STATISTICA, anno LXXV, n. 2, 2015

THE INDUCTIVE APPLICATIONS OF PROBABILITY CALCULUS

Corrado Gini (1964)

Conference held at the Istituto Centrale di Statistica during the ceremony in honor of his eightieth birthday and published in "Rivista di Politica Economica", LIV, series III, VII, July 1964.

On revient toujours aux premières amours. Being appropriate at my age, I must admit that my first love was for the probability theory.

The title of this speech is almost identical to that of my first paper *Contributi alle applicazioni statistiche del calcolo delle probabilit*, published on "Giornale degli economisti", in December 1907. That was not a superficial *flirt*.

The theses presented in it were developed and applied in the volume on *Il sesso dal punto di vista statistico*[1], which at that time was in print, as it was in the following year. I supported those theses at the Congress of Italian Philosophical Society, held in Parma in September of that very year, successfully working alongside with famous experts of philosophy and physics, who honored the Society and the Congress, under the authoritative chairmanship of Federico Enriques[2].

Recently, while looking through the archives for a collection of my publications on the foundations of statistics, which I will call philosophical, I happened to find a thick manuscript in which the logical and psychological theory of probability was dealt with thoroughly, with all the publications, from the first classic ones to the last contributors of that time, being considered.

Enriques, who had the patience to read it, then asked me to write a summary of the main conclusions in the article *Che cos'è la probabilità?*[3] for the "Rivista di scienza" *(Scientia)*, of which he was co-editor.

On reading the paper now, which was then put aside as diverging from the statistical orientation of my studies, I had the satisfaction, almost sixty years later, that my thinking has remained substantially unchanged. Hence, that paper will be added to the publication in which the benevolent colleagues wish to remember this unfortunately venerable birthday.

---

[1] C, Gini, *Il sesso dal punto di vista statistico*, Palermo, Sandron, 1908.

[2] C. Gni: *Il Congresso di Parma*, Rivista filosofica, fasc. 4', vol. 10, 1907; *Intorno al concetto e alla misura delle probabilità*, Bollettino della Società filosofica italiana, anno 3, n. 3-4, 1907; *Sul concetto di probabilità*, in "2 Congresso della Società filosofica italiana" (Biblioteca di filosofia e di pedagogia), Modena, Formiggini, 1908.

[3] C. Gini, *Che cos'è la probabilità?*, Rivista di scienza (Scientia), vol. 3, n. 6, 1908.

I must say, without showing off, that my love did not remain then misunderstood.

In spite of my young age and even younger look, having mostly lived in the country, alternating my visits to the University of Bologna with sports, so unaware of the academic surroundings as to ignore the names of my university professors I became a pupil of a teacher who was a model for his scientific consciousness and honesty, but was not liked by colleagues because of his pronounced political ideas and his harsh character. Without any extra academic support, I was accepted as an unexpected and unwanted intruder by a very compact group of very determined, very influential and protected competitors, among whom some like Bresciani Turroni, Mortara and Beneduce, of an extremely high scientific level. If in spite of all this I manage to find my path without great difficulties and, as an ephemeral goal, to became for some months the youngest professor of Italian universities, was certainly mainly due to the impression aroused from the reading of my publications on probabilities and their applications to the sex at birth and perhaps, even more, from the debate held at the Parma Congress, among the experts of exact sciences who, by their influential judgement, influenced the judges of competitions in Statistics who, although often very competent in other branches of the discipline, were completely lacking in probability theory.

In the above mentioned article I stressed, and I believe I was the first to do so, the fundamental difference existing between the deductive and the inductive applications of probability calculus.

For instance, the deductive applications are those which have probably been made for tens of thousands of years to forecast the results of gambling games. It is in fact certain that, since such games started to be played (that is since 40,000 years ago as is believed from the astragals, that had the role of our dice, brought to light in archaeological findings) some evaluations must have been made regarding the probabilities of the various alternative results.

On evaluating these probabilities of elementary events on the basis of the game structure by probability calculus it becomes possible to forecast those of their various combinations, and this is due to the hypotheses, which are at the basis of probability calculus, appear to be the same as those which governed the construction of the instruments of the game. Probably, gambling games and probability calculus were born together and certainly they developed together long before scientists took an interest in them, as far as we know.

We know that in ancient times and in the Middle Ages people played wildly, so as to provoke the intervention of the clerical and lay authorities for its control. Instead, the first records that can be traced regarding scientifically correct applications of the probability calculus to gambling games, date back to just after 1000 AD and were to be found in Arabic writers who, as it is known, form the bridge between the science of ancient classical times and that of modern times. In fact one can remember al-GazaIi's solution given for some problems which were dealt with in a greatly developed way from the end of 400 AD to mid 500 AD, by a group of eminent Italian mathematicians as Luca Pacioli, Tartaglia, Cardano, Peverone, and later Galileo[4].

Only after some time, such problems were reconsidered and delved into more deeply in France by Fermat and Pascal, to whom, by evident mistake, priority was attributed for a long time, and following in their track they were organically developed by Huygens. It is Cardano's merit, he himself a keen gambler, to have moved much earlier from the practice of games to scientific considerations of a general kind about probability calculus, thus making him the first scholar to deal with such a subject. However, his unrevised and disordered notes, sometimes of difficult interpretation and published more than a century later, when probably the clear considerations by Galileo were already know and latter between Pascal and Fermat had been published and also Huygens' booklet had come out, these of course did not at all influence the development of the theory[5].

Instead in this area, Jacob Bernoulli's work caused a revolution. By starting from Huygens' booklet, that he reproduces, discusses and completes, he extends the applications of probability calculus from gambling games to moral, judicial and economic phenomena, by coordinating the one to the other and arranging them in a publication of fundamental value. Jacob Bernoulli defines the probability concept, by wisely differentiating it from its *a priori* and *a posteriori* measurements: the first on the basis of the structure of phenomenon and the second on the basis of its frequency in a large number of observations. He foresees the difficulties of the two paths and he particularly points out the increasing approximation of the second with the increase in the number of observations and, aiming to measure it, he formulates his famous theorem which, although not really answering the proposed problem, did anyway make the decisive start towards the solution, that a century later would be given by Laplace[6].

Ars Conjectandi, a posthumous publication, constitutes an original and complete work of the applications of judicial, social and economic phenomena (except for the examples that were not developed due to Bernoulli's death), of probability calculus, understood as a tool of investigation suitable for all the phenomena of nature and society. So, Bernoulli is rightly considered to be the founder of the probability calculus. What a benefit would be gained from its reading by the modem statisticians who rarely expand their knowledge to the statistical theory dating before the World War?

But one could not expect Bernoulli to do everything, and it is natural that he, as all innovators do, would be likely to exaggerate the impact of his innovations.

---

[4] C. Gini, *Corso di statistica* (eds: S. Gatti and C. Benedetti), 1954-55, Roma, Veschi; *Origenes y prospectivas de la estadistica*, Supplement to n. 31 of the Boletín de estadistica, Madrid, 1946; *The first steps of statistics*, Educational research forum. Proceedings, New York, Endicott, 1947.

[5] C. Gini, *Gerolamo Cardano e i fondamenti del calcolo delle probabilità*, Metron, vol. 19, n. 1-2, 1958.

[6] C. Gini, *Gedanken zum Theorem von Bernoulli*, Revue Suisse d'économie politique et de statistique, n.5, 1946; *Abhandlungen zur matematischen Statistik*, Berna, 1946, (Italian ed: Rileggendo Bernoulli, Metron, vol. 15, n. 1-2-3-4, 1949; *Concept et mesure de la probabilité*, "Dialectica", vol. 3, n. 1-2.

The weak point of his construction does not lie only in having presented his theorem as the instrument by which to move from the frequency to the probability of events, while it just represents the way to move down from probability to frequency. It is evident that he did not realise the obvious difficulties of extending to the judicial, moral and economic phenomena, the procedures of probability calculus tested by several thousand years experience of gambling. If he had brilliantly overcome the formal difficulties of mathematical calculations, he did not exactly value the substantial difficulties arising from their different nature. This is clearly shown by the correspondence that he had with Leibniz[7].

The correspondence lasted two years and was totally inconclusive. At that time on the continent, Leibniz was considered to be a scientific deity, whom one did not approach however much invited to, without adulation. Stimulated by him to a mathematical approach to various gambles, Bernoulli answered by informing him that for many years he had been pondering on probability calculus and confidently explaining the object of his new theorem. *"Nescio, vir amplissime"*: great man I do not know, he concluded, if you will think that in these speculations of mine there is something valuable.

With consideration and politeness, yet maintaining a certain distance, Leibniz replied by making his reservations with skeptical circumspection on the possibility of empirically obtaining a perfect probability evaluation on judicial and political matters.

A bit upset, Bernoulli replied by stressing that he had given the demonstration of his assumption. Leibniz emphasized his objections, to which Bernoulli did not reply, although he did stress again his point of view. In the meantime he met his death, exacerbated probably by a misunderstanding with the most famous scientist of that time, besides his brother John's ingratitude.

The misunderstanding between the two was total and the motive results from a phrase in the first answer given by Leibniz: "The evaluation of probabilities is very useful, but, in judicial and political subjects, an accurate listing of all the circumstances is more important than the refinement of calculations".

Bernoulli, if he cannot be justified, may be excused; as a good mathematician, he attached more importance to the technical side of his theorem than to the difficulties of his applications to the judicial, moral and political field, to which he intended to extend it. It is probable that, if he had had the time to complete this part of his work, as he had planned, he would have been able to tackle these difficulties and to refine and condition his thinking. Leibniz would be excused if, knowing only the formulation of which he was informed by the author, he did not adequately appreciate the technique of the elaboration and, as a philosopher as well as a mathematician, he was concerned about its applicability.

Such a misunderstanding, apart from making the last days of Jacob Bernoulli very sad ones, had deleterious effects on the development of the probability calculus.

In the hands of the mathematicians, the method, refined and well arranged by Jacob Bernoulli, have given rise to an uncontrollable enthusiasm and led to

[7] C. Gini, *Gedanken zum Theorem von Bernoulli*, op. cit.

indiscriminate applications in all areas of knowledge, sometime with the ingenuity that often accompanies enthusiasm.

Without the probability calculus - Bernoulli had written - *"nec sapientia philosophi nec historici exactitudo, nec medici dexteritas aut politici prudentia consistere queat"*, and Condorcet hoped that, thanks to his work, *"notre raison cesserait d'etre l'esclave de nos impressions"*, an obviously very dangerous freedom because, having dismissed the impressions, what remains is to resort to fantasy.

From the observation errors in physical sciences to the oscillations in the sex ratio at birth and death, from the distribution of the shots on a target to the ratios of illegitimate child-births to the total number of marriages and those of born, from the mechanic theory of heat to the distribution of the various forms of suicide, from the kinetic theory of gases to the percentage of sterile seeds and to the measurements of the purity of seeds, from the frequency of certain aberrant characters in plant sprouts to the verdicts by judges and to the resolutions taken on a majority vote by other corporate bodies, everything was submitted to probability calculus in order to obtain the laws and forecast.

By pretending to deduce from the probability of elementary phenomena the description and the forecast of all natural and social phenomena, however complex they might be, one had the illusion of having found in the probability calculus the lever that might lift the world of knowledge. But the supporting theory was often missing, which should have been common sense.

"La théorie de la probabilité – said Laplace – *n'est au fond que le bon sens réduit au calcul: elle fait apprécier avec exactitude ce que les esprits justes sentent par une sorte d'instinct, sans qu'ils puissent souvent s'en rendre compte"*. But, if he had been still alive then, Giusti would have appropriately said "Common sense, that once was the schoolmaster, is now dead in many schools, and science, his daughter, killed it so as to see what it was made of".

In effect, several of the conclusions reached in the applications of probability calculus were well founded, because the structure and the development of the considered events did, more or less approximately, respond to the hypotheses implicit in the probability calculus; but many others resulted in perplexity and diffidence.

Above all, the conclusions, diverging from author to author, but all not very convincing, which were reached in the applications of the calculus to the deliberations of judging were perplexing to open-minded researchers (Laplace's *les esprits justes*).

Nor, on the other hand, were the refinements by the probabilists appreciated in *alto loco*. It is said that Napoleon told Laplace, removing him from the Finance Ministery he had assigned to him, "you deal with infinitesimal".

In the end the educated public cried shame to the "scandal of geometricians", as the mathematicians were then called, and we can say that the thesis that it caused a halt in the vigorous development achieved by the calculus of probability appears to be true.

It is said that Einstein, faced with the modern theory of the indeterminist physicists, exclaimed that he could not believe that God played dice with His creation. It seems that none of the famous minds, who cultivated the probability calculus in its socalled golden period, realized the unlikeness that nature (as at that time God was still confined to the attic) controls his structure and its evolution through the schemes of combinatorial calculus.

The professional tediousness of two inspectors of the French public administration, one in education and the other in finance, personal friends and ideally like twins, was necessary in order to go against the trend. But after all, would it be out of place, would Cournot and Bienaymé have asked themselves to find out up to which point the deductions of the probability calculus conform to the facts? (if the question was not explicitly formulated, it implicitly results from the type of their publications). But it does not appear that their reservations found a response, because of the overwhelming authority of Laplace's school.

However, while Laplace's school dominated and in a certain sense oppressed the development of thought regarding the calculus of probabilities in his country, it stimulated the initiatives of foreign countries, originating in what can be considered as the second revolution in the area of the probability calculus.

Laplace was still alive when from Belgium a nervous investigator [Adolphe Quetelet] already a mature one in terms of age and study, and with a particularly unusual curriculum, came to Paris. He, who won prizes ever since he was in secondary school for his ability to draw, a subject that he later taught, got his degree by writing a thesis on geometry that earned the admiration of specialists; later on besides drawing and geometry, he taught probability calculus, differential and integral calculus, astronomy and physics, as well as the history of science. On this subject he published a treatise and he followed it constantly as a member, and soon as secretary, of the Science and Literature Academy of Brussels. But he taught these various subjects only in order to achieve economic independence, because his dream was to became famous as a poet and an artist. He wrote poetry, had been in a painter's studio, played the flute, produced plays and would have successfully pursued these artistic and literary inclinations if Gamier, who had came from France to Brussels to take the chair of mathematics and astronomy, had not positively channeled his exuberant activity towards these subjects. Involved with the foundation and organization, and later the management of the astronomic observatory of Brussels, proposed by him, he, in his duty to fulfil the project, undertook surveys of physics and meteorology and, subject to the meteorological phenomena, of botany, wrote the first essays on meteorologic and biometric statistics.

Sent to Paris in order to learn about astronomic instruments, without neglecting his mission, he met Laplace, Fourier, Poisson, Lacroix and stirred his inclinations for the probability calculus, this together with those for statistical surveys. Certainly, if there was a person capable of developing the probability calculus from the theoretical schemes of combinatorial calculus and adapting it to the reality of natural and social phenomena, this was surely Adolphe Quetelet. To him goes the credit not only for having been the founder of the modern orientation of statistics besides that of having organized the census of his country's population and having promoted and organized international statistical meetings, but also one that is of direct interest to us, of providing the first example of inductive applications of the probability calculus, which may be considered to be a second revolution in this field[8].

Thus, how they do basically differ from the traditional deductive applications? For instance, in the case of heights, the old style probabilists, once he has known the standard deviation, would have deducted, on the basis of the relations that probability calculus teaches, among the various deviations, what is the probability of obtaining a deviation equal to a fraction of such a standard deviation or a multiple of it, hence defining a distribution that was anyway hypothetical, as based on the assumption that the character would follow the curve of random errors.

Instead, Quetelet not only studied the distribution of heights, but also of a series of meteorological phenomena and mainly anthropological ones, verifying that they actually followed the curve of accidental errors with a good approximation. So he could work on data observed according to the norms which had been indicated by the probability calculus for this curve, and not just this, but he also could deduct information regarding the structure of the characters, from the overlapping of the two distributions[9].

For such phenomena, as for the measurements of physical quantities affected by accidental errors, he concluded that there must be a constant common cause disturbed in its performances in the single cases by a complexity of unpredictable, mutual independent factors.

In particular, regarding anthropologic characteristics the mean would have expressed the effect of the hereditary factors common to all individuals, while the individual variations from the mean would represent the effect of environmental factors, differing from individual to individual.

It is remarkable that, while Quetelet realized the need for inductive research for measurable quantities or, extensive ones as they are usually called, such as height, temperature and so on, he continued to carry out deductive applications of the probability calculus for intensive or numerable quantities, such as male frequency in child births, that he assumed, without verifying it, followed the Bernoullian scheme.

Thanks to Dormoy, a French actuary, and Lexis, an eminent German statistician and economist, the inductive applications were also introduced in this field some decades later[10].

In particular, Lexis thoroughly studied the problem, leading to that theory which was called the dispersion theory.

He found that the sex ratios in human births show, from one time interval to another, or from a territorial district to another, differences similar to accidental

---

[8] It is curious how Quetelets international reputation greatly surpassed the fame he had at home, this perhaps due to disliking aroused by the indiscreet curiosity subjecting to measurement all the dimension of human body in a period when, except in the bedroom, the anatomy from the neck to the ankles was not allowed into high society.

[9] C. Gini, La distribution des caractères anthropométrique, 1962; Biométrie-Praximétrie, n. 1, Sur la théorie de la dispersion et sur la vérification des schémas théoriques, Metron, vol. 14, n. 1, 1940.

[10] C.Gini, *Sur la théorie de la dispersion*, op. cit.

ones found in the ratios of the colored balls extracted from an urn and put back again after each extraction. He concluded that the probability of a male birth was constant in time and space and it did completely depend on the woman who has produced ovules of the two sexes at a constant or accidentally variable ratio.

Both the conclusions of Quetelet and Lexis were wrong, but this should not obscure their merits. In regard to everything the beginning is the most important and hence an extremely difficult thing, but to add to it is easy, as was well said by Aristotle in a sentence that Tartaglia took as his motto. It is a sentence that should be kept in mind by many of the modern statisticians who believe themselves to be superior to their predecessors for adding something to the method introduced by them. Like flies which, while resting on an eagle's head claim to be flying higher than the eagle.

Quetelet was wrong in assuming that the characteristics of a population, which is distributed according to the curve of random errors, have hereditary causes common to all and only differ among themselves due to environmental conditions. This was experimentally proved to be correct only for pure breeds, that is for self - fertilizing monoecious animals or for parthenogenetic individuals, while for dioecious or monoecious species with crossed fertilization, the individuals differ also as regards hereditary characteristics as well. Nor, did a more accurate investigation show that the curves of anthropomorphic phenomena perfectly reproduce the curve of random errors or normal curve, because in reality they diverge from this due to a greater frequency of the values in the middle and the extremities, and a smaller one of the intermediate values, this originating a more or less hypernormal curve depending on the various characteristics. For the eye's refraction, hypernormality is very high, for which a plausible explanation is given considering that the curve is the result of three curves which may also all be normal: a central one corresponding to the most frequent type, emmetropic, and two lateral ones corresponding to myopic and hypermetrope types.

Similarly, more probed statistical research showed that the sex composition of the offspring certainly depends largely on chance, but not on chance alone, because the various married couples have their own tendency to give birth mainly to the males or females, a tendency which, on the other hand, seems to vary during the course of the generations, while the progress of biological sciences put out of discussion that, as a contrast to Lexis' conclusion, sex is not predetermined in the woman's ovule, but is determined at the moment of fertilization by the male spermatozoon presenting two types: one gymnogynous and other androgynous.

In this case, as well, statistical analyses carried out by more sensitive procedures have also shown how in reality the dispersion of the sex ratio in time does not exactly correspond to that foreseen by the probability calculus, because a certain solidarity exists among the sex ratios of births of successive intervals.

All this demonstrates that the inductive applications of probability calculus are not at all simple, as might appear at first, and it suggests a close investigation of the presupposed conditions and the precautions which need to be taken. It will be much more useful because the inductive applications of probability calculus are only a particular case of those models which were so fashionable in the last few years in the fields of economic and social sciences, but which unfortunately are often applied without complying with conditions and precautions that would lead to reliable results[11].

Regarding this, it is worth recalling the aims of the inductive approximation of probability calculus:

a) The first aim is a descriptive one. They allow us to decide whether or not the considered phenomena conform to the probability calculus, and hence if by this they may be represented with greater or smaller applications and, if they may be corrected by introducing suitable coefficients into the parameters determined by the probability calculus, so as to take into account the divergences. What we want is the simplest possible scheme, true to the distribution of the considered phenomena, that may be taken as the law of its distribution.

b) The second aim is a heuristic one. From the comparison of the observed results with those forecasted by probability calculus, one may realise whether the distributions of the observed phenomena can be explained by the hypotheses which are the bases of probability calculus, or by other ones which may be substituted for these, or if other systematic factors of divergence, which might be determined by further research, exist.

The following considerations may be useful for reaching such aims:

1. when drawing a probabilistic scheme, it is advisable to take into account all the known circumstances so as to obtain the maximum conformity of the theoretical data with the observed data. Not only does this correspond to the descriptive aim, but it makes it easy to reach the heuristic aim, because the slighter the differences between theory and observations, the easier it is to understand their causes;

2. at the same time it is advisable to propose the simplest possible scheme because science has the aim of leading to simplified representations of reality. Evidently, the accuracy of description and the simplicity of the formulae expressing it may be antithetical: depending on the context, it will be convenient to prefer one of the other requirement. I am more inclined to give preference to the second one. Complicated schemes should be avoided, not only because they do not respond to the thinking economy that science intends to fulfill, but also because they make it more difficult to reach the heuristic aim. Hence, simple formulas are to be preferred to the complicated ones involving many parameters. One might say – being attached to faithfulness should not lead to polygamy.
   On the other hand, it often occurs that several schemes may equally be proposed, different ones but equivalent from the point of view of simplicity. The choice among them will be appropriately made on the basis of their accuracy, as well as on the basis of their usefulness from other viewpoints, which may, however, vary from one research to another;

[11] C.Gini,*Intorno all'uso dei modelli nelle scienze e in particolare nella scienza economica*, Rivista di Politica economica, 3 serie, fasc. 1 1953.

3. a third requirement it that of verifying the conclusions so as to make sure that they do not have a value bounded to the observed cases in the research being considered, but they have a more general validity, both in the same research and in other studies.

For this requirement, as well as for above mentioned requirements, it must be noted that the conditions are very different in the social and natural sciences. In particular, in the area of physics that deals with the most minuscule particles, the scientist works on an almost virgin ground and it is hence easy for him to take into account all the known circumstances, while this generally constitutes great difficulties in the social sciences. On the other hand, the physical sciences have the facilities of a large net of laboratories where the proposed schemes may be promptly verified, which instead is not possible for the social sciences, not only due to the lack of a similar network, but also due to a greater complexity of the phenomena studied. These two circumstances should be considered by the experts of social sciences, who believe that they make progress by imitating the experts of physical sciences in the model building. In reality, their imitation is very poor, either because they generally do not take into account all the known and often very complicated circumstances, or because once having proposed a scheme and obtained certain conclusions, then they do not bother to verify them. Any scientific branch must make use of methods suitable to it. This is an essential condition for success.

For this purpose it is necessary to state the importance of verifying the adopted probabilistic scheme[12].

As the verification results positive, this allows us to state that the scheme supplies an explanation of the phenomena observed, which is sufficient but not necessary.

The mistake made by Quetelet, like that made by Lexis, was not having realized it.

It is true that if the manifestations of a phenomenon are governed by a constant cause for all the observations carried out, but these are disturbed by a series of different causes intervening in a reciprocally independent way from group to group or individual to individual, the distribution of the phenomena takes the shape of the curve of accidental errors, but the opposite may not be true at all.

A distribution similar to that of random errors may also occur under other hypotheses. In reference to anthropometric characters, it was seen that hereditary factors, which according to Quetelet should have been constant for all individuals of the species, instead differ according to subspecies, races, families, chromosomic combinations occurring during fertilization, probably following a curve that, if not exactly like that of random errors, is of a similar type. The influence of hereditary factors, added to that of environmental factors, determines the observed distribution that conforms approximately to the curve of random errors, however differing from it because, as we said, it is more or less hypernormal.

Similarly, it is true that, if the probability of an event remains constant in time and space, and the occurence of the event in one case does not influence its

[12] C.Gini, *Sur la théorie de la dispersion*, op. cit.

occurrence in the successive cases, the dispersions of the frequency of the events is normal in the various intervals of time and in the various territories, but the opposite is not true.

In the majority of cases, the inductive applications by Quetelet and Lexis responded with sufficient approximation to the descriptive aim, but they totally failed their heuristic aim.

For the inductive application to respond to their heuristic aim, it is necessary not only that the correspondence between the observed distribution and the expected one on the basis of the probability model occur, but also that the single hypotheses assumed by the scheme are proved by the facts. For the distribution of heights and other anthropometric characters, the hypothesis of a constant cause represented by hereditary factors did not correspond to the facts. For the variations of the sex ratios at births or deaths in time, the hypothesis that probability remained constant in the single cases, did not correspond to the facts, nor did that the successive cases were totally independent from one another.

However, for social phenomena, it is not at all easy to carry out the verification of all the hypotheses implied by the scheme. This explain why for only a few phenomena the inductive applications of probability calculus have been used from a heuristic viewpoint. As far as I know, so far, a useful application has only taken place for the sex ratios at birth, but it is certain that as our knowledge progresses and statistical methods are refined, this orientation will be able to make notable contributions to science.

However, for social phenomena, it is not at all easy to carry out the verification of all the hypotheses implied by the scheme. This explain why for only a few phenomena the inductive applications of probability calculus have been used from a heuristic viewpoint. As far as I know, so far, a useful application has only taken place for the sex ratios at birth, but it is certain that as our knowledge progresses and statistical methods are refined, this orientation will be able to make notable contributions to science.

I believe it is worth recalling the results that have been progressively collected regarding the sex ratio in human births.

The idea that the weak correspondence of the actual dispersion with the theoretical dispersion would show that the hypotheses implicit in the calculation of the theoretical dispersion respond to reality, led Lexis to admit, as we have mentioned, that a probability of a male birth was constant for all the women, or, at most, it varied accidentally from one to another, and that the differences, which are observed among the sex ratios at births of different countries, or different groups of population (e.g. legitimate or illegitimate births) were caused by a different frequency of abortions, among which a higher number is for males. This thesis has been widely accepted and confirmed by valuable statisticians such as Tschuprov and Wedervang, but on the other hand, it had been criticised by Ciocco and Boldrini on the basis of more precise studies. They demonstrated, if not its lack of foundation, certainly its insufficiency.

On the other hand, Poisson and von Bortkievicz had pointed out that a normal dispersion of a chronological series of frequency ratios generally, and of sex ratios at births in particular, may occur even if the probability of the event vary from one group of population to another when partial probabilities, related to single group, do show normal dispersion in time and the groups of the population to which they refer also vary with normal dispersion. Because, if such groups remain numerically unchanged, or vary with a hyponormal dispersion, the dispersion of mean probabilities of the series is also hyponormal.

On the other hand, another application of the probability calculus has already shown that the probability of conceiving the two sexes is not only different from one group to another, but also from one family to another. In effect the profound differences which are found in the sexual composition of brothers and sisters are instead, as I have already mentioned, largely dependant on chance, but for a non-negligible portion (that would result, from some investigations, between 10 and 15 %) from a different tendency of the various couples to conceive the two sexes[13]. A further inductive application of the probability calculus, by using the method of results, showed that such a tendency of single couples to conceive one rather than other sex, does not remain constant during the life span periods, but it does show a tendency to vary, i.e., for future births, there is a probability that the sex which was previously less represented, will occur more often[14].

Incidentally, the dispersion of sex ratios in hatches of multibirth mammals does instead generally result as hyponormal, probably in relation to the fact that, among the androgynous or gymnogynous contiguous spermatozoon in the female organism a trace still exists of the rigorously equalitarian repartition that occurs among them at the moment of maturity[15].

Another tendency to compensate occurs among sex ratios at births registered on successive days. It also emerged from an inductive application of probability calculus, based on the comparison of theoretical probabilities with the actual frequencies of increasing or decreasing sex ratios for successive days, as resulting from the records office[16]. It is deduced that the sex ratio at births on one day is not independent from that of the previous day, but tends to vary in the opposite sense, which can be explained by the fact that if on one day more males appear to have been registered, either due to an accidental variation or to a shifting in registration, less remain to be registered in the days immediately following.

But, as the interval of time gets wider and wider, from one day to ten and from ten days to longer intervals, very soon the compensating tendency disappears and for period of months, years, five years, ten years, leads to a continuously increasing tendency, which finds its natural explanation in the tendency of the sex ratio to vary in a specific direction. When these compensative and continuous tendencies

---

[13] C. Gini, *Il sesso dal punto di vista statistico*, op. cit.

[14] *Il sesso dal punto di vista statistico*, op. cit.

[15] C. Gini, *Combinations and sequences of sexes in human families and mammal litters*, Acta genetica et statistica.

[16] C. Gini, *Sulla probabilità che X termini di una serie erratica siano tutti crescenti (o non decrescenti) ovvero tutti decrescenti (o non crescenti) con applicazioni ai rapporti dei sessi nelle nascite umane in intervalli successivi e alle disposizioni dei sessi nelle fratellanze umane*, Metron, vol. 17, n. 3-4, 1055.

act within the single terms of the series, the dispersion of which is being measured, they may, as we said, be compensated by other tendencies, without altering the normality of the dispersion. When instead they involve more terms of the series, they cause variations in the series which make the dispersion hypernormal. The slight hypernormality, that has been found in the chronological series of the sex ratios in human births, is probably due to such variations.

If one admits - and it is not difficult not to do so - that the tendency to conceive one sex or the other is influenced by the conditions of the natural environment, the finding of almost absolute stability of sex ratios in time suggests that both the parents influence the sex of children, and the variations in their attitudes, determined by environmental factors, compensate because otherwise the incidence of the sex that has a prevailing influence wnould be increasing through generations[17].

Such a conclusion is not in contrast with the well-established fact that in the human species, as in many other species, there is a male etherogamethogenesis, so that two types of spermatic mature (androgynous and gynmogynous) because the conditions of the ovule, and everything leads us to believe it is so, may have a decisive influence on the sex of the foetus, giving more or less access to one type or other of spermatid[18].

But why - one may ask - not only in social sciences, but in physics as well, are schemes continuously renovated and, shortly after a scheme has been adopted, is it put aside as insufficient or erroneous?

This depends on various causes, partly due to deficiencies in the human mind and partly to the progress of science.

First at all it depends on the progress of our knowledge. One cannot accuse Quetelet of being wrong if he believed that all the individuals of one species had the same hereditary patrimony, nor one can accuse Lexis if he admitted that the sex of the newborn depended on the female ovule and the probability of conceiving one sex or the other, for the various women, was constant, or at most varied only accidentally. If these theses had to be abandoned, it was due to the progress of the biological sciences, which could not be foreseen by Quetelet and Lexis.

Another reason for which a scheme is often abandoned is because we have become stricter regarding the agreement that theoretical schemes must have with the observed data, either because the observed data have become more accurate, or because our studies have become more refined. On this depends the fact that Quetelet, and after many others, for a long time assumed that the curves of height or other anthropologic characters exactly followed the curve of random errors, while for height, given the multiplicity of its applications, it can be stated for sure, and for many other anthropologic characters, that it is likely that the distribution is hypernormal.

---

[17] C. Gini, *Il sesso dal punto di vista statistico*, op. cit.

[18] C. Gini, *Sul rapporto primario dei sessi nella specie umana*, Atti della 19 Riunione della Società Italiana di Statistica, Roma, Failli, 1961; *Maggiore frequenza del maschi nei concepimenti e maggiore mortalità dei maschi durante la gestazione, nel parto e nelle prime settimane di vita*, Atti della 21 Riunione della Società italiana di statistica, Roma, Failli, 1962.

But one must not hide that other times the scheme must be abandoned because in reality, while formulating it, one did not consider a hypothesis which is instead essential to it, or because one hypothesis was perceived, but it was considered of negligible importance, while it is an important one, if not a decisive one.

Now I am reaching the subtle part of my speech.

Are we sure that in the applications of probability calculus, not only to economic, judicial, moral and social phenomena, but also to physics and to gambling games, its unchallenged ultramillenial dominion, some hypothesis that so has far eluded us, is not present here?

Let truth win. It is said that the probability calculus has the task of determining the laws of chance, however realizing that the expression may appear as a contradictory one. In effect, the essential task of the probability calculus is to determine the impact of accidental factors. However, the probability calculus is only the subjective expression of combinatorial calculus: there is no formulation of the probability calculus that cannot be expressed in terms of combinatorial calculus.

Chance does not enter into the combinatorial calculus. All the combinations are ruled by strict, intransgressible laws. How can one assume to be able to calculate the impact of chance on the basis of calculation that excluded it? Where does the logical jump lie, for which one goes from the combinatorial distributions to random ones? This is the problem.

I believe the jump is in the so called theorem of joint probabilities[19].

In 1908 I was already noticing that this theorem is only an approximate one and I deduced, although without going deeper into the problem, that almost all the conclusions based on it must be approximate as well[20]. Let it be so: I believe one would not find it very difficult to admit it, but there is another more difficult problem. Are the following divergences accidental, hence negligible for a great number of observations, or are they systematic ones? It is known that there are authors who supported that results of gambling games do not conform to the forecast of probability calculus. Among others, a professor of psychology of Wurzburg University, Prof. Marbe, wrote hundreds of pages on the matter with conclusions that have been criticized and opposed in some other hundreds of pages by famous probabilists. Regarding this specific situation, I feel that the critics were right, although in effect it was the case of applications to particular measurements of collective phenomena, which do not solve the problem. This deserves to be reconsidered.

Why is the theorem of joint probabilities only an approximate one? Its proof is known...One starts from the measurement of the probability given by the so-called mathematical definition, but the conclusion does not change if starting from the empirical definition.

If m are the favorable cases and a the possible cases of an event a, and r and s respectively the favorable cases and the possible cases of an event b, which is the probability of their concurrence, assuming that the two events are independent,

[19] C. Gini, *Alle basi del calcolo della probabilità (Considerazioni sul problema della probabilità composta e sulla previsione dei fenomeni aleatori)*, Metron, vol. 23, 1964.

[20] C. Gini, *Che cos'e la probabilità?*, op. cit.

that is to say randomly associated? As the two events are independent, any of the possible cases n of event a and in particular any of the m cases favorable to event a will equally associate with any of the s possible cases of event b and, among these, with any of the r cases favourable to event b. Hence there will be mr cases favorable to the occurrence of phenomena a and b on ns possible cases. However the conclusions "there will be" does not logically follow from the premise. It would follow if we did substitute the expression "it might equally associate with" by the expression "must equally associate with", but, if the cases of phenomenon a were to combine with the cases of phenomenon b in a specific way, then they would not be any more independent and, if they are independent, one cannot say that they must combine in one way rather than another.

Hence, in the theorem of joint probabilities the random distribution for which the theorem is not valid is inadvertently substituted to the combinatorial distribution for which the theorem is valid.

What it appears that we can state is that the random distribution tends to the combinatorial distribution with the increasing number of the considered cases. In other words, the random distribution may be considered as a sample taken randomly from the combinatorial distribution representing the universe.

Let us now recall that the expected value or the sample average corresponds to the mean value of the population. Let us also remember that Prof. Marbe's applications and their related criticism substantially concern the probable or mean frequencies of the various results in gambling games. Hence, without closely studying his lengthy publications and their voluminous criticism, however I am *a priori* inclined to acknowledge that for the specific case Marbe was wrong and his famous critics were right.

However, the probability calculus does not only apply to the determination of mean or probable values.

Its task is to determine other constants as well, for instance the measurements of variability and association. As it is known, the probable value of the sample variance, measured with respect to the sample mean, is not equal to the variance of the population, but it is systematically smaller, and the expected value of the correlation in the sample is not at all equal to the measurement of the correlation in the population, but it is systematically greater.

It may be said that the probability calculus also teaches us how to measure the difference between the probable value of the sample variance and the value of the population variance and similarly the difference between the expected value of the sample association and the value of the correlation in the population. However, one must not forget that the validity of the instrument for such a measurement is already made false by the suspicion of error. Substantially, by thinking well about it, when the probability calculus gives the measurement of the amount of accidental factors, it sets a more complicated combinatorial distribution against a combinatorial distribution; it does not set a random distribution against a combinatorial distribution. For instance, when one speaks of a sample distribution (meaning random samples), in reality one says something which does not exactly correspond, because the samples are not randomly chosen, but they are built from the population in all the possible ways, according to the rules of the combinatorial calculus, which exclude the intervention of chance. One has not got a random distribution, but a combinatorial distribution. From this, a random distribution may be obtained by a random extraction of a more or less large number of samples; but the divergences identified for the various statistical constants among the random distribution obtained in this way, and the combinatorial distribution from which it is derived, cannot be defined theoretically, but only empirically. It must be acknowledged that to speak of laws of chance implies not only a verbal but a substantial contradiction, and that is an illusion to try to eliminate it.

But, if the impact of the divergences cannot be avoided, can we hope to minimise it by multiplying the applications? Certainly, it may be done. I hope so as well. With the hope comes a comforting fact that often, in the procedures by successive approximations, a consciously a known incorrect quantity is introduced in the calculation, the disturbing influence of which might however be progressively reduced as the calculus proceeds. This may well happen for the substitution of the combinatorial distribution to the random distribution, as one made in the theorem of joint probabilities. Objectively though, one must also remember that, in other cases, the introduction of an incorrect quantity has perturbing effects, which get worst in the following elaboration. It remains only to appeal to experience, while carrying out more detailed verifications.

Our lottery games would be ideal for such verifications, because its mechanism and the rigorous control to which it is subjected, exclude that, as on the basis of any reasonable opinion, it is not a case of mere gambling, as was objected to some of Marble's tests carried out on the results of roulettes. I have started some elaboration on data of lottery extractions, that I could get hold of. I hope to be able to complete the data collection and the following elaborations. Those so far carried out do not contradict either the tendency of the probable values in random distributions towards the mean value in combinatorial distribution, either the systematic divergences among the values of variances and correlations. However, some of the results need to be analyzed and interpreted. I believe it is worth carrying on.

Mathematicians did not only apply the probability calculus to gambling games, but they also applied it to the decimal series of some numbers. The best known applications are those for the series of the first 707 decimals of (calculated by Shanks), applications made either by Sorel or, with more details, by Czuber and Cassinis. These last two authors separated the first 660 decimals in groups, each of 60 decimals, and they made 10 series of 11 terms, each of which expresses the frequency by which the related figure appears in the 11 groups, and vice-versa 11 series, each of 10 terms, each expressing the frequency by which the 10 figures appear in the related group.

The result was that the distributions of these decimals fit rather well the theoretical distributions according to the Bernoulli scheme. The actual standard deviation for the two groups of sequences remains slightly below the theoretical standard deviation, leading to a slightly hyponormal dispersion index and the number of figures within the probable divergence equals, or is slightly above that of the figures left out.

What conclusion can be drawn?

It is certain that the decimal series of π is not a random one, but it does occur according to a specific law (precisely the law that allows for its calculation); but the complication of the law is such that the figures distribution is barely uniform with a combinatorial distribution. However, it was noticed that the frequency of figure 7 resulted as exceptionally low, either in the first 660 decimals, or in all the 707, so far calculated. In the series of the 707 decimals, it represented a divergence equal to 2.335 times the standard deviation, which would correspond to the theoretical probability of occurring, not once in 10, but once in 1000 terms.

After the publication of these applications between 1920 and 1930, many other decimals of it were calculated, reaching the number of 100,000. Recently I have begun some elaborations, so far limited to the first 3000 decimals. Already in the first thousandth the frequency of figure 7 ceases to be exceptionally low, being higher than that of figures 4 and 6 and, in the second thousandth it goes over that of figures 3, 0, 1 and 8.

After separating the 3000 decimals into 30 groups, each of 100 terms, the frequencies of the 10 figures, for each of them, were classified. Then, 10 series, each of 30 terms, were set up, each series expressing the frequency with which the respective figure occurs in the 30 groups. The dispersion indexes remain higher than one for two figures (1 and 9) and lower for the other 8, with an average slightly lower than one (0.946), thus confirming the slight hypernormality of the dispersion. If one examines how the figures follow one another, however, the result is that this does not occur exactly according to the combinatory calculation, or, as it is usually said, they are not completely independent. If we separate the 10 figures from 0 to 9 in "low" (from 0 to 4) and "high" (from 5 to 9), we find that more often a high figure or a respectively low one follows a low figure or a high one (contrasting sequence). In the first thousandth, the sum of the contrasting sequences would be 520 against 480 concordant ones: in the second thousandth, 518 and 482 respectively; in the third, 534 and 466.

The lack of independence, with a tendency to contrast, among the successive figures, is confirmed by the frequency by which the repetition of the same figure occurs in two successive decimals. According to the theory, this should occur 100 times on 1000 decimals; it does instead occur 98 times in the first thousandth, 93 in the second, 76 in the third.

Hence the independence among successive events, that is one of the hypotheses of the Bernoulli scheme, is not verified.

Certainly, if the compensating tendency among successive figures does not cause the hyponormality of the dispersion, it does contribute to it.

More elaborations could be carried out, however the results do not seem doubtful to me. They confirm our thesis that the inductive applications of probability calculus must not be limited to the comparison between the actual distribution and the combinatorial distribution, globally considered.

This helps us to realize the superficial and external likeness of the two distributions, thus corresponding to the descriptive aim of inductive applications of the probability calculus, but it does not help us to understand the internal structure of the two phenomena, that is, the heuristic function of such applications. For this I believe the progress of statistical researches must expect a lot.

Summary

The Author goes back to Founders of Probability calculus to investigate their original interpretation of the probability measure in the applications of the probability theory to real problems. The Author puts in evidence some misunderstandings related to the inversion of deductions derived by the use of probability distributions for investigating the causes of events.

Keywords: History of probability; Statistical induction; Frequency and probability