MISUSES OF STATISTICS IN AN HISTORICAL PERSPECTIVE
Reflections for a Course on Probability and Statistics

René Guitart

Institut de Mathématiques de Jussieu
Université Paris Diderot Paris 7
175 rue du Chevaleret 75013,
Paris, France
Rene.guitart@orange.fr

ABSTRACT

This paper is a kind of guide for a historical introduction to the beginnings of probability and statistics (limited to the end of XIXth century), through the "Reading of Master", introductive to a mathematical course on these subjects. We provide the minimal historical bibliography. We claim that in many historical masterpieces everyone could recover the original meaning of notions and computations, understand the constitutive relations between probability and statistics, and so avoid the confusion of interpretations made very often today. Indeed, any given practice of statistics or probability is a tailpiece of a historical sequence of fixations, hesitations, remorse, errors, of "mathematical pulsation", and this sequence is their true meaning. To be aware of that it is necessary, as well for new future developments as for applications. Especially it is important to know the history of major misuses and difficulties of probability and statistics in past.

History teaches us that statistics lay on several notional difficulties, as: what are hazard and chance, their dynamics, what are stochastic process and random variable, what are the good processes for limits or infinite gluing of data, from where data do come, what are the good descriptions of populations, characters and samples? All these subjects could be elucidated from classical books. So when doing statistics we have to construct interpretations of results with a serious critical open eye.

We propose to emphasize this historico-epistemological point at the beginning of any course in statistics, especially for the benefit of future mathematics teachers.

1 Introduction

1.1 Reading the Masters

This paper provides an historical introduction to probability and statistics, via the "Reading of Masters". Today we can observe a lot of misuses and misunderstandings of statistics. At first of course these misuses are due to a general ineptitude induced by a bad understanding of the true nature of any given mathematical tool, and the false conception that such a tool could be used blindly. But they came also from specific difficulties of statistics, which can be understood by its history, and the reading of the decisive historical works. We propose a minimal list of such references to read or to consult. Our explanations here are directly related to these historical materials.

1.2 Starting with elementary problematics, and coming back to history

Clearly it could be difficult to read directly the original papers, even if we know that some good explanations are written there. A preparation for such a reading could be at first to study directly two or three very elementary and very good books, as (Granouillac, 1974), (Moroney, 1974), (Lévy, 1979), about very concrete problematics of today in statistics, and with (Boll, 1941,1942) on probability. Secondly we recommend to begin the study of an important historical material, the book (Bertrand, 1889), which is very intuitive, with a lot of significant exercises. Thirdly, we could read historical analysis as (Desrosières, 1993) and (Edwards, 2001) on statistics, and (Barbin & Lamarque, 2004) on both statistics and probability. This last book is synthesized in (Barbin, 2004) in a preface explaining that the type of problems is not the same in probability and in statistics. For example on the question of the relation between causes and events: for Laplace the problem of causes is a question of probability (of causes), whereas for Cournot the problem is to investigate statistically on causes by analysis of effects. Also there are differences between decision and prediction, extrapolations and results, modelling and exploring, etc.

1.3 Two sources, one pulsative mathematical subject

Two subjects seem to be different and not to come from the same source: on the one hand the question of probability or "geometry of the hazard" (Pascal, 1654) regarding uncertainty (Pascal), or likelihood (Leibniz), for applications to equitable judgements, and on the other hand the question of statistics, analysis of data for a state, concerning health, trade, taxes etc., and beneficial decisions in these areas.

But the exploration of the history shows that from a mathematical point of view the two subjects became much related and intertwined; comings and goings between the two subjects is vital in their historical developments. It is perfectly possible to teach separately probability and statistics; but our claim is that the true mathematical subject is in between, and a deep understanding of the correct use of both theories assume a main attention to this "pulsation" (For the idea of "mathematical pulsation" see (Guitart, 1999)).

2 Average, probability and expectation

Today the scientific subject of probability and statistics is the construction of the laws of chance and of distributions of data and their use for analysis in various other sciences, for presentation and analysis of phenomenon and data, for explanation and prediction. In such a scientific use, mainly as a 'logic' to guide experiments and observations, four implicit credits or beliefs are to be examined: the belief in the 'average' as meaningful and significant in reality, the belief in the existence of a real structural organisation for any experimental data, the belief that the average is the natural minimal summary of the organization, and the belief that the variation in statistical data is an effect of chance and so it is a matter of probability.

2.1 Addition, proportion, multiplicity and average, percentage

If a multiplicity of n weighted data (p_i, x_i) is given, then the pondered average or pondered mean value is (p, m) with

$$(p, m) = ((p_1 + \cdots + p_n), (p_1 x_1 + \cdots + p_n x_n)/(p_1 + \cdots + p_n)).$$

The average or mean value is m. Almost all concrete applications of probabilities and statistics could be enough directly reduced to computations of such convenient averages. And a lot of misuses come from the oversight that the average is not so natural and obvious.

The fact to adopt an average in place of a multiplicity of values is not a mathematical principle, but an empirical decision, related to the fact that it amounts to compute proportions, or linear functions, and indeed it is the simplest of arithmetical rules. It could be justified also by the fact that the average minimizes the weighted sum of square distances to the given data. But simplicity does not imply adequacy. The basic objection to average m of the x_i is that if f is a function of x, with $y = f(x)$, why it is not better to compute an average M of the y_i, and then to take, in place of m, the value $f^{-1}(M)$? For example why not to compute the average of x_i^2 rather than the average of the x_i? So it is left to the user to choose the good average (on x or on $f(x)$, and, furthermore, with which weightings). A good discussion on this point is in (Bertrand, 1889).

It is clear that $2 + 3 = 5$ is a theorem, and that 2kg $+$ 3kg $=$ 5kg is a scientific law in physics of pondered bodies, experimentally proved through the Law of Lever or by the determination of barycenters (Archimedes). But on the other hand 2\$ $+$ 3\$ $=$ 5\$ is not a theorem or a scientific law: It is a convention for an exchange accepted today by everyone to enter in the game of finance and trade. So everybody accepts this rule at the root of his practise with money, estimation of salaries, debts etc. Looking at the history of money and bank, it is easy to understand how this rule did not always exist. In fact this rule imposed the transformation of objects of exchange as magnitudes, comparable and additive. Then the exchange could be performed just as balancing in calculus, and henceforth the calculus became the obvious medium of trade. A similar point is in the principle of vote (for decisions in justice, for election of representatives, to choose a product in a market, etc.): the votes are added, and the convention is that the decision is taken according to a numerical calculus (for instance the principle of majority). It is recommended to read (Condorcet, 1785).

In fact these reductions to arithmetical practices of liberal market and liberal democracy could be seen as application of a meta-rule of "reduction to average". So for the addition of dollars we have: (2\$ $+$ 3\$)$/(2 + 3) =$ 1\$. A good reading here is (Cournot, 1835).

A first source of errors with an average is the ignorance of the concrete source of where it comes from (kg, \$, votes, etc.): then a true criticism of the result is impossible. And then the worst consequence is that different averages, and percentages, in different heterogeneous areas, are compared, and do conduct to fallacious correlations and illusory understanding.

2.2 Probability and expectation

Usually it is considered that probability theory starts with the work of Pascal on games (Pascal, 1654), its correspondence with Fermat in 1654, and, that the first treatise is the booklet of Huygens, in 1657: "du calcul dans les jeux de hasards" (Huygens, 1657). At this stage, we get the notion of expectation, while the notion of probability will be introduced explicitly later. The expectation models what is equitable, in presence of a true random situation; it is not what is good, what is beneficial. The analysis by Pascal could be formulated today, in a very anachronistic way, as the analysis of a stochastic process, as the description of conditional expectation (the introduction of the idea of "martingale"); none of these words is from Pascal' s work, but this interpretation is plausible. On the other hand we can insist on the fact that it is not a calculus of probability, a proportion, the quotient of number of favourable cases by the total number of cases [according to Laplace' s view (Laplace, 1812), (Bertrand, 1889)]. A fortiori it is not a calculus of probability via frequencies!

In modern terms, the expectation $E(X)$ of an alea X (a random variable concerning a stochastic process) could be defined as the average of the possible values x_i for X, weighted by the probability p_i that X takes the value x_i. It is the average of the possible values, but it is not the most probable value.

So, on the one hand, historically the expectation is a primary notion, preceding the notion of probability, and a fortiori preceding the notion of frequency; and on the other hand, in its modern expression, this notion seems to be a derived notion, constructed with the notions of probability (the p_i), of statistical distribution (the i), of average. This too analytical view could be a source of misunderstanding.

2.3 Relativity of probability during the time of the process, logical aspects

A difficulty, well illustrated in the Bertrand' s treatise (Bertrand, 1889), is the question of relativity of chance in time, a priori and a posteriori probability, reversions from probability of effects to probabilities of causes, in relation to the converse Bayesian calculus (Bayes, 1763) as reformulated by Laplace (Laplace, 1812). The relativity is also a question of dynamics in a stochastic model (question of martingales), and is related to incomplete information. The difficulty is that these relativities are not often announced explicitly in a concrete problem.

These relativities have also to be mixed with some logical questions, when we compute the probability of a logical combination of random events or variables. This logical point is treated in (Boole, 1854), and is at the basis of the Kolmogorov axiomatics in (Kolmogorov, 1933, p.2). So the modern axiomatic, with a set E of elementary events, a probability function P on a field F of subsets of E, with random variables seen as functions X on E, etc.) is able to support these aspects.

3 Statistics, combinatorics, asymptotic calculus, normal law

3.1 Statistics versus probability

Let us start with a warning: from a mathematical point of view, in principle, statistics need not to be constructed through a probabilistic interpretation. For instance the approach by Jean-Paul Benzecri in his factorial analysis of correspondences is based on a geometrical analysis of the shape of a cloud of experimental data. The observation of symmetries, of repetitions of a motive, of frequencies of a phenomenon, is not necessarily related to a causal interpretation in terms of chance.

Nevertheless, very often it is the case that the variation in a data set is a consequence of chance: in these cases (and in these cases only), the frequency (a notion in analysis of statistical data) could be related to the probability (a notion in doctrine of chance).

The key point is the following. Starting with an elementary reproducible stochastic process, we imagine as an experiment that this process works several times, and we look at the sequence of results. This sequence is a statistical data, with frequencies, etc., and of course, in such a sequence, the variation is due to the stochastic nature of the elementary process, and we can ask for a mathematical relation between the probability of the elementary process and the frequency in the sequence.

So, the probabilistic interpretation of statistics consists in the converse: given a sequence x_i of data, we pose the hypothesis that it is produced as iterated values of an unknown random process X, according to a probabilistic law which is to be revealed.

The difficulties with this point of view become very serious in several directions. The information (the x_i) could be incomplete (in fact a sample in a population). The true nature of the process underlying X could depend in fact on the possible modification in time of this process according to its repetition (the various X_i are not independent). And when we would like to correlate and to compare several sequences x_i, y_j, associated to random variables X and Y: the underlying stochastic processes associated to X and to Y are not necessarily the same, or at least easily correlated. In order to surmount these difficulties we could read analysis on the theory of measurement (cf. § 3.4).

3.2 Probability and frequency, weak law of large numbers of Bernoulli

A central difficulty comes from the confusion between *a priori* probability and frequencies. In fact, history shows that these two aspects are closely interconnected at the heart of the subject: this interconnection finds its mathematical expression in the law of large numbers. We learn this from Laplace and Bertrand' s books (Laplace, 1812), (Bertrand, 1889), where they discuss hazard and chance.

The weak Law of Large Numbers is a mathematical result expressing that the mathematical model of probability is consistent with the frequency interpretation of probability. Informally, and in a rather vague way, this law says that: when the number N of independent repetitions of an elementary stochastic process increases, then the observed frequency f_N of favourable issues "probably approaches" the probability p of the favourable issue, that is to say that the probability $P_N(e)$ that the difference $p - f_N$ exceeds a given positive number e converge to 0, as N increases indefinitely. It is a fundamental result by J. Bernoulli (Bernoulli, 1713).

This beautiful result constructs a relation between three terms: an unknown probability p, an observable frequency f_N, and another unknown variable probability $P_N(e)$; the beginner has to be careful to distinguish among these three terms.

3.3 Asymptotic calculus, de Moivre's Normal Law, Central Limit Theorem

After the treatise of Huygens, three decisive steps in probability where the books by (Bernoulli, 1713), (de Montmort, 1708), (de Moivre, 1718). There the combinatorics is well developed, with a thought towards the statistics of (Graunt, 1662) and (Halley, 1694), around the Binomial Law, and even is pursued towards asymptotic calculus (e.g with the formula of de Moivre-Stirling) and the normal law.

At first we reach the delicate question of the natural extension from finite combinatorics towards probabilities and statistics considered as potentially infinite combinatorics. Hence the asymptotic calculus and variations on limits which has to be defined and put forward. So to use the law of large numbers is not trivial.

The “Normal Law” (so named only in the XIXth century) was discovered by de Moivre, as a limit case of the Binomial Law (Freudenthal, 1957). It was also studied again by Laplace and Gauss, and so it is also known as the “Laplace-Gauss Law”. Its graph is the so called “Bell Curve”. It is related to the elaboration of so called “Central Limit Theorem”, a deep improvement of the weak law of large numbers, resulting from works of de Moivre, Laplace, Gauss. This is to be read in (Laplace, 1812), (Bertrand, 1889). Today, this Central Limit Theorem is considered by probabilists as the central object of probability theory.

3.4 Measurement: Least Squares Method

The theory of measurement by the method of Least Squares could be studied through the works of (Mayer, 1750), (Legendre, 1806), (Gauss, 1855). In Mayer a very interesting empirical method of grouping observations is used, for the analysis of astronomical observations of the Moon. But the more decisive step was the emergence of the so called Least Square Method. The first justifications of this method (Laplace, Gauss) passed through the Normal Law and the Central Limit Theorem. But in fact in (Legendre, 1806) and in the second attempt of Gauss (Gauss, 1855), the justification is outside the scope of probability, even if related to the Bell Curve. It is very instructive to read the story of this subject (Bertrand, 1889), (Derosières, 1993).

3.5 The Average Man

In the XIXth century we get the theory of standard deviation with respect to the average, the development of various statistical or probability laws, motivated by various domains of applications. The development of the subject of probability and statistics is mainly about applications of the method of least squares and the Bell Curve. It is instructive to read various utilizations of this material.

In these applications, a basic difficulty is the confusion between data of several measurements of a given object and values of a given character in a given population. It is perhaps the root of the difficulties with probabilistic justification of the mean square method, and this confusion is excessively admitted by (Quételet, 1835). With his construction of the “Average Man”, Quételet’s basic assumption is the following: there exists an ideal man, and each concrete human is a measure of this “Average”; furthermore this measure is a random variable, made by chance, and according to a normal distribution.

4 Conclusion

In this brief approach, we stop at the end of the XIXth century, even omitting Galton, Pearson, and then Fisher, and so the true birth of modern statistics in the 1920’s, through the theory of samples. There, the central question would be the construction of a good poll: how to construct a representative sample (Fisher), in such a way to get reasonable predictions? Clearly, on the question of the probability that a poll is a good one, the statistics get a new link with probability. We have also omitted the birth of the statistics of laws which are very different from the normal law, and the statistics of extremes (Levy, Gumbel). So, in some sense a theory of exceptions (improbable values) was created, and this is again a new link between statistics and probability.

Nevertheless, today the calculus of probability is in a "pulsation" in itself, between two presentations: on the one hand "à la Kolmogorov" with an E, F, P etc., and on the other hand as a direct manipulation of random variables and of their laws. This has been observed judiciously in (Mazliak, 2002). We think that this pulsation is easy to perceive in History.

And furthermore, on reading the classics, we realized here (although stopping in the XIXth century) that between probability and statistics, another real pulsative knot had been constructed by History. We think that it is important for future teachers to know that; especially this could help them not to reduce the idea of probability to the idea of frequency; otherwise it will be a real fault with respect to the true nature of the subject. Here is the fundamental misuse of statistics, to forget its link with probability as an a priori theory of chance.

REFERENCES

- Barbin, E., 2004, "Préface", in Barbin, E. & Lamarche J.-P. (coord.), *SHistoires de probabilités dt statistiques*, Paris: Ellipses, pp. iii-v.
- Barbin, E. & Lamarche J.-P. (coord.), 2004, *SHistoires de probabilités dt statistiques*, Paris: Ellipses.
- Boil, M., 1941, *Les certitudes du hazard*, Que sais-je ? N^o 3.
- Boil, M., 1942, *L'exploitation du hazard*, Que sais-je ? N^o 57.
- Bayes, T., 1763, An Essay towards solving a Problem in the Doctrine of Chances', communicated by R. Price to the Royal Society.
- Bernoulli, J., 1713, *Ars conjectandi*, Basel. (1801, L'art de conjecturer, Caen: G. Le Roy).
- Bertrand, J., 1889, *Calcul des probabilités*, Paris: Gauthier-Villars (repub. Chelsea Publishing Company).
- Boole, G., 1854, *The Laws of Thought* (French trad. 1992, Paris: Vrin).
- Condorcet, M. J. A. N. de Caritat, 1785, *Essai sur l' application de l' analyse à la probabilité des decisions rendues à la pluralités des voix*. Paris: Imprimerie Royale.
- Cournot, A. A., 1838, *Recherche sur les principes mathématiques de la théorie des richesses* (reed. 2001, Paris: Dunod)
- Desrosières, A., 2010, *La politique des grands nombres Histoire de la raison statistique*, Paris: La Découverte (1ère éd. 1993).
- Edwards, D., 2001, *Annotated Readings in the History of Statistics*: Springer.
- Freudenthal, H., 1965, *Probability and Statistics*: Elsevier (Dutch ed. 1957).
- Gauss, C.-F., 1855, *Méthode des moindres carrés Mémoires sur la combinaison des observations*, trad. J. Bertrand, Paris: Mallet-Bachelier (ré-impression: 2009, Ed. Jacques Gabay).
- Granouillac, A., 1974, *Comprendre la statistique sans mathématique*, Paris: Chotard et ass.
- Graunt, J., 1662, *Natural and Political Observations Mentioned in a following Index, and made upon the Bills of Mortality*, London: J. Martin, J. Allestry and T. Dicas.
- Guitart, R., *La pulsation mathématique*, L' Harmattan, 1999.
- Halley, E., 1694, "An estimate of the degrees of the mortality of mankind, drawn from curious Tables of the Births and Funerals at the City of Breslaw; with an attempt to ascertain the price of annuities upon lives", Phi. Trans., London., n^o 196, 1692–1693, pp. 596–610.

- Huygens, C., 1657, *Du calcul dans les jeux de hazard*, in Œuvres Complètes, La Haye, 1888–1950, t. XIV. (first ed. in Latin 1657). Reprint in *L' espérance du Hollandais ou le premier traité de calcul du hazard*, Paris: Ellipses.
- Kolmogorov, A., 1933, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer (1956, Foundations of the theory of probability, 2d English Edition, AMS Chelsea publishing).
- Laplace, P. S., 1812, *Théorie analytique des probabilities*, Paris: éd. Mme Vve Courcier.
- Legendre, A.-M., 1805, *Nouvelles méthodes pour la détermination des orbites des comètes*, Paris: Didot (reed. 1806, Paris: Courcier).
- Lévy, M. L., 1979, *Comprendre les statistiques*, Paris: Seuil.
- Mayer, J. T., 1750, "Abhandlung über die Umwälzung des Monds um seine Axen und die scheinbare Bervegung der Mondsflecten", Kosmograpische Nachrichten und Sammlungen auf das Jahre 1748,1, pp. 52–183.
- Mazliak, L., 2002, "Sur la pulsation probabiliste", *Bull. APMEP*, **440**, 287–292.
- De Moivre, A., 1718, *The doctrine of chances: or, a method of calculating the probability of events in play*, London: printed by Pearson for the author (3rd ed. 1756, reprint 1967, New York: Chelsea).
- De Montmort, P. R., 1708, *Essay d' analyse sur les jeux de hazard*, Paris (reprint 1980, New York: Chelsea).
- Moroney, M. J., 1974, *Comprendre la statistique*, Coll. Marabout Université n^o 203, Verviers: Marabout s.a.
- Pascal, B., 1654, Usage du Triangle Arithmétique pour déterminer les parties qu' on doit faire entre deux joueurs qui jouent en plusieurs parties, in *Œuvres complètes*, 1963, Paris: Seuil.
- Poisson, D.-S., 1837, *Recherche sur la probabilité des jugements en matière criminelle et en matière civile*, Paris: Bachelier.
- Quételet, A., 1835, *Sur l' homme et le développement de ses facultés, ou Essai de physique sociale*, Paris: Bachelier, I,II.