How Old is Statistical Inference?

Ralph deLaubenfels

Ohio State University, Columbus, Ohio, 43210

email: delau@math.ohio-state.edu

ABSTRACT. This brief note has three parts, tied together by being related to the age, or apparent age, of statistical inference. The first part observes how easily one may get the impression that statistics is only about one hundred years old. The second part gives examples to demonstrate that an age of three hundred years is more accurate. The third part argues that many of the origins of statistics may be found in the mathematics and philosophy of the classical Greeks. Detailed references are given throughout, to enable the reader to flesh out what are meant to be only quick summaries.

Key Words and Phrases: history of statistics, history of mathematics, classical origins

1. HOW CLASSICAL IS "CLASSICAL STATISTICS"?

It is easy to get the impression, especially when beginning graduate study in the subject, that statistics is about one hundred years old. Particularly misleading is the common use of "classical" as being synonymous with "frequentist," especially as formulated by Fisher and his contemporaries, in the basic partitioning of statistics into two camps of varying partisanship. The excellent textbook Casella and Berger (2002) introduces Bayesian statistics with the following sentence, at the beginning of Section 7.2.3, top of page 324: "The Bayesian approach to statistics is fundamentally different from the classical approach [frequentist, Fisher et al.] that we have been taking." Consider also, on page 2 of Lehmann and Casella (1999) "frequentist (the classical approach of averaging over repeated experiments),". The following more detailed exposition may be found in Stuart and Ord (1991, p. 1197, lines 15–18): "...the *frequentist* paradigm, sometimes known as the *classical* approach, which has been the dominant school of statistical thought for most of this [the twentieth] century. However, the *Bayesian* viewpoint has grown in popularity since the 1950s, ..." (italics theirs). In that same book, under "classical inference" in the index, one finds "*see* Frequentist inference."

Thus we seem to see frequentist statistics, beginning in the late nineteenth or early twentieth century, as the Parthenon of statistics, with Bayesian statistics a new avant garde approach arising in the lifetime of the currently middle-aged. Some discomfort with this modernist conception of statistics may occur if one hears of Thomas Bayes, 1701–1761, but this coincidence of names could be explained as another esoteric joke, analogous to the statistician formerly known as Student, and his distributions. A superficial "Google" search of Thomas Bayes yields "British mathematician and Presbyterian minister," not really the model of a modern statistician.

In apparent support of this view of statistics' chronology, Stigler (1999), in Chapter 8, proposes to "advance and defend the claim that mathematical statistics began in 1933."

2. EARLY APPEARANCES OF POPULAR STATISTICAL ENTITIES.

The second third of this brief note is meant to show that statistical inference is closer to three hundred years old; that, in particular, most of the objects that now appear in an introductory statistics course were in standard use in the eighteenth century.

Let's begin with the central limit theorem, really a probability result, but so fundamental to large-sample statistical inference that it may be considered on the interface of probability and statistics. DeMoivre proved it for binomial distributions in 1733 (see Hald (1998, Sec. 2.3, p. 17)). Laplace proved the general case in 1810 (Hald (1998, Sec. 17.2, p. 307)). See also Kendall and Plackett (1977, pp. 101–104), for a description of Daniel Bernoulli's work on the central limit theorem, 1770–1771.

Least-squares regression (not by that name) was introduced by Legendre in 1805 (see Stigler (1986, p. 55) and Hald (1998, Sec. 6.4, p. 118)) and Gauss in 1809 (see Stigler (1986, p. 140) and Hald (1998, Sec. 18.1, p. 351)). Gauss asserted, probably correctly, that he had gotten his results as early as 1795. Gauss's formulation included probability distributions on the parameters and the error. Throughout the second half of the eighteenth century, the same idea of choosing parameters that minimize error was presented, with either the sum or the maximum of the absolute deviations replacing the sum of the squares of the deviations, in other words, L^1 or L^∞ replacing L^2. See Hald (1998, Chap. 6), Stigler (1986, Chap. 1), and deLaubenfels (2006), and the references therein.

Lambert in 1760 solves a maximum likelihood equation to get the MLE (maximum likelihood estimator) for a population mean μ, in a probability distribution $p(x) = f(x - \mu)$, with the function f specified (see Hald (1998, pp. 81–82)). Daniel Bernoulli also constructs an MLE in a 1778 paper, although there is reason to believe he wrote up the results as early as 1769 (see Pearson and Kendall (1970, pp. 155–172) and Hald (1998, pp. 84–85)). Lambert gives no motivation for using the MLE, while Daniel Bernoulli calls his motivation "metaphysical rather than mathematical." Laplace's "Principle" of inverse probability,

$$p(\theta|\vec{x}) \propto p(\vec{x}|\theta),$$

for a probability density or mass function p, data $\vec{x}$, parameter θ, which follows by Bayes' theorem from his uniform formulation of parameter spaces (see Stigler (1986, pp. 102–105)) was stated in 1774 for θ discrete and in 1786 for θ continuous, although the continuous formulation was used by him in both 1774 and 1781 papers (see Hald (1998, pp. 160–161)). Besides launching a general theory of Bayesian statistical inference, note that Laplace's principle makes maximizing the posterior probability equivalent to choosing the MLE. Gauss's presentation of least-squares regression, mentioned above, included the observation that, for a normal distribution on the error, the least-squares estimators are MLEs.

Credible sets, the Bayesian analogues of confidence intervals, are virtually automatic once one has explicit representations of posterior distributions. But authentic frequentist confidence intervals also appear in the eighteenth century. Lagrange in 1776 uses the asymptotic normality of $(p - \hat{p})$, for binomial parameter p, sample proportion $\hat{p}$, with mean zero and variance $\frac{\hat{p}(1-\hat{p})}{n}$, to assert, for any positive t, that "the probability that the value of p is enclosed between the limits"

$$\hat{p} \pm t\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{*}$$

equals $(\Phi(t) - \Phi(-t))$, where Φ is the cumulative distribution function for the standard normal variable Z (see Hald (1998, p. 23)). The interval (*) is a (frequentist) confidence interval for p, with confidence level $(\Phi(t) - \Phi(-t))$. Laplace also gets this confidence interval from the central limit theorem in Laplace (1812), and comments there that the same result is obtained by the Bayesian approach of treating p as a random variable with uniform prior distribution, and using the asymptotic normality of the posterior distribution, as he did in 1774 (see Hald (1998, pp. 24–25)).

Arbuthnot in 1710 tested the null hypothesis of male and female births in London being equally likely; he called this hypothesis "Chance" (see Kendall and Plackett (1977, pp. 30–34)). The data consisted of there being more male births than female births each year for 82 years in a row. The probability of this data, given the null hypothesis, is $(.5)^{82}$, sufficiently small for him to conclude that "it is Art, not Chance, that governs." That was a frequentist hypothesis test (Stigler (1986, pp. 225-226), Hald (1998, p. 65)). Other statisticians, such as Nicholas Bernoulli, suggested that the hypothesized "chance" (that is, the binomial parameter p defined to be the probability that a birth in London is male) could be something other than .5 (Stigler (1986, p. 226), Hald (1998, pp. 65–66)). Laplace, in 1781, addressed a similar problem in Paris, but with Bayesian techniques: given 251,527 male births and 241,945 female births from 1745–1770, he calculated a posterior probability

$$P(q \leq .5 \,|\, \text{ data }) = 1.1521 \times 10^{-42} \ \ (q \equiv \text{ probability that a birth in Paris is male}),$$

concluding that male births were more likely than female births. Even more extreme numbers gave a similar conclusion in London. He also tested the null hypothesis of male births in London being less likely than male births in Paris, getting a posterior probability of

$$P(q > p \,|\, \text{ data }) = \frac{1}{410{,}458},$$

and concluding that London was more prone to male births than Paris. Interestingly, when a similar comparison to Paris and Naples gave a posterior probability of $\frac{1}{100}$, he said it "is not sufficiently extreme for an irrevocable pronouncement" (Stigler (1986, pp. 134–135)). It is stimulating to one's historical imagination to contrast this with the modern habit, at least in many circles, of a significance level of .01 always representing sufficiently strong data for rejecting a null hypothesis.

Speaking of historical imagination, Arbuthnot's 1710 paper was titled "An argument for Divine Providence, taken from the constant regularity observ'd in the births of both sexes." Here we have two subjects a prudent modern speaker or instructor might be hesitant to address: anything as openly religious as "Divine Providence" and formulating birth gender as a binomial problem–which gender do I choose as "success" in the traditional binomial counting of "successes"?

3. TRULY CLASSICAL ORIGINS.

In the last third of this note I would like to argue that many origins of statistical inference may be found in certain aspects of classical Greek mathematics and philosophy. This is not meant to imply these ideas *are* statistics, but that statistical inference shares them with calculus as a common ancestor, so that calculus and statistics may be considered supportive cousins.

For a rough working definition of statistical inference to refer to, I quote introductory definitions from two webpages of Departments of Statistics:

University of California, Los Angeles: (www.stat.ucla.edu) "...make decisions in the face of uncertainty";

The Ohio State University, Columbus: (www.stat.osu.edu/dept/emphasis.html) "...science of decision making under uncertainty..."

Here is an outline of my argument.

1. Historically, Greek mathematicians were the first to deal with the style of uncertainty that statisticians deal with.

2. Closely correlated with this uncertainty was their dealing with continuous models.

3. They were the first to not only make approximations but methodically estimate the error in these approximations, in particular controlling error with inequalities involving the quantity being estimated (see Resnikoff and Wells (1984, Chapter 4)).

4. Platonism, roughly, the philosophy that there is a perfect immaterial world, which our observed material world can only approximate or be shadows or reflections of, nicely describes the relationship of statistics to probability, a sample to a population, or a statistic to a parameter, and, more generally, the incorrigible and always potentially infinite error and uncertainty involved in statistical inference.

5. Although Greek mathematicians were not heavy users of inference (their astronomy and geography may be an exception; see Struik (1987, pp. 54–55), Resnikoff and Wells (1984, pp. 93–103)), they were responsible for the necessary precursor, rigorous deduction. Techniques of deduction are a necessary first step before inference can be effective; deduction (in particular, for statistical inference, probability) is how one tests the candidates for hypotheses arrived at by inference.

6. Although the Greeks did not develop probability, they understood how essential it was to knowledge or truth, as illustrated by Aristotle's question of whether the statement "there will be a sea battle tomorrow" is true or false (Anglin and Lambek (1995, p. 68)).

Now I will briefly elaborate on some of the points above.

Unlike probability, where both discrete and continuous models appear routinely, statistics seems fundamentally continuous, both literally in the sense that a continuous parameter space is usually the most natural, and subjectively, in the elusiveness of the quantities sought.

At least as early as the statement "all is water" ascribed to Thales (624–547 B.C.), historically the first real mathematician in the sense of addressing "why" in addition to "how," many Greeks had a continuous model for the world (not all; Pythagoras (570–500 B.C.) stated "all is number," meaning natural or at most rational numbers, a discrete model). Besides their belief in truth as an end in itself, the classical Greeks' great interest in geometry led to considering continuous phenomena, and thence to encountering systemic uncertainty. Where, exactly, are you on a line?

Irreconcilable confusion was exacerbated by their insistence on allowing as "numbers" only rational numbers (This perspective persisted even with many mathematicians up through the end of the nineteenth century; consider this quote from Kronecker, in 1886: "The integer numbers were made by God, everything else is the work of man." See Struik (1987, p. 162.)

The earliest Pythagoreans were forced into intimacy with this discord between their continuous geometry and their discrete arithmetic merely by drawing the simplest right triangle, with both legs of length one, hence a hypotenuse of length $\sqrt{2}$, not rational (see Struik (1987, p. 42), Stillwell (1989, Sec. 1.5), Anglin and Lambek (1995, Chap. 10)).

Probably the Greeks' most famous encounter with uncertainty is Zeno's paradoxes (see Anglin and Lambek (1995, p. 55)). Here is one perhaps less well known than the tortoise that could never be overtaken: an arrow in flight, at a fixed instant, travels no distance, therefore has no motion. As with the tortoise, Zeno's surely ironic conclusion is that motion is an illusion. As with other paradoxes of the time, the seeming contradiction comes from the conflict between continuous (geometry; specifically, length) and discrete (numbers and time) models.

The style that the problems above share with statistical inference is the pervasiveness of the uncertainty. Perhaps especially motivating, with their sense of incompleteness, are those problems that weren't solved at the time, such as Zeno's paradoxes (although, in fairness to the classical Greeks, impressive successes, such as the method of exhaustion and "atoms"; see Struik (1987, pp. 46–47); must be mentioned). These problems set the stage for both statistics and calculus. They led to the practical philosophical orientation that approximation is necessary and error and imperfection are inevitable; hence the desire for models that are not too sensitive to imperfect measurements or assumptions, denoted "robust" in statistics and "well posed" in mathematics. Statistics and calculus represent different, but closely related, responses to that outlook: calculus with limits, statistics with inference from samples.

Speaking more literally of philosophy, Dale (1999) suggests, in footnote 3 of Chapter 1, that Plato's parable of the cave could be considered the first example of an inverse problem. More generally, according to Platonism the material world that we see consists merely of shadows of an unseen immaterial world of "forms." See Anglin and Lambek (1995, pp. 67–68 and elsewhere), for a quick description of Platonism. For example, the idea, or "form" of a circle is (only) *approximated* by a drawing of a circle, an earring, or an Olympics symbol. Reasoning from these material approximations to a conception of the immaterial idea is inverse reasoning or inference.

More directly to the point, any continuous model must be considered in the realm of the immaterial. Any information achieved materially (that is, data) is finite, and the set of all possible such information is therefore countable. Continuous objects are uncountable. For example, the support of a continuous random variable is a union of intervals of real numbers, an uncountable set. The data that we actually calculate or work with is decimal approximations of real numbers, a countable set.

The acknowledgment of the existence of approximation and error is equivalent to accepting the presence of uncertainty. A sign of this equivalence is the development of a "theory of errors" (Hald (1998, Chap. 5)) in the foundations of statistics that developed in the eighteenth century. A key step in the development of statistical inference is Thomas Simpson's decomposition, in 1755, of a measurement into

$$\text{measurement} \;=\; (\text{true value}) + (\text{error})$$

(see Eisenhart (1961), Plackett (1972), and Stigler (1986, p. 90)).

Note that the extension from approximations of a desired quantity to error bounds on that quantity, alluded to in point 3, is very analogous to the important step in statistics of putting a margin of error around an estimator of a parameter, to form either confidence or credible (depending on one's statistical orientation) intervals.

Early origins of probability may be found in games of chance, including interesting precursors of dice; see David (1998). I think Aristotle's comment in point 6 shows that the Greeks understood more than others of the time how fundamental probability is. See Kendall and Plackett (1977, pp. 1–14), for a discussion of the classical Greeks and probability.

The intimate relationship between calculus and statistics, supported by this common ancestry, is important, because the ultimate justification for statistical constructions and techniques, frequentist or Bayesian, really lies in their asymptotic behavior, that is, their convergence, the calculus response to uncertainty, to the desired quantity.

For the penultimate paragraph, I briefly allude to two dissenting opinions. I've already mentioned Stigler (1999, Chap. 8), giving a "point estimate" of 1933 for the beginning of statistics. A different sort of disagreement may be found in Kendall (1960), objecting to excessive stretching of historical origins into the past ("The writer on any modern idea who can claim the Chinese thought of it first in the Shang period is usually regarded as having scored a point.")

I argue that origins of the sort I've described, of a people (Greek mathematicians and philosophers) identifying problems whose solutions were often left to their intellectual descendants (modern statisticians and mathematicians) is the sort of historical connection that is particularly valuable, in the traditional sense of not repeating mistakes of the past. Here is an example. Greek mathematicians eventually got the area of a circle (see Archimedes' proof, Anglin and Lambek (1995, pp. 98–100)); much earlier phases of the analysis began with upper and lower approximations, by, respectively, circumscribing and inscribing regular polygons. In an early phase of this reasoning, the sophist Antiphon (about 425 B.C.) asserted that a circle actually *is* the inscribed polygon, at least for sufficiently many sides. This was closely related to Zeno's paradoxes, sharing the same confusion about continuous versus discrete; in this case, Antiphon was visualizing space as being discrete, so that the true nature of the circle—smooth, instantaneous changes of direction–was degraded into something subjectively quite different, the finite number of abrupt changes of direction of a polygon (see Anglin and Lambek (1995, pp. 60–61)). This equating of an approximation with the real thing is analogous to treating 99% probability (or the more indirect confidence or significance) as being equivalent to certainty.

REFERENCES

Anglin, W. S., and Lambek, J. (1995), *The Heritage of Thales,* Springer, New York.

Casella, G., and Berger, R. (2002), *Statistical Inference* (2nd ed.), Duxbury.

Dale, A. I. (1999), *A History of inverse probability* (2nd ed.), Springer, New York.

David, F. N.(1998), *Games, God and Gambling,* Dover, New York.

R. deLaubenfels (2006), "The Victory of Least Squares and Orthogonality in Statistics," The American Statistician 60, pp. 315–321.

Eisenhart, C. (1961), "Boscovich and the combination of observations," in *Roger Joseph Boscovich,* ed. L. L. Whyte, Allen and Unwin, London, 200–212; see also (Kendall and Plackett 1977, pp. 88–100).

Hald, A. (1998), *History of Mathematical Statistics, From 1750–1930,* Wiley Series in Probability and Statistics, New York, NY.

Kendall, M. G. (1960), "Where shall the history of statistics begin?," *Biometrika* 47, 447–449; see also (Pearson and Kendall 1970, pp. 45–46).

Kendall, M. G., and Plackett, R. L., eds. (1977), *Studies in the History of Statistics and Probability,*(Vol. 2) London: Griffin.

Laplace, P. S. (1812), *Theorie Analytique des Probabilites*, Courcier. Paris.

Lehmann, E. L. and Casella, G. (1999), *Theory of Point Estimation,* (2nd ed.), Springer, New York.

Pearson, E. S., and Kendall, M. G., eds. (1970), *Studies in the History of Statistics and Probability,* (Vol 1) London: Charles Griffin.

Plackett, R. L. (1972), "The discovery of the method of least squares," *Biometrika* 59, 239–251; see also (Kendall and Plackett 1977, pp. 279–291).

Resnikoff, H. L., and Wells, R. O., Jr. (1984), *Mathematics in Civilization,* (2nd ed.), Dover, New York.

Stigler, S. M. (1986), *The History of Statistics: The Measurement of Uncertainty before 1900,* the Belknap Press of Harvard University Press, Cambridge, MA.

Stigler, S. M. (1999), *Statistics on the Table: The History of Statistical Concepts and Methods,* Harvard University Press, Cambridge, MA.

Stillwell, J. (1989), *Mathematics and its History,* Springer-Verlag, New York.

Struik, D. J. (1987), *A Concise History of Mathematics* (4th ed.), Dover, Ontario.

Stuart, A., and Ord, J. K. (1991), *Kendall's Advanced Theory of Statistics* Vol. 2, Oxford University Press, New York.