequency and, conversely, that relative frequency is an unbiased observation of probability. Under Plato's theory of forms (see Section 2) probability is the Platonic ideal form of relative frequency, and relative frequency is the manifestation of probability in the world of senses. Indeed, relative frequency is an observed phenomenon and it therefore has a Platonic ideal form, the term of which is probability. And because probability is an ideal form it cannot be observed directly, but it casts a shadow into our world of senses referred to as relative frequency.

This interpretation of probability, which traces its origins to Bernoulli (1713), complements Kolmogorov's mathematical definition and can for example be of value in the teaching environment. It is a historic predecessor of the so-called limit-interpretation of probability, which it also contains in the sense that the limit-interpretation can be derived from the present interpretation through an application of the law of large numbers.

6. Discussion

The Gauss-Pearson decomposition is an important part of the history of statistics; but its role in modern day statistics is arguably of even greater importance. Specifically, the Gauss-Pearson decomposition is the detailed premise of Gauss' and Pearson's methods for generating and testing hypotheses through empirical observation, as well as the premise of many other statistical methods. An understanding of the Gauss-Pearson decomposition benefits the understanding of the statistical methods, how they are derived, and what they can and cannot be used for.

Interesting in and of itself, Gauss' observation postulate fundamentally builds upon Plato's theory of forms. Note that the present article does not argue that the notion that each phenomenon has an ideal form is universally valid. On the contrary, there may well be situations in which the notion of an ideal form does not make sense, and hence Gauss' observation postulate does not hold. In such a situation, it follows by definition that the observation does not have a Gauss-Pearson decomposition and thus methods which rest on the decomposition cannot be meaningfully applied. In short, it is not meaningful to use methods designed to make inference with respect to an ideal part which does not exist.

Because of its simplicity, the Gauss-Pearson decomposition has an inherent pedagogical value. In the teaching environment, discussing Gauss' observation postulate and detailing the Gauss-Pearson decomposition can facilitate understanding of many aspects of statistics. Understanding the statistical hypothesis test, for example, can be made easier through a detailed explanation of the premises from which it is derived. Considering the prevalence of the statistical hypothesis test in modern science, it can well be argued that the Gauss-Pearson decomposition is not only a pillar of statistics, but a pillar of modern science.

Acknowledgements

This work was supported by the Swedish Research Council, project 435-210-565. The author is grateful to Andrew Bray for valuable comments and suggestions for improvements.

References

Archimedes (c. 287-212 B.C.). *On floating bodies, book 1 and 2*. English translation by T. L. Heath, 1897.

Arnauld, A., & Nicole, P. (1662). *La logique, ou l'art de penser*. Paris: Savreaux. English translation by J. V. Buroker, 1996.

Bernoulli, D. (1778). Diiudicatio maxime probabilis plurium obseruationum discrepantium atque verisimillima inductio inde formanda. *Acta Acad. Scient. Imper. Petrop., pro Anno 1777, Pars prior*, 3–23. English translation by C. G. Allen, 1961.

Bernoulli, J. (1713). *Ars Conjectandi*. Basel: Thurnisiorum Fratrum. English translation by E. D. Sylla, 2006.

Einstein, A. (1905). Zur Elektrodynamik bewegter Körper. *Ann. Phys.*, *322*, 891–921.

Ekström, J. (2011). On Pearson-verification and the chi-square test. UCLA Statistics Preprint.

Galilei, G. (1632). *Dialogo sopra i due massimi sistemi del mondo*. Florence: Per Gio, Batista Landini. English translation by S. Drake, 1953.

Galilei, G. (1638). *Discorsi e dimostrazioni matematiche, intorno à due nuove scienze*. Leiden: Elsevier. English translation by S. Drake, 1974.

Gauss, C. F. (1809). *Theoria Motus Corporum Coelestium in sectionibus conicis solem ambientium*. Hamburg: F. Perthes und I. H. Besser. English translation by C. H. Davis, 1858.

Gauss, C. F. (1821a). Anzeige: Theoria combinationis observationum erroribus minimis obnoxiae. *Göttingische gelehrte Anzeigen*, *33*, 321–327.

Gauss, C. F. (1821b). Theoria combinationis observationum erroribus minimis obnoxiae: Pars prior. *Commentatines societatis regiae scientarium Gottingensis recentiores*, *5*. English translation by G. W. Stewart, 1995.

Grube, G. M. A. (1935). *Plato's thought*. London: Methuen.

Hald, A. (1990). *History of Probability and Statistics and their Applications Before 1750*. New York: John Wiley & Sons.

Irwin, T. H. (1999). The theory of forms. In G. Fine (Ed.) *Oxford Readings in Philosophy: Plato Vol. 1*. Oxford University Press.

Newton, I. (1726). *Philosophiæ Naturalis Principia Mathematica, 3rd ed*. London: G. & J. Innys. English translation by A. Motte, 1729.

Pascal, B. (1665). *Traité du triangle arithmétique, avec quelques autres petits traités sur la même matière*. Paris: Guillaume Desprez.

Pearson, K. (1900). On the criterion that a system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Phil. Mag. Ser. 5*, *50*, 157–175.

Pearson, K. (1911). *The Grammar of Science, 3rd ed*. London: Adam and Charles Black.

Ross, W. D. (1951). *Plato's Theory of Ideas*. Oxford: Clarendon Press.

Taylor, A. E. (1936). *Plato: The man and his work*. New York: The Dial Press Inc.

White, N. P. (1976). *Plato on knowledge and reality*. Indianapolis: Hacket Publishing Company.

UCLA Department of Statistics, 8125 Mathematical Sciences Building, Box 951554, Los Angeles CA, 90095-1554

E-mail address: joakim.ekstrom@stat.ucla.edu