STATISTICS IN THE BIOMEDICAL RESEARCH AND CLINICAL STRATEGY (*)

Italo Scardovi

1. To deal with statistics as a methodology of living sciences means in fact to delve into a strange event of the history of knowledge, because the statistical method, the statistical mode of thinking, came out of those sciences and nowadays, when techniques which are the result of that method, intervene in certain biomedical research, it sometimes appears as one is witnessing a sudden clash between two "cultures"; almost as the research on the living and the methodology of such a research belong to different and distant lines of thought. If it is so, such has not always been.

After the Galilean revolution, when science and the scientific method were turning into culture and philosophy – critical culture and experimental philosophy – life phenomena did find difficult to identify themselves with the rigid determinism of the physical gnoseology. However, if the great theoretical system by Newton-Laplace did then give way to a more universal intuition of nature, it was because in the living processes, even earlier than in atomic and sub-atomic events, signs did emerge of a new phenomenology, a new natural philosophy: a "statistical philosophy" that, leaving its mark on the sciences of matter and life, created whole bodies of doctrine, from thermodynamics and radio-activity to the biology of evolution, from genetics to the quantum mechanics.

Yet, in several areas of research, the medical one in particular, the gap is still felt. The motivated skepticism of those more alert, who find the refined analysis of quantitative and often gross determinations to be improper, is added to a certain instinctive *horror numeri*, as a result of a perpetuation of human misunderstanding.[1]

(*) The text reports the integrated outline of a Conference held at Bologna University, Medical Faculty, taken from a basic lecture invited by the President, Giuseppe Leti, and presented at the XXIII scientific meeting of the Italian Statistical Society (Bari, 1985).

[1] Emanuele Padoa hit the target when he made ironic comments about some biological research which is pleased with – as per his words – "... refining the mathematical method in order to add a second decimal to a coefficient calculated on the basis of the record of an experiment that, had it been carried out one year later or in another laboratory on a slightly different batch, would basically have given similar results, but different for the first decimals and perhaps some integers ...".

To the vague lack of communication between statistical techniques and biomedical research also contributes the attitude of those who, in charge of the method, address those who are not, as unbelievers to be converted or, even, as ignorant to bear with. This concurs to the separation between method and sciences: a separation detrimental to these and, even more, to that. If many methodologists seem to have forgotten the scientific genesis of several formal procedures, quite a few researches, driven by an almost irrational urge for objectivity, seem to turn to the statistician as one might have turned to an exorcist: searching for an apotropaic ritual, for moral support. Unfortunately, due to lack of scientific knowledge and of a logical outlook on their methods, many statisticians end up accepting this role, almost as enjoying something magical, almost making their instruments, not just a critical protection, nor the result of the historical development of thinking about the phenomena of which the study of the living has shown the first significant paradigms, but rather an heterogeneous cookbook of readymade solutions, devoid of an authentic heuristic reason.

Hence, since the fashion of statistical embellishment added to any, more or less, scientific publication has spread in medicine; one has witnessed a frantic running for the possession of rules and formulae, with passive imitation by some people and with mystical exaltation by others, the most dangerous ones. These and those use the formal techniques, without any interest whatsoever for the theoretical principles from which they are derived, without knowing the phenomena which originated them; they take statistical inductions as self-evident truths, thus reminding us of certain medieval demonstrations on the existence of God. Although one welcomes the introduction of canons and ideas, once upon a time unacceptable or considered as sacrilegious, into many disciplines, yet, because of this, one should mention the risk of irrationally accepting any method, any language. As Leonardo said: "Each instrument should be used with the experience which created it".

These are not new considerations, but still up-to-date ones, which on purpose I digged out of many old papers: texts of seminars in 1975 at Academy of Lincei, invited by the President Beniamino Segre, and dealing with the relationships between biology and statistics, old academic publications, opening lectures of meetings, and lastly the outlines of a conference held in 1979 at a Faculty of Statistics, invited by the Dean Bernardo Colombo. In the spirit of an experimental scientific tradition and of a quantitative and rational way of reasoning, it became then easy to find a starting point for a method in Andreas Vesalius' work. In the full swing of Renaissance he founded the Italian anatomy school. In *De humani corporis fabrica* which was published in 1543, the same year in which an even more innovating book came out: *De revolutionibus orbium coelestium* by Nicolaus Copernicus, I had spotted a passage that I found worth quoting while speaking at the University of Padua and in the very hall dedicated to Vesalius. In claiming the synthesis of theory and observation, of empirical research and methodological interpretation from the medical science, it maintains a transcendental up-to-datedness of its own, even after four and a half centuries. In his *Praefatio* Vesalius despises the deplorable division within the medical art which has introduced in his time schools the fashionable, deplorable system, according to which one person dissects the human body while another describes the parts: this other person is perched on a high pulpit, like a crow, and disdainfully repeats, over and over again, information about facts that he has never observed directly, but has either learnt by heart or describes reading straight from books written by others.

This rebellion against the authority of the scholastic tradition – in which Leonardo's lesson reappears and methodological theses by Galilei and Bacon are forewarned – reflects a past that from many points of view is long gone. Since quite a long time the "physician" has descended from the pulpit and has immersed him into the phenomena, drawing his experience from things. But the vocation for a kind of an overhanging speculation does not seem to have completely disappeared, if it is true that he often addresses the cultivator of the method asking for a "superior" interpretation of his data. So that the statistician would be the new crow: perched high (as he does not have to see), busy developing his own saying and dispensing, from above there, his powers that he appears to have acquired by the method he cultivates. That is a recurrence of the old dualism between who observes and who explains, almost making of statistics a new metaphysics, holding the rules of knowledge.

But it is not in this way that statistics appears today to be in a privileged position. If a whole scientific reasoning – beginning with physics, the leading science – adopts its language and instruments, it is because these have become the logical backbone of a new rationality, the preliminary step for the reconciliation of science and philosophy, to which the "investigation of the living has greatly contributed.

2. By now it should practically be out of the question that the method is a problem common to positive sciences, including the research stage of medicine. Yet, let the statistician come forward, who has not been asked to fill in the blank half page left of an already completed paper, conclusions included, the title "Statistical analysis", prominently at the top of the page. Or the statistician who has not been asked to find some novelties in reams and reams of figures, whatever they were. Or the statistician who, in the first incidence, might have caused some embarrassment by asking what the statistical analysis would do, as conclusions had already been drawn; or, in the second case, statisticians who pretended (unless they were unwary) to understand the hypothesis of the research work, the nature of the variables and the cognitive aim. It is not easy to say in what context of discovery or proof, a certain, almost notary-like, passion for clinical or laboratory case studies, one which not even the most obstinate empiricist would approve of, finds its allocation. However, it is not difficult to see in this a misunderstanding of statistics as inductive logic and scientific method. Science is not data accumulation. It is the development of hypotheses, the struggle with doubts; it is the continuous and often despairing research for objective supports to reasoning. It is not a matter of patiently gathering figures and news, from which one can magically draw some transcendental truth for the artifice of formal, statistical and non-statistical rules.

If I remember these academic *topoi*, it is not because of their strange episodic aspect, but because of their typical aspect and because they reveal a dangerous lack of critical spirit. To turn an archive inside out is not engaging oneself with science. Nor, if the "abacus man" is allowed to ask a question, one can see how a clinic should be held when the diagnosis is found as an automatic consequence of a mass of standard analyses, mechanical controls and instrumental readings, behind which any representation of the living is likely to get lost. Just too bad for him, if his personal "parameters", each of them taken individually, do not fit within the limits imposed to the "normal" in the operational definition of each variable. Checks, analyses, findings have a very precise and, if I may say so, Galilean function: to test a hypothesis, in this case the diagnostic hypothesis.[2] Otherwise measurements, tests, checks no longer play the role of auxiliary instruments, of indispensable help to man's intelligence. Due to a strange-methodological inversion, they often become mechanical substitutes for reasoning, intuition, mental synthesis: synthesis of observations and abstractions, of data and ideas.

Going back to the past, then? Glorifying the old-fashioned witch-doctor, a bit of a healer and a bit of a clairvoyant? Not, surely. It is instead the refusal of a new witchcraft that no longer evokes the *aura vitalis* or the *virtus dormitiva* ("It is precisely there, where concepts are lacking, that the word makes its timely appearance" – says Mephistopheles to the young person who is about to enter University), but succeeds in making of the measurement, the test, the numerical datum, an end rather than a mean, almost an excuse to avoid thinking. In fact, a substitute of the ancient, esoteric words in the survival of a moral and intellectual tradition that from the medical practice goes back to the medical research. It is the tradition that still compels so many people involved in this important area of knowledge to ask the methodologist for some ornamental little formulae, so as to add a touch of "scientificity" to publications overflowing with names, if not always with findings, or to ask him for the elaboration of an original thesis from an inventory of cases, as by a kind of superior conjuring. Francois Jacob said: "Some people seem to believe that in order to follow a scientific procedure it is enough to measure anything and to feed the obtained data to a computer. As if in the debate between theory and experience, facts had the right to speak first, or as if, by piling up data and handling them statistically, one might have some reasonable expectation of enucleating a general principle".

[2] Because, if it is true that medicine is first of all a practice directed to the therapeutic purpose, it is also true that, at least since Claude Bernard, it has tried to achieve a scientific asset, to make a pragmatic use of the important discoveries in biology. But if physiology, histology, embryology, microbiology, molecular biology and genetics set up the theoretical bases for a medical knowledge that, through them, becomes science, experimental science, so as to avail itself of the important discoveries in biochemistry, chemistry and physics itself, precisely for the mediation of the biological research, yet, one cannot see how medicine might adopt those results without deriving a method associated with them.

3. Within the limits of this talk it is not possible to get deeper into the naturalistic genesis of the statistical method, the taking over of a statistical intuition of life phenomena, the logical and historical reasons of statistics as methodology of biomedical sciences. Certainly there is no lacking of significant precedents: from Davenport's "zoo-statistics" to Farr's "vital statistics", to that "medical statistics" which owes so much to the eighteenth-century actuaries, and which in the nineteenth century divided academies in "numberists" and "anti-numberists", causing unending logomachies. It is not without significance that the benefits of antisepsis have been identified after statistical observations; that polemics regarding vaccination and serum therapy were argued on statistical data; that the unhappy effects of German measles in pregnant women were perceived and statistically proven many years before medical science could draw the appropriate information; that the study of twins left the dark side of myth when, for it, a statistical content became essential. But, to indulge only on these aspects it would mean to neglect the most authentic conceptual gain obtained by statistics from the life sciences: it would mean not to mention that the first statistical vision of reality, the intellectual choice, of which a certain type of physics will make a concept of the world, was born in the biological research, with Darwin and Mendel.

Neither, after the fashion of our times so much more oriented towards information rather than formation, more towards news than concept, would I know how to change my subject into a more or less up to date chronicle of principles and techniques. There are already so many descriptive catalogues, even by esteemed biostatisticians concerned about scrupulously recording the latest "application", the latest algorithm, pleased with the ever-growing attitude of entrusting so much of detailed research to the automation of statistical calculation. As if the problem was that of more calculations, rather than of more reasoning; of using so many techniques, rather than thoroughly carrying out one method. It might well be that I am old-fashioned, but I cannot join the chorus of easy enthusiasm for the "statistical" reform of some biomedical research branches.

Surely risking to be charged of heresy, I dare say that I rather like the situation of time when certain statistical assumptions regarding the phenomenon of the living were hotly debated and rejected than the present one, when every doubt seems to have been overcome, every hesitation to be without reason, every calculation possible and every algorithm a solution. This because of two reasons: the conceptual link between phenomenal reality and mathematical abstraction cannot be either superficial or contingent; and all this naive trusting of computerized mechanics, all this damaging falling for the undoubted fascination of certain syntaxes, without realizing their inner strong logic, their inner and deep scientific content, cannot but turn into a sterile and useless turmoil, that has got nothing to do with science and leaves fundamental problems unsolved. The signs of this may, perhaps, are foreseen in some recent outcries which, with some motivations in their preliminaries and none in the conclusions, are really and truly rejection crises. Naturally it is like this! It is natural and worrying, because from the criticism of certain careless uses of a method, one may end up challenging the method itself, from the reasoned rebuke of improperly using mathematical statistics, one moves to attack mathematics as such, without the benefit of doubt only given by true culture. Such is the fate of a research without theory, one which ignores its own logical and phenomenal roots and is typical of a human society in which "having" has taken over "being", action has obscured thinking and profit has replaced truth.

This avoiding the very roots of the problem is not just the result of an extreme fragmentation of knowledge, of ever growing barriers between cultures, behind which so much of the present intellectual laziness, disguised as specialization, hides. This is the refusal to critically investigate problems and methods, is the illusion of winding up knowledge into stereotyped little rules, the more vain and false the more they are dressed up with formal ornaments, that appear to be arranged in order to pretend some profound thinking.

The technical moment is not everything in the scientific research, where statistics is first of all interpretation norms. One must consider the logic of the inductive problem and the phenomenal context that must be analyzed again in the light of a new reading code. The scientific experience, the adventure of research, the application of hypotheses to the scrutiny of phenomena, are just what make us understand how the formal abstraction is the final goal of a process of analysis-synthesis aiming to identify that simplifying element, without which the mathematical-symbolic representation is a vain disguise. The technical-formal moment is certainly a fundamental landing, but it cannot leave out of consideration the logical-methodological moment, neither can it be translated into an inconsistent substitution of operationalism to thinking, with the illusion of ennobling a concept by the makeup of abstracts symbols. It is an easy fashion, so widespread that has created outward symbols of an assumed scientific prestige, has even dictated hierarchies which, because of the meaning attached to them, sometimes remind me of the "pecking hierarchies" amongst hens, whose mysterious social meanings have been explained by ethologists.

Regarding the method, even in the medical sciences the eternal duality concept between real and rational, concrete and abstract, inductive and deductive, empirical and theoretical, reappears. These alternatives go all the way through history and find in statistics their methodological setting. The *grand siècle* literature deals with these alternatives with anecdotes related to Descartes and Bacon: the first described tucked up in bed, keeping warm with the help of a huge stove, while writing his elegant *Discours*; the second, while taking a dive into the snow to prove the refrigerating effect on a chicken, until he caught a cold that ended up being fatal. By this curious thermal contradiction (Descartes will also die of cold) it is allegorically shown the antithesis of rationalism and empiricism, of theory and experience, and at the same time the intellectual twofold aspect of statistics, which is abstraction as well as observation, logos and *empeiria*.

It was precisely the difficulty of setting both the moments within the actual reality or medical and biological research, that arose the fiercest questioning and the most successful methodological solutions in statistics: profound basic queries regarding the meaning itself of the scientific knowledge and the heuristic reason for each cognitive technique to exist; non contingent solutions destined to become really and truly theories. These laid again the foundations of knowledge in which, for centuries, to a "physiology" ready for a kind of measuring experimentation and some daring mechanical analogy[3] it was opposed a choice of speculative activities bordering with myth, wherefrom a coherent research on matters struggled to get started. An entire false culture, full of pretense and mystery, found excitement in nominal listings, true monuments of verbal erudition. The naturalistic and medical way of thinking often lost the meaning of facts and ideas in the darkness of those classifying mazes.

One cannot avoid to think of the Manzonian character of don Ferrante, of that classification of his listing the genera of things in *rerum natura*, which seems to be the witty invention by a writer, but is instead the copy of a page taken from a medical treatise:[4] a cultural summa that from the seventeenth century – the century of Galilei, Kepler, Newton, Boyle and Harvey – stretches up to a kind of medical and naturalistic eighteenth century; a preliminary sketch of scholastic philosophy that found "qualities" in listings.

4. This is neither the time nor the place for retracing the logical and historical genesis of a scientific way of thinking more and more oriented towards objectifying living phenomena. I must, however, speak briefly about this extraordinary intellectual adventure. It is the time for great debating, one involving the very meaning of the statistical method. A debating of which one must know the precise terms, and this for two reasons: because it is very modern in its content and because nowadays it is not so much explicit in its form, it is more like a corporative uneasiness, rather than an open confrontation of opinions.

Let us, therefore, go back to the mid nineteenth century. Following a path largely alien to the becoming of quantitative sciences, naturalistic research is conquering the biological idea of variability, and physiology, an already experimental discipline, is still struggling within the grasp of a suffocating contradiction: the more it advances by imitating the criteria of physical sciences, the more it stumbles over immanent differences. Evidence of this is the scientific work of Claude Bernard, the founder of modern physiology, the author of the most classic publication on the experimental method in the biological and medical research. It is worth to stop over those pages for a little while, not only because they mark the highest moment of methodological discussion in the medical field – a rethinking of methods and contents as ever before – but because the problems of that time are those of today; the terms of that debating are surviving, sometimes without being expressed, in the scientific happening of our times. In his *Introduction à l'étude de la médécine expérimentale*, Claude Bernard maintained the necessity for

[3] Heart as a pump, blood circulation as a hydraulic model, many such studies had their place more in the field of hydrodynamics, optics, acoustics and even mechanics, than in the living sciences. After all, the mentioned fields of sciences were chapters of physics of those days.

[4] Viani's "Dialoghi su i rimedi efficacissimi per guardarsi dal male contagioso" (1630), of which Manzoni gained knowledge thanks to one of Achillini publication.

translating vital processes into numerical laws, for bestowing upon living phenomena the role of the Galilean proof of an hypothesis. He wrote: "The application of mathematics to natural phenomena is the aim of science, because the phenomena law must always be expressed mathematically". Measure, therefore, in order to put it mathematically, as done in physics and in astronomy. In physiology, however, to measure meant to accept the reality, with which Quetelet (from a formal viewpoint) and Darwin (from a heuristic viewpoint) had to deal. When *Introduction* was published in 1865, the intellectual world had already been buzzing for six years about *The Origin of Species*, a publication destined to the remaking of biology. Darwin had understood the individual variability amongst species and drawn a casualty oriented and population based image for the evolution of human entities. As the Nobel François Jacob said, such variability is one of the main motors of evolution and without it we should not be in this life world. Claude Bernard, instead, used the physics of his time as a model: he asked for univocal data, precise measures, and invariable facts. "Statistics can only offer probabilities" stated, in an unexceptionable way, that determinist, eager for certainties and unaware of what was about to happen to science. "The law of large numbers never says anything about a single case" – he concluded with obvious logic, adding soon after, as in extreme defense "In the indeterminism there are no laws".

At the same time, when *Introduction* was being printed by a Parisian *imprimérie*, in Brünn, Gregor Mendel, yet unknown, was illustrating, before the local Naturalistic Society, his discoveries concerning the biological memory linking generations together, the laws of an indeterminism essential to life, according to a combinatorial algebra that no one would have been able to understand. Indeed, Bernard wrote that "All the morphological generalizations (...) which represent the supporting point for the naturalist are (...) not enough for the physiologist and the physician". A coherent argument, yet, also the physiologist and the physician avail themselves of comparison values and conventional syntheses. Statistical averages – he added – "... confuse by wanting to bring together and falsify by wanting to simplify (...). They only dress results with a misleading accuracy". Exactly, accuracy; that need of accuracy which leads to "... not understanding how the statistical results can be called laws: because (...) the scientific law is based only on certainty and on absolute determinism, never on a probably". It was, however, in those very years that measurement and probability began to become inextricable realities in the same physical sciences taken as model by Bernard. In him one senses Comte's lesson, even more than Laplace's influence. (In fact for the leader of positivism, the probabilistic analysis of reality was considered as a "shameful scientific aberration", an "absurd logical utopia", a "philosophical monstrosity").

From here arises the refusal of any statistical intuition of reality, the ironical attitude towards the laws of tendency ("true, in general and false, in particular") and the disdainful anathema against the method of variable phenomena. Had Bernard read the *Lettres sur la théorie des probabilités*, that Adolphe Quetelet had published in 1846, he would have found, between the seventeenth and the twentieth letter, a desperate attempt to put some order in that disorder, to give a law to that irregularity. If he had known about Mendel's numerical findings, about the dice game, by which life passes on creating variety, his anxiety for certainty and his desire for determinism would have rejected that new science based on chance. He seems unable to grasp the profound meaning of variability: that natural variability he always met in his daily work as a physiologist, that variability which was made non-essential by the astronomer Quetelet, who turned it back to the formal theory of errors, and which, instead, the naturalist Darwin elevated to *raison d'être* for the evolution of species. In Bernard there is a coming back of the themes of the long diatribes which already arose in the first decades of the nineteenth century about researching for figures in medicine, of the disputes created in 1828 by the *Recherches sur la fièvre typhoïde*, that were presented by Pierre Alexander Louis at the Parisian medical academy, amidst general hostility. This is the point at which I intended to arrive. Louis had dared to introduce his "numerical methods" for a diagnostics based on the listing of symptoms and findings and on the typological identification of clinical pictures: in order to identify the constant from the variable, unity from plurality. This meant controversy, harsh controversy. The *Académie de médecine* and the *Académie des sciences* would shortly become animated places of excited debating between those who supported, sometimes roughly, diagnostic suppositions and therapeutic rules with numerical data and those who, with equal tenacity and belief, appealed to the unrepeatable singularity of the clinical case.

It was a dialogue between positions as much legitimate, as irreconcilable. One supported the need of defining laws in human pathology, and laws transcend contingency. The other objected about the inevitable particular aspect of each diagnosis and each therapy. The supporter of the former position often brought the statistical way of investigation down to a kind of childish collection of records, one not often helped by a coherent methodology. The advocates of the latter seemed to ignore to what extent the formulation of their diagnoses and prognoses depended upon an implicit synthesis of previous experiences. The debate is still a current one, the antithesis deserves more thinking, must be reconsidered. We are quite aware of the effects of the controversy between those who, naively, trusted a kind of empiricism without hypotheses and those who maintained that vital conclusions had often been drawn on the basis of few cases, while more important findings had no outcome. That debate seems to put forward the epistemological discussion of our times, regarding law like suppositions and casual links. (In fact, during an epidemic in Gottingen, Röderer and Wagler laid down an anatomical-pathological theory on thirteen autopsies, while the coincidence of two thousands cases of haemoptysis and lung tuberculosis were considered by many as a precarious concomitance. It is an old story. It is the history of science: behold Tyco Brahe, gathering an infinite number of astronomical measurements and on this basis rejecting the Copernican hypothesis; behold Kepler, building up the celestial mechanics on some of those figures).

In the naive attempts by the first "numberists" there was a new awareness: the renunciation of the absolute, the idea of the empirical force behind non-totalitarian statements. Another physician, Jules Gavarret, in 1840 published in Paris the *Principes genéraux de statistique médicale*, in which he translated into theory a methodology of the medical research within individual differences. Gavarret did realize that the empirical claim of natural variability, of accidental fluctuation was pending on every statistical synthesis. In dealing with this struggle with chance, he had arrived at daring and refined applications of Poisson's "*erreur possiblé*", by taking a multiple of it as neighborhood of reliability for the statistical constants, and basing on it a criterion to evaluate the differences among different groups, sometimes excluding from the casual ambit the non-coincident result of two different therapies, some other times assigning to spontaneous chance the, however different, impact of an epidemic in the two sexes. That is an ambitious way of reasoning, followed by echoes of presumptuous jeer from the majority of people. Nobody would consider those fascinating statistical outbreaks and a long methodological night would fall upon medical sciences. Paradoxically enough, the experimental and quantitative rigor of Claude Bernard contributed to that; his disdainfully rejecting any statistical image of life and health phenomena. A clever attempt for *ante litteram* analysis of statistical significance was outlined in those first probabilistic treatises of epidemiological and clinical-therapeutic data. It is an inductive concept with perhaps only two timid and remote previous scientific results: the research on comets by Daniel Bernoulli and Maupertuis's calculus on the non-accidental aspect of the presence of sixfingered individuals in the genealogy of a Berlin family.

In strictly carrying forward the epistemological model of a knowledge that could not prove totally right in the phenomena of life (it soon will show its limits also in the sciences of which, for over two centuries, he supported the conceptual layout), Claude Bernard remains entangled within an incurable contradiction. He sees variability and refuses the mean value ("... never to be found in nature"); he sees variability and solves it in the "type", that is also mean value, that is always abstraction. Yet, the subtle irony of this determinist, searching for absolutes in the silly application of statistical calculation, often hits the target. In his derision of the extravagant fiction of a "*urine moyenne européenne*", a witty prelude for many future anecdotes, there is the warning that a statistical parameter – an extreme form of abstraction and synthesis – does not draw its only justification from a formal principle, and that the *reductio ad unum* implies the critical assumption of the plurality of the single values. Otherwise, variability is only disorder, so as the multiplicity of cases is confused empirical experience. Goethe wrote: "What is the universal? It is the single case. What is the particular? It is millions of cases". In Claude Bernard one can find the awareness that science reasons on models that models are intellectually constructed sections of reality, where a complex of circumstances is isolated and a phenomenon is represented in an extremely simplified form. Although he is not aware of, such are the statistical models as well, because they settle the different singularities in the spirit of classic experimentation.

5. In fact, there is no epistemological antithesis between experimental method and statistical analysis when this last one abstracts and formalizes in the Galilean

scientific tradition; which means, indeed, to make artifacts out of phenomena, to reproduce reality fragments by eliminating the elements that make them contingent, special and unrepeatable. Statistical analysis obtains this abstraction by choosing variables and setting up classes of ideal continuity with the Platonic idea and the Aristotelian category: abstract typologies of events, of which empirical reality offers examples that are all the same, yet all different. Hence, the common name and the natural number; hence, the elementary statistical data: a natural number of objects gathered together in the common name of a classifying paradigm. The classifying listing of cases (the figures of which, so much of the self-styled "applied statistics" is too proud) is the natural preliminary of every statistical analysis. Which means to consider those data as empirical determinations of an aleatory variable in a virtual universe that allows other possible determinations, with defined probabilities. (If Shakespeare's Hamlet could say "There are more things in heaven and earth than are dreamt of in your philosophy", the modern methodologist offers the researcher a philosophy containing more "things" than do occur in the empirical reality). The statistical intuition of events has offered the physics of elementary constituents and the biology of elementary processes the key for a reading of the phenomenal reality on the basis of populations and chances. Similarly, for the research on man's health, that sets his models in the immanent variability, founds its theories, and finds the presuppositions for turning pathology into science and clinical medicine into a coherent art.

Which rational stand can the medical sciences take in the light of individual variability? Must they abstract typical pictures, to which compare the single? Must they divide the values interval, so as to define a dichotomy in relation to the threshold (before it the "normal" beyond it the "abnormal")? Must they acknowledge the changeable continuity of moving from physiology to pathology, so as to solve the antinomy of a medical practice necessarily addressed to the individual as *unicum*, because organic synthesis of inseparable forms and functions, and of a medical research which is necessarily aimed at identifying, in each variable, rules, classes, types? All this and a lot more.

But, first of all, one must draw the line between individual and statistical laws, between the properties of the single and the properties of the group. One must understand that in the non-deterministic contexts, once the universe model, that stood for three long glorious centuries of science, became obsolete, a statistical constant is not a vague epistemological expedient, a temporary approximation of an otherwise perfect knowledge, as some kind of philosophy of science still seems to believe, almost ignoring the change that took place in the scientific thinking, from Darwin to Boltzmann, from Mendel to Born. It is an intuition of nature inspired by the kinetic theory of heat and even more by the biological evolutionism, by the mechanics originated from Planck's and Einstein's quanta and even more by the genetics developed on Mendel's and Johannsen's numbers; in fact by all sciences which learned to deal with "populations" and to identify their statistical properties, to observe events that escape any individual necessity, collective phenomena in which an immanent fortuitousness plays its role.

"It is hardly necessary to taste the water in order to conclude that the sea is salty" – the Port Royal logicians used to say. No doubt about it. Yet, if the subject of a research was, for instance, the salinity gradient of waters, then it would be the case of a quantitative, and not invariable, reality, of the requirement for multiple trials. Hence, in the statistical phenomenology, the repetition of trials, that has mainly a psychological value in testing absolute terms, has an objective, heuristic role. Claude Bernard realized this. The more he tried to base medicine on totalitarian, categorical terms ("Each A is B") and on, univocal hypothetical statements ("If A, then B"), the more he met with realities which contradicted every syllogistic schematism. He then stopped over typologies for which the empirical cases are repetitions of the same event. Sarcastically, he observed: "There is no need for statistics in order to know how much of oxygen and hydrogen are required to make water (...) or how often the cutting of a specific nerve causes the paralysis of a certain muscle". This is true, as it is for any other "totalitarian" law.

But it is not so if the law is a "statistical" one. How could one understand, without statistical analysis, the numerical proportions of genotypes resulting from a cross-fertilization of heterozygote? How could one deduce "rules" from phenomena subjected to the immanent variability and to that inertial play of large numbers, from which they draw – as statistical properties, laws of tendency – a nomic inductive value that seems to elude Hume's logical trap? In this thesis, now typical of all positive sciences which met with the phenomenal variability, the consideration on statistical knowledge finds a first answer to the questions regarding induction. Virtually, if only one normal human being is the expression of the chromosome number, in the diploid state, in *Homo sapiens sapiens* (an inductive inference from the individual to the class in a uniform context, in which Virgil's "*ab uno disce omnes*" rules: each further observation is a proof of invariance), one hybridization 'only is not enough for the discovery of laws governing character heredity. In this case the empirical plurality of cases has a particular epistemic meaning and relates to the very nature of events: casual combinations of the possible allelic states of a hereditary character, in fact, express "statistical phenomena" for which, in the very context of the discovery, the variable plurality of the determinations is essential. It is the plurality from which laws are derived.

Like any other science, medicine as well uses "singular" and "plural" statements, "deterministic" and "statistical" properties. Certainly, in both cases, the antinomy concerns more the gnoseologic meaning of multiple observations, than the actual process of research. Yet, the conceptual difference remains and goes deeper into the investigation on diseases, as well as their clinical identification in an individual. The pathologist's statistical terms translate into the diagnostician's probabilistic evaluations, and only in the case of strictly bi-univocal connections they can lead to certainty. By now, this is the assumption of phenomenology which cannot be related to the classic rules of causal logic. This thesis offers new reasons and new perspectives for thinking over again the most discussed problem of knowledge philosophy and marks methodologically all positive sciences, starting from the research on the living, from the studying of man. From there, with Darwin and Mendel, came the first awareness of an immanent variability in nature, from there, with Quetelet and Galton, came the idea of a conformity to distributive laws, from there came the concept of statistical phenomenon. Within this context, statistics becomes cognitive technique, inductive canon, nature intuition. If in the past, in its dealing with classification, it could represent the methodological garrison of disciplines which could not be directly tried by experiments and turned into mathematical expressions, and by a descriptive typology preparing to change over to quantity, it brought into them an empirical world for reaching the objectivity of Newton's natural philosophy, now, having conquered variability, statistics sets up the canons of a new logic. That is logic of the probable for a science giving probable formulations, in the presence of the fortuitousness of reality. From elemental particles to galaxies, from nucleotides to the man, the presence of chance, with time dimension, has entered by now into the phenomena.

To an epistemology struggling to overcome the schemes of the syllogistic deductive tradition, the statistical method has offered an immediate inductive reason in a new image of reality – the immanent variability – and tendency laws, collective properties, because of the need to draw synthetic expressions for knowledge and action. It is the crisis of a world concept, Claude Bernard's drama, and his insoluble contradiction: while he daringly struggled to drive the phenomena of human physiology and pathology to the physical deterministic paradigm, a new theoretical paradigm, arisen from life phenomena, came to subvert the moral and intellectual foundations of that reassuring vision of nature. From this originates the incomprehension between Bernard and Gavarret, a kind of persistent discomfort. It comes from this the invitation to humbly thinking over again about principles and rules, canons and instruments, deeply considering the *raison d'être* of the method of variable phenomena in the adventure of knowledge, finding again the heuristic meaning for a logic of propositions, no longer totalitarian, statements, no longer causal or univocal: a logic of statistical terms and plurivocal probabilistic assertions.

6. This method is still based on the line of the Galilean philosophy, the meeting point between theory and practice, phenomenal reality and abstract model, on the logical canvas of that hypothetical-deductive canon which, particularly in the *modus tollens*, has laid down the positive disciplines. While Galilei theorized the criterion of "sensible experiments" and of "necessary demonstrations", Bacon suggested the comparison of different situations for the presence and, respectively, the absence of a specific circumstance: a criterion of which examples, more and more significant, can be found in the "*grand siècle*" and in the "Age of Enlightenment" and one which, in the nineteenth century, will be codified in Stuart Mill's logical system. If the Galilean Redi did compare two groups of ampoules containing organic tissues, some closed and some others open, in order to prove the *ex ovo* generation of insects, the Baconian Petty had tried to compare the survival rates for two human groups, only one of which received medical assistance, in order to assess doctors' ability and the benefit of their treatments. Just as Redi's seventeenth century experience, that was to repeat a century later in Spallanzani's researches, carried out as well on spontaneous generation, and again another cen-

tury later, in Pasteur's experimental adventure, mirrors the classical principle, so Petty's one puts forward the most pliable re-proposal of that principle in a quantitative and variable context. It is the methodological paradigm of an analysis becoming more and more a statistical one of reality, one which, from the nineteenth century has begun to be the nerve of entire sciences. One can find it in the aetiopathogenetic investigations by Louis, Gavarret, and Farr, as in the botanical experiments by Charles Darwin: each of these scientists equally concerned to free, first of all, the testing time from the most dangerous logical doubt for a research set into the natural variability: the suspicion of an accidental origin for experimental, non-coincident results.

The hypothetical-deductive canon still goes after the methodological pattern, but, while offering a new semantics, variability and chance impose a new syntax. The formulation of the new knowledge does not imitate any more the "causal hypothetical statement" of the classic science (a science of invariable properties, repeated events, deterministic uniformity), but the "statistical hypothetical statement": "If A then $B_1 \cup B_2 \cup ... \cup B_i \cup ... \cup B_m$ with respective probabilities $p_1, p_2, ... p_i, ... p_m$ such that $\Sigma_i p_i = 1$". This is not the epistemic remedy of classic determinism: it is the canon of a strictly statistical knowledge, one that has replaced necessity with possibility, univocal way with plurivocal ways. The Fisher's "significance analysis" – wherefrom so much of the biomedical research has drawn and draws its intellectual instruments keeps in line with that knowledge and turns the "null hypothesis" into the preliminary of a natural philosophy in which chance has penetrated, of a research that has met with the spontaneous variability of the elemental events of life and matter.

When medical research deals with random trials when it uses the combinatorial solutions of the design of experiments when it compares the statistical properties of different empirical groups and checks their significance, it finds a first help in the Fisher methodology, somehow coherent with the rules of the scientific tradition, although transferring from the logic of certainty to the logic of probability eliminates any causal necessity. It is still the scheme of hypothetical assertion, but its antecedent is the hypothesis of total chance (H_0), supposed as true, and the consequent is a random variable upon which a unitary probability is divided: the variable defined by the disjunctive and exhaustive plurality of the possible results. Therefore, "If H_0, then $X|H_0$ with law of probability $f(X|H_0)$" (being $\int_{-\infty}^{+\infty} f(X)\mathrm{d}X = 1$). To refuse H_0 when the empirical observed value, X_E, is within the region of $X|H_0$, to which a minimum probability is attributed, thus becomes a conventional deciding factor in the praxis of an experimental research woven with statistical assertions: the assertions of a new logic, regarding which the traditional lack of symmetry between rejection and acceptance disappears. The principles of the Galilean science, hence, live again in the statistical method, so as to overcome the contradiction on which Claude Bernard's experimentalism collapsed.

But, actually, it is mainly the lacking of Galilean natural philosophy that appears to affect those who, also in the medical world, today resort to statistics as a rational witchcraft, indiscriminately using techniques and leaving it all to computing, while nervously waiting for the fatal answer to the "significance" of a result. However, those techniques, derived from the biological research, have become part of the logical asset of many researchers, in the biomedical fields as well; however, the research on living phenomena, at the time of the most significant discoveries of this century, has decisively resorted to the spirit of those techniques. Besides the initial fascinating and probative questions regarding the hypnotic effects of a drug, analyzed by Gosset's test, one must recall the intuitions of the first geneticist, like Bateson and Punnett, when dealing with statistical results that did not comply with the Mendel independence (beyond the limits of sampling error), and their drawing, from one of Pearson's criterion, one more reasons for the refusal of the hypothesis of total fortuitousness of that discrepancy, for their turning, along with Sutton and Morgan, the statistical significance of the difference between theoretical and empirical data into the introduction to a "chromosomal theory". One must also recall the puzzlement of hematologists, in the Twenties, facing growing contradictions between the blood phenotypes of family genealogies and Landsteiner's genetic model. One must think of the decisive statistical confutation of the hypothesis of fortuitness for that discrepancy and of the following research for a genetic scheme that is Bernstein's triallelic model.

Yet, the statistical control of the "null hypothesis": H_0 – to be allowed within the context of a falsifying philosophy – always conceals the danger, when the hypothesis is not confuted, of underestimating the experimental information, of leaving aside intuitions not without foundation, of improperly diverting a line of research that might be successful. When one adopts the value of probability that an accidental process produces the observed result, which sometimes happens, like a kind of "litmus paper" to which ask for definitive answers – and not as one of the many judging elements that a researcher uses to back up his tests – this becomes the fatal limit for the techniques aiming to confute H_0. On the other hand, it is true that a null hypothesis, that has not been rejected, remains open to the confutation of new experimental data: the new data of an uninterrupted research. But, when one does not imply the instrumental provisional character of the missed confutation of H_0, one indeed risks compromising the "positive" hypothesis of the scientist. Moreover, who does not stick to a rigorously falsifying criterion, within the hypothetical-deductive context, and admits the possibility of "accepting" the hypothesis of total chance for an event E, ends up accepting the probability $p(E|H_n)$, that an event similar to the one observed occurs by chance instead of probability $p(H_0|E)$, that the observed event has to be attributed to chance.

Man has always sought security by resorting to quantitative methods (traditional expression of objectivity); and these, by now, offer him only probability values. If this has contributed to give science a more "human" aspect, it has also given a new space to the irrational philosophies and horoscopes. And by horoscopes, to some statistical criteria, 'to which it is too often resorted as to augur rituals, transforming the "statistical significance" of a result into an almost thaumaturgic expression of undeniable truth. The ritual of significance tests is certainly essential to the hypothetical-deductive knowledge in probabilistic contexts, but this because it helps and stimulates the critical thinking, not because it pretends to put it to sleep by substituting a kind of methodical lottery to the methodical doubt.

Far too often statistical calculations are seen to appear in the middle of dissertations which could do very well without them, and which actually do without them. What it is not always clear is if such exercises respond to aesthetic requirements or somehow pretend to be propitiatory formulae spells. In either case, science, along with statistics, ends up being: humiliated. The danger lies not in the statistical methods as such, but lies in the naive attitude of so many who fall for the very beautiful tests: the trade instruments of every cautious researcher, which one hates to see sometimes turned into an alibi for the lack of culture. It is in those hypothetical-deductive canons that lies the never ending shifting of the methodological dialogue between hypothesis and data, experience and reason, born with science and still the focus of a wide theoretical debate. However, as the classical experimental method cannot give ideas to those who have none, similarly, the statistical method, that considers again the principles in new contexts, has no reason to be, except because it is control of hypothesis. To hide the vacuum of the hypotheses behind algorithmic curtains of some assumed modern bio-mathematics, which causes biologists to cry and mathematicians to laugh, is no less sterile an activity than patiently piling up data, hoping for a sudden revelation, while doing so believing (or pretending to believe) to honor the logic and the method of statistics: some collections, to speak the truth, are to statistics what a quarter of an ox is to biology.

There is a significant fact that must attentively be looked at the settling in of bio-statistic disciplines, from biometry to genetics, into the curricula of the medical faculties. This is particularly true for a country like Italy, where (having been here, more than anywhere else, so devastating the effects of what Fisher called the "unfortunate misunderstanding") the university teaching of the statistical method has been confined for a long time within the non-scientific disciplines. Provided the researching instruments are taught with critical awareness of their logical premises and their empirical limits, so as not to turn statistical analysis into a kind of simple ritual, closer to the obscure divination of the soothsayer than to the Galilean reasoning of the scientists.[5]

[5] One hopes to never have to regret times past when one was happy enough to narrate the virtues of quantitative knowledge by adding to them some demographic curious note: e.g., that the numerical proportion of sexes in the birth of human species, oscillates around a stable ratio, with a slight systematic prevalence of the male sex; or that at population census, women declaring to be 39 years old are many more than those declaring to be 40; and so forth. These are commendable news, destined to rule out, the first, a popular prejudice, and the other the reliability of official documents.

To recall the logical foundations and the scientific roots from where the inductive statistical techniques draw heuristic meaning, does not mean to deny their instrumental, contingent validity. To do so would be like to prevent the surgeon from using his scalpel because it could be used as a murder weapon. Besides, we all, for Fisher or against Fisher, "believers" or "non-believers" are tempted to use certain statistical-inductive techniques in the empirical event of research. I could not, indeed, say if these people's motivation might, by chance, be found in the ironical answer given by Niels Bohr, the leader of "Copenhagen Physics School". To those amazed people asking why he, the rationalist free of any superstitions, kept an horseshoe on the door of his house in Tisvilde, he answered: "No, positively I do not believe that a horseshoe can bring you good luck, but, if I have put it there it is because I am told that It brings good luck even to those who do not believe it".

7. Biomedical research finds an essential principle in the hypothetical-inductive canon of science and in the intellectual instruments of its statistical reproposal within non-deterministic phenomenology, clinical strategy can find out some food for thought within the inductive procedures of statistics. Diagnosis and prognosis, as they are uncertain inferences, appear to recall and sometimes to use intuitively the elementary eighteenth century relationship between probabilities, known as Bayes-Laplace's theorem. A theorem that since a few years has found its space, in relation to diagnosis, even in some glossy magazines dealing with that area. The most subtle theorizing on probability as subjective expectation takes the theorem as a rule for inductive inference, for changing an *a priori* opinion in relation to experience: the habit typical of life, knowledge, action, wherefrom the most controversial and difficult problem of logical thinking – the famous "Hume's problem" – and the hardest and most effective chapter of statistical methodology – the inductive inference – were drawn.

As well known, if in a population the probability that one suffers from disease a is $p(a)$, and the probability that such disease produces the symptom s is $p(s|a)$, according to Bayes-Laplace theorem, the probability $p(a|s)$ that an individual with symptom s suffers from disease a, is proportional to the product: $p(a) \cdot p(s|a)$; i.d. $p(a|s) \propto p(a) \cdot p(s|a)$.

If $p(s)$ is the probability that symptom s appears in an individual, the Bayes-Laplace relation is: $p(a|s) = p(a) \cdot p(s|a) / p(s)$.

If diseases which could present symptom s are $a, b, ..., i, ..., m$, the equality $p(s) = \Sigma_{i=1}^{m} p(i) \cdot p(s|i)$ holds.

All these probabilities can be assumed as statistical frequencies, *i.e.* as numerical ratios of a part to the whole.

To put it into non-mathematical terms, the Bayes-Laplace's formula assesses the probability of disease, given the symptom, as proportional to the product of the symptom probability, given the disease, multiplied by the probability of the illness occurring.

In scaling the validity of contrasting hypotheses and assessing the arrangement of observations and opinions under the leveling effect of experience, the Bayesian algorithm offers an immediate and simple logical pattern for a problem of hypothesis probability, such is in fact diagnosis. This is a thesis which gets repeated in text after text, one which received a clear conceptual mark by Corrado Gini more than thirty years ago[6]. However, for some time now a new schematic representation of diagnosis tends to prevail and by many quarters the idea of a Bayesian clinical reasoning is now rejected.

8. It is however clear, in my humble opinion, that it will not be possible to talk about inductive inference until one will thoroughly know the structure that makes the daring leap from the known to the unknown, that is the brain: without understanding its process of evolution, its phylogenetic and ontogenetic adaptation to the living. It is the relationship between environment and species that conditions the evolution of the latter, it is the individual's adaptation to a natural and social environment that develops its inductive abilities and shapes its instinct to generalize, to decide regarding the destiny of an inferential choice. It is on this natural reality, on the awareness of the unavoidable uncertainty of further extension of knowledge, that the theoretical debate about the foundations of inductive logic has to be reconsidered and same for the very technical-formal development of statistical methods for the approximation of properties observed in a limited number of events. It is still largely an obscure question how the central nervous system takes in and elaborates the data of the external world, absorbs and translates images, chooses and works out the real: but, first of all and more than philosophy the question involves biology and neurobiology. It will be this last one to define the mental reflection of phenomenal data, the classifying synthesis implicit in the taking shape of information within the logical asset of species: a selective and reductive process in a structure that receives and deciphers, translates and orders, extracts analogy from diversity, regularity from casualty; one which is capable of carrying out abstractions and analogies, inferences and selections.

The antitheses between idealism and realism, objectivism and subjectivism, empiricism and rationalism, nativism, and sensationalism, the debates regarding the idea of probability, of induction, which permeate the history of the philosophical thinking, will have to be tested on this organic reality. From species to species, the Darwinian "fitness" of the various cerebral assets cannot be but the developing expression of a long evolution of trials and errors, successes and failures, through which they have been genetically defined into even more complex organized in a systems of neurons, due to the interplay of the accidental variability within species, capacity and abilities, continuous reciprocal updating between the

[6] C. GINI, *The statistical bases of diagnosis and prognosis*, "Acta genetica et statistica medica", III, 32, 1952.

sensorial apparatus of perception and the central unity of elaboration. It is in fact, the process of knowing and reasoning, as results of their codifying in the alphabet of nucleotides in the great adventure of adaptation, that has forced the choice within the enormous number of information, the order in the complex disorder of events, the abstraction of invariants from unrepeatable singularities, thus conveying things to symbols of events considered as numerable and mathematically treatable. It is the natural history of human intellect that is becoming more and more capable of seeing and foreseeing what the evolutionary opportunism of the species has taught it to see and foresee, capable of reacting to the stimuli which have become relevant in the acid test of phylogenies, of developing, in ontogenesis, abilities essential to the fitness of the species. Then, nativism? Not, certainly in the sense of innate information. Perhaps in terms of "innate" aptitude, or even better, aptitudes which have been genetically codified by the evolutionary history of the species: the aptitude for seeing uniformity and similarity where other cultures, other species, see none?

If induction is an instinctive act which is refined by the natural process of evolution of species and by the experience of living beings within their world, nevertheless it is not devoid of sense to search for its rational roots in the scientific research as well as in diagnostic investigations and therapeutic decisions: uncertain inferences which make use of statistical properties of populations.

9. The medical diagnosis, as inductive process, choice among hypotheses, can draw from Bayes-Laplace's formula an essential pattern for the different contribution of pre-existing opinions and observed data in reaching a judgment. However formal, induction maintains its logical *status* of non demonstrative inference, more so in the clinical field, where no formula succeeds in reproducing the complex and sometimes eluding act of diagnostic choice; where the true problem lies in the comprehensive clinical picture, that cannot always be taken easily back to an elementary grammar of symptoms, diseases and homogeneous cases. Those who claim to put forward the Bayesian relationship between probability – $p(s|a)$ – of symptom s, given illness a, and probability $p(a|s)$ – of the illness, given the symptom, like using a magical modern mechanism for "automatic diagnoses", do not appear to be totally aware of the unavoidable individual component.

That formula rather represents an *ab externo* simplification of the diagnostic process: the symbolic expression of what is done naturally, more than a prescriptive rule. The knowledge of the intervening probabilistic factors can supply the diagnostician with an indispensable critical awareness and make him perceive the different gradual stages played in the synthesis by theory and clinical data, general information and specific situations. Moreover, it also allows to understand the risk of over – or under – estimating probability $p(a)$ of the pathological factor arising.

In the setting up of a logical-empirical criterion for the distribution of a unitary probability in a disjunctive and exhaustive multiplicity of diagnostic hypotheses –

so as to propose again the classic problem of "probability of causes" – the Bayesian algorithm builds up a schematic network (perhaps one too insidiously schematic) of the clinician's inductive process when risking an inference: a necessarily reductive vision excluding any complexity and gradualness (only one symptom, a more or less sensible one or a more or less specific one, or a syndrome taken as a whole, which may only be present or absent and, if present, a non-necessary expression of one and only one form of disease within an alternative plurality of forms considered as reciprocally incompatible although all compatible with the symptoms picture). A vision which postulates inadmissible independences and restricts the choice to a closed causal space of mutually excluding hypotheses in which no disease is necessary condition of the observed symptom and this one is not sufficient condition for any disease.

Having accepted a scheme of symptoms and defined the congruence of each symptom in relation to each one of the pathological factors forming an alternative plurality of diagnostic hypotheses it is any way possible to find again the logical-formal rooting of some criteria of the medical practice in the link between "direct" and "inverse" probability. Each diagnostician may value differently the single "Bayesian" components and one is not surprised by Gini's conclusion after a survey amongst doctors and an analysis of papers by famous clinicians, starting from those by Augusto Murri, glory and pride of Italian medicine and the University of Bologna: no one seems to be able to define how he mentally combines objective and subjective elements of the diagnostic supposition.

10. By laying down a scheme for the dialectic game of data and hypothesis in relation to the probative aspect of data for the hypothesis and to the acceptance of the hypothesis in itself, the Bayesian algorithm makes evident the importance of values attributed to initial probabilities and reveals the statistical conditioning – when any alternative has not been excluded from the start – of inverse probabilities $p(i|s)$ to the experimental reality from which the "direct" probabilities $p(s|i)$ are drawn. If facing the same empirical evidence, that is the same symptoms, different physicians may reach different diagnoses, due to the different personal evaluation of *a priori* probabilities $p(i)$, this is because the most unstable and ambiguous component of the diagnostic moment is expressed in these ones. Such appears to be proved also by exaggerations of the probability of occurrence typical of specific situations and by the very behavior of specialists who tend to overestimate (though involuntarily, in good faith) the initial probabilities of pathological entities attributed to the respective specialized areas. Texts hand down the warning by a German professor to his students: "And above all you must remember that rare diseases are not frequently observed". Almost instinctively, the subjective partition of initial probabilities $p(a), p(b), ..., p(m)$ is modeled upon experience in a demographic context. Generally, the diagnostic evaluation will be more indifferent to the context the more such partition is uniform, more part of it the more it is differentiated.

However, clinical strategy cannot stop with the identification of the most probable diagnostic hypothesis: a result is involved, value systems intervene. First of all, the different seriousness of the possible diseases must be taken into consideration. Thus, the problem switches from the methodology of knowledge to the methodology of convenience. It is no longer a question of pure inductive logic: a decisional marking component is added which is, although alien to the diagnostic problem as such, always relevant and often decisive in guiding the physician's work. One thing is to identify a pathological form by its symptoms and another is to decide about a therapeutic strategy. To differentiate between the identification of a (probable) illness and the decision for the best treatment (e.g. the one involving minimum risk) is conceptually fundamental. If already at the time of diagnosis certain operational choices regarding opportunity and utility (and risk) of some tests may have a priority, they become essential at the time of therapy for the strategic choice: that is a decision aiming to maximum advantage and minimum risk. (These are the problems of modern decision theory, but the question is older: it's the famous debate between d'Alembert and Daniel Bernoulli, repurposed by Laplace, about smallpox vaccination).

In this light, every possible risk for the patient must be taken into consideration, including the danger linked to the therapy. It is no longer the case of the mere diagnostic conjecture, elements intervene which are extraneous to the inductive process. It may even happen that – when the therapies are not incompatible with one another – the therapy for a more serious illness is undertaken, even if this is less probable than a less serious disease. In such a utilitarian context, it makes sense to differentiate – as Neyman did – between "first type error" and "second type error" that is between the error of rejecting a true hypothesis as false and the error of accepting a false hypothesis as true.

If it is more dangerous not to treat an already existing illness or to treat one which does not exist depends on the illness and the therapy, as well as on the hope of survival and the type of life after this. Not to mention other elements of opinion, other assessments of values where ethical, hedonistic, ideological, religious, principles intervene, factors which are always tending to some aim and are often in contradiction with one another.[7]

11. In suggesting a critical re-reading on the theme of diagnosis of certain logical-formal development of inductive inference, these considerations offer the clue to set a distinction – I believe an essential one – between scientific investigation, cognitive identification and strategic decision. The first of these is the search for laws governing reality, the second is the typological attribution of an actual condition and the third one is the choice of a suitable behavior.

[7] Is it not true, in fact, that the very deontological moral which rejects any suggestion for euthanasia and cleverly prolongs extremely painful agonies as it considers improper to arrogate to itself the right to end one's life, agrees to the fact that a life can be shortened by removing a vital organ useful to some other person?

If the dividing line between knowledge and strategy is identifiable due to the evaluation of advantages and disadvantages occurring in the latter, the difference between the other two conceptual categories appears as more subtle, yet not of a lesser meaning. Deciding about a suitable behavior is one thing, identifying a circumstance is another and yet a different matter is to test a working hypothesis in relation to the phenomenal determinants of an event. The adoption of a therapeutic protocol is a decisional strategy in terms of utility and risk; the identification of the illness affecting a patient is a probable inference resulting in the allocation of a case to a phenomenal class. This, by itself, is not scientific research, nor the identification of natural laws, of statistical properties. In this case the scientific research results by inferring the quantitative links essential to the formulation of probabilities $p(a)$, $p(s)$, $p(s|a)$: probabilities used by the diagnostician, even if unconsciously, to go from s to a, finding out from a direct probability its "inverse", in relation to his own ability of synthesis and his grade of information about probabilities which are not invariant in time and place.

These data are drawn from the different domains of medical research and have a precise operational content in the individual situation. However, their heuristic extent keeps to pluralities, expresses statistical properties of populations; in such ambits they become parameters of phenomenal realities and sometimes of real and true laws, if conditions exist such as to reach a homothetic stage. Naturally, these are the laws of tendency, as for any event deep down into the individual variability, for any reality subject to chance games. Which is the statistical context of the medical scientific research: a research that cannot be but hypothetic-deductive development of hypotheses and comparison with empirical evidence, a research that cannot however provisionally, base itself but on theories which have not been denied by real data.

In comparison to these, the diagnostic act, though aiming at contingency and individuality, may become a critical control, a moment of proof, a testing of the evocative value of a symptom, the predictive impact of a data. Actually, the clinical practice works within the individual and the contingent, and the diagnostic act is the utilization of the results of scientific research, of values $p(s|a)$, $p(s)$, $p(a)$ found by pathology, semyology, epidemiology. Without this knowledge the diagnostician cannot infer $p(a|s)$ and he has no elements for choosing (comparing) the different probabilities $p(i|s)$, for $i = a, b, ..., m$, amongst the pathological forms considered to be possible. (The prevailing of one probability value, in relation to symptom s in the space of alternatives, will not be the only factor towards the definition of diagnosis; some pathological forms can be excluded following the identification of other symptoms with no connection to s). When symbol s also indicates, more than a specific symptom, a real and true syndrome, it is also true that symptoms and syndromes almost never have a univocal semeiotic value, leaving aside authentically pathognomonic symptoms. From this the complexity of the inductive diagnostic problem arises.

Which is not, anyway, exclusive of the medical practice? Not even the anthropologist, for instance engaged in the identification of the sex of a human skeleton, found buried in a ground, will rely on one element only. Just like the medical research has specified "constellations" of symptoms and precise hierarchies among these, so the anthropological research has identified a multiplicity of morphological and metrical characteristics, able of supplying tendency information. Just like the clinician, the anthropologist will try to use a myriad of signs (symptoms) and his inductive opinion (diagnosis) – a cognitive act that is not by itself scientific research – will be the less uncertain the more the report allows him to identify a multiplicity of findings. But his conclusion will be founded if some, however tendential, statistical relationships between sex and skeleton characteristics have been identified (this is the scientific research) – starting from the laws of variability and transvariation in relation to sexual dimorphism of the human bone structure – and if it mainly uses the general information on the history of human settlement where the finding happened. Towards these, the diagnostic act may sometimes be considered within the logical context of the justification. The analogy with the clinical diagnosis – as heuristic strategy – may perhaps be pushed so far as to identify the morphological characteristics with the symptoms and the metrical characteristics with the values drawn from the quantitative-analytical surveys used by the physician in order to support the diagnostic opinion. It is, however, mainly expressed in the examination of incomplete finds, typical of fossil remains. The anthropologist then finds himself in the shoes of the physician facing a partial symptomology and he has, eventually, to try to guess the sex from only one bone segment.

However forced, the comparison still gives a clue for the clarification of initial differentiation. The identification, *more statistico*, of the evidence of sexual dimorphism in the human skeleton is scientific research; the sex diagnosis in a skeleton find is cognitive utilization of the test results. The various scaling of a priori probabilities always uses multiple evidences. (As the physician, examining an anemic looking patient, will be more inclined to suspect a genetic fault in a microcythemic sense if he knows that the person is from a place where heterozygote for thalassaemia are widespread – a genotypic expression of selective event following past cases or malaria – so the anthropologist might find reasons to expect one sex rather than the other under the circumstances of time and place of which the bones being examined bear indirect witness).

12. Hence diagnosis is a logical-intuitive combination of various components which are subjectively perceived, if not objectively defined. Nevertheless, they are statistical quantities and, what is more important, quantities intervening according to an unconscious inductive mechanism, of which a relationship between subordinate probabilities may suggest a rational paradigm, but in what logic? That which prefers the most probable hypothesis on par with data? This is the problem. When the Bayesian diagnosis is translated in choosing the hypothesis to which the maximum or the *a posteriori* probability function corresponds. i.e. the

highest value $p(i|s)$ for $i = a, b, ..., m$ it makes sure, by a series of diagnostic actions, to hit the target with a frequency fundamentally equal to such a probability – obviously if probabilities $p(i)$ and $p(s|i)$ for $i = a, b, ..., m$, are not "biased estimation" – so as to do wrong in all the statistically "marginal" clinical cases for which the least frequent pathological entities and the least frequent combinations between symptom and illness occur. It is clearly true that the diagnostician is free to choose within the probable space of alternatives, but, as soon as the Bayesian inference becomes a standard rule – either followed by man or computer – the criterion tends divert the choice towards the most probable causal hypothesis. Which guarantees the practical predominance of correct inductions in a "*long run*" (as by Neyman) and at the same times the risk of making very serious errors regarding the individual. The Bayesian probability, although subjectively meant, draws inductive and predictive strength from the statistical weight attributed to the combination illness-symptom in the aetiopathogenetic case studies, codified by usual protocols.

Here comes then the Bayesian criterion as a diagnostic criterion the more efficient, the more the investigated subject, as to the considered characteristics, falls within the values with the highest probabilities $p(i|s)$. (That is called "pathological normality"). Thus comes then the above mentioned risk, the difficulty to draw the most unusual diagnoses, to solve the least typical cases, to infer a coherent individual and subjective translation of involved statistical variables by a constant re-adjustment of the *a priori* assumptions. If science, as Monod wrote, "can only search for invariants", it is natural that medical science – as all sciences – identifies typologies: "normal" is, in fact, a type. It is essentially important to codify phenomenal properties in order to present the ideal term of comparison for every science modeled on individual variability. However the clinical decision – that also finds indispensable information in the statistical properties or populations – keeps to the individuals, the variant. Opting for diagnostic hypothesis with the highest Bayesian probability, the physician hence meets with a "risk function" that has to be defined and graded for the decision of a therapeutic strategy. These are considerations presenting more ground for the already anticipated conceptual distinction: the distinction between scientific knowledge and decisional strategy, between search for truth and choice of convenience. The second alternative is not a rhetorical repetition of the first.

Diagnosis itself offers an example: it is a cognitive act, but it is not by itself a scientific research. This is even a more pressing distinction today when – in a jerk of managerial utilitarianism – the tendency is to see in the statistical method a kind of nonspecific chapter of the decisional theory. It is the same in medicine, where, in the light of discovery, it makes sense to doubt about the heuristic meaning for the substitution of decision to knowledge, betting to hypothesis, convenient strategy to probable induction; but where, decision, strategy, betting become the guiding criteria for every clinical choice in uncertainty conditions. In both the situations the idea of chance intervenes: in the first one as an expression of the natural immanent variability, of the fortuitousness of the intervening processes, of the intrinsic vagueness of phenomena; in the second one as expression of insufficient individual knowledge, an information fault, not a fault of phenomenal determinism. However vaguely, the Carnap's distinction between probability-1, as degree of belief and probability-2, as empirical frequency, may offer on this matter a suitable logical scheme.

13. Lately we find a revival of interest for the problem rationalizing the diagnostic process on the stimulus of an ever more refined, automatic handling of information. Automation becomes indispensable for the translation of research data into practical data. A computer is able to consider every single hypothesis within the known range of alternatives, it may update itself about statistical frequencies which are not invariant from population to population, and it is capable of perfect syllogistic inferences without incurring in formal errors or logical fallacies. It is greater than man in the rigorous respect of grammar, in following axiomatic rules: if the premise is true, the following conclusion worked out by the machine is also true.

Such is not the case with inductions, for which it is still the natural intelligence, informed by the evolution process, that more often hits the target. Inductive inferences are not "relations of ideas", to put it as Hume did: they are reasonings which venture into what the skeptical logician called "matters of fact". Without our distractions, our prejudices, our vices, the computer can diligently work out a logical syntax and, within the terms of such a syntax, it can deal with the semantic contents inlaid in a store of information, so as not to leave out any element, any knowledge of the clinical praxis. Hence, it can draw, in the defined sense, rapid and coherent diagnostic information. But – question returns – in which logic? Is the logic of the highest probability that is the prevailing of "statistical normality"?

Some people believe it is more in conformity with the clinical praxis – which is not very inclined towards the express use of probability, moreover if restrained within the Bayesian scheme – a "cognitive" logical process based on classifying canons reduced to the choices of "expert systems", within programs oriented towards the reproduction of the rules, usually followed by the clinician with a lot of experience, so as to codify in systems of decisional chains the models for the simulation of mental processes. The same logical links between events can find timely helping of natural intelligence in a system of "artificial intelligence". More than regarding the possibility of simulating messages between neurons, when the chemical and electrical fundamentals of the most elementary cerebral synapses are not yet completely defined, the urgent question lies with the meaning of intervention of the "fortuitous" element in cognitive processes, natural and unnatural ones. But, first of all one must identify the structural sub-strata of behavior and reasoning, specify the neuronal circuits programmed in the genetic code: from the transmission of the nervous impulse (electro-physiologic expression of an adaptable evolution) to the working out of information in the upper centers of the brain. Moreover, from experience to experience, from neurogenesis to neurogenesis, the meaning of necessary and fortuitous, of certain and possible, cannot but accept the widest differences of evolution.

The ideological *querelle* of Claude Bernard appears to be revived in the conflict between the diagnosis based upon probabilities and the diagnosis based on "expert strategies". It is an antithesis whose logical-statistical content should be considered again. If "statistical" diagnosis utilizes collective properties which are put into categories and translated into probabilities, with all the limitations and the risks involved in the adaptation to particular situations, the "expert" diagnosis, as well, does not seem to guarantee, by itself the relevance of every result. Anyway, the basic problem remains and it is shared by any computerized process. The problem is the "datum", with its changeable and eluding empirical contingency. This contingency, if it can uproot the chance of a general use of the Bayesian criterion, equally affects the results of any other logical inferential process.

14. There is a lot to learn by observing how today, in statistics, the classic problems of knowledge are being proposed again, not always with the required awareness, and, even the question reappears whether statistics should be founded more on rational or empirical bases – such was a report, from the *Journal of the Royal Statistical Society*, to the *Statistician*, and precisely regarding the clinical diagnosis. After years and years of frenetic technicalism, methodology is hence returning to questioning on fundamental matters: even in the fascinating comparison between different strategic principles. Amongst these: the Bayesian theory, the mathematical theory of evidence, the theory of the certainty factor, are all attempts to emulate the cognitive process of thinking, to make a logic out of common sense, to formalize inferential acts of which not even the expert appears to be rationally aware. Upon the conjecture of this one, the "expert systems" are in fact modeled, by now, almost as a fashion, much preferred due to their apparent simplicity and their satisfying univocal results. Perhaps, even for their avoiding "probabilized" spaces. Quantities which usually intervene in the scaling of alternative are simple numbers, *ad hoc* scores. By means of these quantities the deductive and categorical systems have recently moved from a qualitative to a quantitative stage, always within a deterministic or causal context. On the other hand, it is not impossible to scale at different stages the "decisional trees" by means of truly real probabilities; it is not impossible to translate those diagrams into Bayesian inferences, provided that in the space of alternatives the hypotheses are made as exclusive.

However, for these systems it is preferable to avoid the probabilistic language and to lay down specific programs for different pathologies, structured in causal hierarchical networks and then move from sets to sub-sets solving any decisional knot in relation to specific evidence. These are procedures aiming at the simulation or the diagnostician's process of elimination by successive classifying alternatives, sequential modular schemes in which each segment of knowledge carries out the logical canon "if, then", so as to find again a kind of deductive necessity. In spite of their misleading simplicity, these are complex structures which require automatic systems capable of governing a combinatorial redundancy of alternatives.

In contrast with such imposing deterministic and prescriptive plans, supported by systems of artificial intelligence, the more agile indeterministic models are set, models coherent with the immanent statistical variability and orientated towards the formulation of probabilistic forecasts. These forecasts, however orientated towards the highest probability, do not exclude, because non-univocal, any atypical event, do not hide the least frequent clinical case. It is true that the Bayesian scheme postulates a closed space of alternatives, but it only adduces probability. This is not so for the deterministic decisional tree. In its causal proceeding, it sequentially neglects bundles of clinical hypotheses that are not immediately converted back to the codified information, on which the system allocates those conventional quantities in which it appears to show the difficult relationship with the uncertainty of a logic aiming at forecasting certainty, at inducing by deducing. Similarly, the mathematical theory of evidence, that is very much a fashion nowadays, attributes real numbers, from 0 to 1, to hypotheses (numbers that may perhaps be understood as degrees of confidence in the Carnap sense, although not forced to satisfy Kolmogorov's axioms).

15. In expressing the difficulty of abstractly formulating normative methods, of drawing positive inferences from empirical contexts, these theories have a tendency to escape (*à la* Claude Bernard) statistical probability as intellectual instrument, as line of thinking. Yet, one needs to be educated about probability in relation to knowledge and action, when knowledge and action avail themselves of inductive inferences, particularly in the medical field. Diagnosis is probability, prognosis is probability, and research draws probability distributions in terms of statistical induction: at time as probable knowledge, at time as convenient bet. It is logic of uncertainty, in which the categorical statement is not any more totalitarian as the major premise of the Aristotelian syllogism; and the hypothetical statement is no longer univocal and causal. It does not translate itself; it cannot translate itself, into the syntactic distinction between fallacies and correct logical figures. These are the canons of a strictly statistical vision, of knowledge deep down into natural variability, of a philosophy in which the ancient duality of deduction and induction even appears to melt-away.

More than a pragmatic difference, the contrast between statistical-probabilistic diagnosis and the diagnosis based on cognitive schemes of expert systems reflects a conceptual alternative. The diagnosis which uses statistical probabilities atones for the objective non-transferability to the single situation of information which is, and remains, statistical and probabilistic within a reference class. The "expert" diagnosis aims to the identification of individual causal correspondences, but it faces the non-repeatable aspect of events, the inherent variability. Each clinical case becomes a *unicum*, an incomparable singularity, a class by itself. The more each scheme of "causes" and "effects" is made in details, the more it tends to vanish. To identify cause-phenomena and effect-phenomena means always to transcend singularities: the categorical repetition of the world events is a classifying abstraction, linguistic artifact, semantic allegory. However, pathology and clinical medicine have codified (in the medical and surgical field) realities and criteria, have identified typologies. Are they conventional categories? Surely they are. They are the categories which shore up scientific knowledge, as well as common knowledge: and they are statistical categories.

To move from those categories to single situations it is neither always easy, nor always possible to convert it to automatic procedures. If the mechanical simulation of the reading of an instrument result, of laboratory data appears as straightforward, it does not look equally simple to reproduce the diagnostic induction. The mental process that, in the first century A.D., was defined by Cornelius Celsus as the art of conjecture, and two centuries B.C., by Carneades of Cyrene was assimilated to the judicial inquiry and taken back to his skeptical probabilism, to the logic of assigning "degrees of persuasiveness" to the πιυανότης of Plato and Aristotle, with the meaning that Cicero will give to *probabilitas*. This is, neither in view of rejecting any attempt towards the rational objectification of clinical reasoning, not in order to find an excuse for the common-place that considers diagnosis purely a practical art (mantic art, divining art) but rather in view of calling for the reconsideration of a thoughtful pondering *in re probabili* of data and methods.

It is not an easy thing to say where these procedures may lead, procedures grown from the bonding of informatics and nevertheless tending to become autonomous doctrines by means of pragmatic canons of thinking which are based on the always fascinating games of combinatorial mathematics. It is not an easy thing to foresee the destiny of those stimulating intelligent, computerized structures which are known as "connection machines". But do allow a humble "abacus man" full of curiosity and doubt to wait for an imminent reo turning to some fundamental logical questions, in the search for profound meanings. The recent sudden interest in the problems of foundations, along with a growing feeling of impatience with so much of exasperating and dulling specialization, appears to bring some omen.

16. Whenever certain pragmatisms, whatever the logical criterion, repeat prescriptive rules, without any spark of critical spirit, they end up in puerile normative schemes: in the naive "Bayesianism", as well as in the naive "cognitivism". This is the risk involved in the mechanical handling of any technique, even the most tested one: this is true within the context of research as well as in that of strategy. Has not Penrose written that a statistical analysis on multiple births, carried out without critical intelligence, would inevitably lead to conclude that human twins genetically inherit even clothes? This is an irrational application of the covariance criterion that dates back to Francis Galton, the inventor of the method. Even he, the founder of the biometric movement, met with a serious fallacy regarding heredity in jumping from syntax to semantics, from methodology to phenomenology, from statistical covariance to biological determinants. But, if Galton, the

biometrician, was suggesting not to leave out *a priori* any correlation, Huxley, the naturalist, was spreading an amusing story aimed to warn whoever tried to reduce the world representation to far too easy causal connections. It went, more or less, like this: There is a correlation in the English counties between the amount of milk produced and the number of spinsters. It is known that cows produce a lot of milk if they can feed themselves on sainfoin and clover, plants that are pollinated by hornets, of whose honey and larvae country mice are very glutton. These mice are stalked by cats and these are notoriously taken care of by spinsters.

Obviously a computerized system can handle better than the human mind those particular deductive procedures related to the so called "mathematical induction".[8] A community of *robots* could perfectly solve the famous logical problem of the forty unfaithful wives (to my knowledge the most amusing example of mathematical induction); it would not be so easily done by a community of human beings, especially of suspicious husbands. However, the logical-deductive capacities that would be of such a great help to these, are instead – such an irony – totally useless for a community of robots, because, like angels, they are sexless. Perhaps, if it is still allowed to carry on jokingly, in some of the robotic exaltation of our times one can see the aspiration of a restless humanity looking for alternative links for evolution, since the famous geneticist J.B.S. Haldane has noticed that man cannot hope to evolve himself into angel, because he does not have sufficient genetic variance for the moral perfection and for the wings.

Likewise, the *robot* does not seem to be superior to man in relation to empirical induction when dealing with probability, playing with chance. It has now become indispensable to use pragmatic canons of thinking when deciding within closed systems, codified in all the rules of the game, all the possible strategic alternatives. Not so when the changing reality imposes a choice facing a novelty that eludes any programmed combinatorial virtuality. For instance, how could possibly an "intelligent" computer solve the painful doubt of Louis Pasteur, forced to take a quick decision about using for the first time his experimental vaccine against rabbis on a child who had been bitten by a dog? How would it have handled the alternative between uncertainty on the chance of infection and uncertainty on the results of vaccination?

Without the filter of a culture that is critical knowledge', creative spirit, daring imagination, no automatism can still replace the human mind, especially when dealing with chance. We all smile at the little story about the intelligent computer that, questioned in view of a travel by plane about the possibility of death by an implanted bomb, faultlessly answers: "The probability that a terrorist has hidden a bomb in the baggage compartment of that flight is 10^{-2} (one hundredth)". Then, after a quick and correct calculation, it adds with axiomatic rigour: "but there is a much smaller probability, exactly equal to 10^{-4} (one tenth of thousandth), that two

[8] Deductive rigor and the possibility of a quick synthesis of knowledge heritage are the positive advantages making automatic and informatics systems, by now indispensable in the medical research and practice, instruments such that it would be irrational to neglect completely. It would be irrational and perhaps even immoral, and for so many reasons.

terrorists, one unaware of the other, have installed each one bomb". Finally, efficient and zealous it adds: "hence, I advise you to put a bomb in your luggage so as to make very small the probability that there is another one".

Obviously only a little story, an amusing *boutade*, but I would not like to think that an "intelligent" system, involved in the calculation of probability for a specific illness in an individual, could accurately calculate a value p, realizing afterwards that the empirical association frequency of such an illness with a specific contagious form is one tenth of thousandth of p, and would hence advise, with its unarresting logic, the automatic therapeutic system to infect the unfortunate individual, so as to reduce sensibly the probability of the other pathological entity. But, at least for now, this is only a joke.

Department of Statistics ITALO SCARDOVI
University of Bologna

SUMMARY

Statistics in the biomedical research and clinical strategy

This paper presents some reflections on the biomedical and clinical research, from their origin to the present. A certain mathematic and statistic rationality may help to give a logical addition to the biomedical research and to change clinical strategy into a coherent decision making process.