Statistical model, sample size, and sample composition

1.Mr. Gorekh Prasad Nayak 2.Soubhagini Mohapatra
Nalanda Institute of Technology, Bhubaneswar
Dept. of Basic Science & Humanities
E-mail ID: gorekhprasad@thenalanda.com
SoubhaginiMohapatra@thenalanda.onmicrosoft.com

ABSTRACT:

For a subjectivist, a single event describes a "fact," which is something that takes place in a specific location at a specific time and is preceded, surrounded, and followed by facts that are fundamentally distinct from one another because they take place in various locations or at other times. For instance, the phrase "Mr. Tito Rossi, born and lived in Milan in Ronchiano square 25, has obtained white hair" refers to a particular event. If our Mr. Rossi has white hair or not will determine whether this assertion is accurate or incorrect. These kinds of statements provide the most details on a person and a modality. a collection of related claims for each member of a population, often known as individual vectors.

INTRODUCTION:

In a previous paper [Costantini, D., Garibaldi, U. and Monari, P. (1998)] we have given some conditions that allow to transform single events in statistical or generic ones. It is impossible to do statis- tics limiting attention to single events. Neglecting information about individuals while collecting information about the population the indi-

viduals belong to, single events are transformed into statistical ones. This transformation ensues from the formal conditions discussed in the paper quoted above. These conditions essentially determine the invari- ance of probability with respect to certain descriptive changes. Thus statistical events refer to individuals only as belonging to a popula- tion. By using the term statistical unit we intend to speak about such individuals. At the basis of the analysis of the concept of statistical event lies the conviction that this notion can be usefully applied in order to give a satisfactory foundation of statistical inferences able to overcome the dilemma between orthodox (frequentist) and Bayesian statistics.

This is the first reason why we want to work out the present pa- per. The second one is that in modern Physics, namely in quantum mechanics, there are fundamental laws, such as the Schödinger equa- tion, referring to single statistical units. May be the best example is the one given by Dirac: "Some time before the discovery of quantum mechanics people realized that the connection between light waves and photons must be of a statistical character. What they did not clearly realize, however, was that the wave function gives information about the probability of *one* photon being in a particular place and not the probable number of photons in that place" [Dirac, P. A. M. (1931),p. 9].

On the basis of our previous analyses, which we will take as known, the present paper shall discuss a new interpretation of proba- bility, essentially Popper's notion of propensity. This interpretation is very close to the probability interpretation used naively by physicists, whereas it seems to be quite unknown to statisticians. Secondly, we shall try to interpret the notion of statistical model making reference to the probability of statistical units.

1. PREDICTIVE ALgORITHM

In what we are doing predictive probabilities are crucial. Thisterm, at least in the Italian literature, is normally used in a very narrow sense. In order to avoid this restricted use, we prefer talking about predictive algorithm. First of all, we shall hence see what this means. We consider a sequence

We call predictive algorithm all procedures that, staring from suitable con- ditions aim at determine the value of the probability distribution .
Moreover by using the term predictive algorithm we intend also to refer to all deductions that starting from try to characterize more com- plex probability distributions, for instance the ESF (Ewens Sampling Formula) [Costantini, D. and Garibaldi, U.]. We shall call the con- ditional probability an 1-predictive law. Obviously, when n
is to be understood as an absolute probability, that is Pr X_1 x .
Moreover we shall call *m*-predictive law the joint distribution

It goes without saying that, once for each n is known, the observation stochasticprocess — as suitably Daboni and Wedlin.

From an abstract point of view, is neither bound to one nor to the other probability interpretation. Therefore we want to point out that (2) may be interpreted both subjectively, as a degree of belief in the occurrence of the event X_{n+1} x, and as the probability that at time =
n 1 a physical system is in the state x. To avoid any misunderstanding,
let us add that we prefer the second interpretation, because, at leastsince the first probabilistic explanation of the Brownian motion carried out by Einstein almost one century ago, it is known how well stochastic processes describe the (probability) dynamics of physical systems. This shows, beyond any doubt, how well the predictive

algorithm adapts itself to the treatment of physical problems. This kind of problems are certainly not linked to the mind of any human being nor to its degree of belief but simply to how things are.

Hence, predictive algorithm has nothing to do with the subjective philosophy of probability, namely not with the extreme subjectivism,

very near to the solipsism, worked out by de Finetti. This philosophy permeated and stirred many fundamental discussions taking place when the predictive algorithm was developed. For obvious reasons of spacein the present paper we cannot dwell on the reasons which clearlysupport this statement. We will confine ourselves to remind the reader the following:

One the one hand, the determination of is possible after having introduced clear (probability) conditions stating the role of the indi- vidual vector $\mathbf{X}^n$ $\mathbf{x}^n$ upon the probability of the random variable X_{n+1}. On the other hand, we have been able, by using , to laybоth the probabilistic foundations of elementary particles statistics and to give a probability representation for the dynamics of a gas of parti- cles. Interested readers are asked to consider our analyses [Costantini, D. and Garibaldi, U. (1997); Costantini, D. and Garibaldi, U. (1998); Costantini, D. and Garibaldi, U. (2000)].

2. ENSEMBLE AND SAMPLING DISTRIBUTION

Taken for granted the possibility to speak of statistical events in the sense specified in [Costantini, D., Garibaldi, U. and Monari, P. (1998)], we have to interpret these events. In order to do this, it is necessary to discuss the propensity interpretation of probability. This is the last, in order of time, interpretation of this notion which has been taken into account. But before this we have to face a firm conviction widely spread among orthodox statisticians. It is about what we could call the frequency prejudice, i.e. the belief according to which the probabilities which do not refer to collectives are meaningless; in other words, the belief according to which probabilities are always tied to frequencies. As a consequence, of this prejudice the referenceto a statistical unit compels probability to be subjective. Hence the probability that a statistical unit bears an attribute cannot be used in scientific researches. The frequency prejudice rest on two points: the notion of collective refers to an infinite sequence of observations [von Mises, R. (1961)] or to a hypothetical infinite population [Fisher, R. A. (1950)]; the notion of probability refers to relative frequencies. As a matter of fact, this way of seeing probabilities and the frequentist prejudice can be dated as they reflect an approach to the statistical- probability disciplines linked to the debates, which took place in the first half of the last century. If fact, the more shrewd modern literature on statistical inference has abandoned both that interpretation and the frequency prejudice. But this has taken place in silence, without really facing the questions it raises. On the contrary, we intend to discuss frankly about the implications of the above renounce.

It is enough to glance through a few books on modern physics to realize when probability referring to statistical units are used. This causes the rise of the propensity interpretation of probability of K. Pop- per. This author puts forward "*a new physical hypothesis* (or perhaps a metaphysical hypothesis) analogous to the hypothesis of Newtonian forces. It is the hypothesis that every experimental arrangement (and therefore every state of a system) generates physical propensities which can be tested by frequencies" [Popper, K. R. (1959)], consequently "I propose [. . .] to look upon probability statements as statements about frequencies in *virtual* (infinite) sequences of well characterized experiments" [Popper, K. R. (1967)] and therefore "we can look upon probability as a *real physical property of the single physical experiment* or, more precisely, of the *experimental conditions* laid down by therule that defines the conditions for the (virtual) *repetition* of the ex- periment" [Popper, K. R. (1967)].

In order to fully understand this interpretation, it is necessary to consider it in the context of the researches of quantum mechanics. Let us return to Dirac's light waves and photons by considering abeam of linearly polarized light going through a Polariod filter whose polarization plane is oriented at 45° with respect to the light [Ghirardi,
G. C. (1997)]. As we know from experience, if we measure the
intensity of the beam emerging from the filter, we find, according to Malus' law that tied the transmitted intensity to the incident one, thatthe intensity is reduced by a factor $\cos^2\theta$ being θ the angle betweenthe polarization planes. Everything is quite clear, if light is understood, as it happened in the 19 th century, as a classical electromagnetic field and hence as a continuous quantity. The situation becomes unclear, however, if, according to quantum mechanics, the beam is understoodas a population of indistinguishable photons which cannot be subdivide into parts.

If the assertion put forward by quantum mechanics regards the whole population, it would be possible to maintain that half of the photons get absorbed by the filter while the other half continues onits way without being disturbed. But this theory does not make any statement for the population. It answer the question: what happensto a single photon that runs into the filter? The reason is that it is

possible to launch such a weak beam of photons at the filter that each photon has enough time to be absorbed or to go through the filterbefore another one cuts it. The answer of quantum mechanics is that each photons has a probability $1/2$ to pass through the filter.

Physicists refer to the probability of a particle. This gives rise to a probability notion objective but not frequentist. This is the no-tion of propensity.

On the other hand we know that from Maxwell to Boltzmann, from Gibbs to Tolman everyone taking an interest in statistical mechanics maintains that probability is (the limit of) a rel- ative frequency. It seems to be unnecessary to discuss about this if itis clearly known what relative frequency refers to. The answer seems easy but it is not. The temptation to give simple answers has caused a great number of misunderstandings and on top of the list is the frequency prejudice. We are convinced that the best way of dealing with the question is to reflect upon Fisher's sample space or, better, upon the ideas of Gibbs. What we are about to say has a double pur- pose. Firstly, to show that relative frequency is a very sophisticated notion, that cannot be reduced to a fractional number of individuals. Second, that both in orthodox statistics and in statistical mechanics the 1-predictive laws play a fundamental role.

While introducing the notion of sample space, Fisher states "If the individual values of any sample of data are regarded as co-ordinates in hyperspace, then any sample may be represented by a single point, and the frequency distribution of an infinite number of random samples is represented by a density distribution in hyperspace" [Fisher, R.
A. (1950)]. Therefore, according to this author, the distribution of (relative) frequencies is represented by a density function linked to an infinite number of random samples. But which is the nature of this collective (of samples)? The best way of answering this question isto directly refer to the *ensemble* of Gibbs, i.e. to the notion which Fisher and all statisticians after him have adopted by calling it sampling distribution.

For disciplines, like statistical inference, which refer to density on a sample space, the interpretation of a macroscopical physical system as a population of particles is not important. As a consequence statis- ticians do not look out to the structure of the collectives of statistical mechanincs. This caused problems of interpretation, in fact in statis- tical mechanics a collective is not a population of particles, a system as physicists say, but the whole set of diverse dynamic configurations that the system, globally interpreted, is suppose to be able to assume.

Thus Gibbs frequency interpretation does not depend upon the size of the considered system, that could perfectly well been made up of only one particles. This is an extremely important acknowledgment on which most of what we are going to say will be based. Therefore,it is necessary to justify it in depth.

At the basis of mechanics lies to notion of microscopical state. This state has traditionally been understood as a point in a space of an F-dimensional space, the phase space Γ, a point denoting the configuration of the physical system of interest. If the system is a monoatomic gas the dimension of the phase space will be $F = 6N$, where N is the number of gas atoms and 6 are the degree of freedom of each atom, three for the coordinates and three for the momenta. According to Tolman [Tolman, R.

C. (1979)], statistical mechanics "can be employed when we need to treat the behaviour of a system concerning whose condition we have some knowledge but not enough for a complete specification of precise state. For this purpose we shall wish to consider the average behaviour of a collection of systems ofthe same structure as the one of actual interest but distributed over a range of different possible states. Using the original terminology of Gibbs we may speak of such a collection as an *ensemble of systems*[. . .] The istantaneous state of any system in the ensemble can then be regarded as specified by the position of a *representative point* inthe phase space, and the condition of the ensemble as a whole can be regarded as described by a "cloud" of such representative points, one for each system in the ensemble." furthermore "the condition ofan ensemble at any time can be regarded as appropriately specified by the *density* ρ with which representative points are distributed over the phase space." This means: if for each state a of Γ, $M(a)$ is the number of the systems in the state a, then when M grows beyond any limit, what happens to the fraction $\frac{M(a)}{M}$ identifies the density ρ on Γ.

If, as usually happens in statistical mechanics, the considered sys- tem is a population with a finite number of particles, the ensemble can be imagined as an infinite (super)population whose units are population (finite ones) of particles where the dynamic characteristics, the position and the momentum of each particle, are perfectly defined. It follows that "By the *probability* of finding the system in any given condition, we mean the fractional number of times it would actually be found in that condition on repeated trials of the same experiment" [Tolman, R. C. (1979)]. Therefore, to (probabilistically) treat the physical systemof interest, we think of it as surrounded by an imaginary — Popper

would say virtual — population of systems having the same structurebut distributed over the whole of all initial states compatible with the constraints of the system. Hence in statistical mechanics probabilityis seen as a (fractional) frequency resulting from the repetition of the same trial and not the (fractional) number of individuals bearing an attribute. Furthermore the density on the phase space is an 1-predictive law applied to a statistical unit denoting the considered physical sys- tem. The latter is a population of N particles that we suppose to have randomly drawn from a virtual (super)population. The distribu- tion of this virtual (super)population is given by the above density ρ.It is worth noting, and here we shall anticipate what we are going to say, that a similar hypothesis amount to the fundamental postulate of statistical sampling.

Now, if we consider the density ρ on the phase space and remind that the considered system is a population with $6N$ degrees of freedom, it becomes clear that the density $\rho(x_1, \ldots , x_F)$ is completely similar to the sampling distribution of a sample whose size is F . Fisher's sample space is the statistical transliteration of Gibb's phase space. The relative frequency specifies the condition of the sample space, i.e. the density with which the samples are distributed in the sample space. The sampling distribution is the fractional number of times a given sample would be found if the same trial was repeateda very large number of times. In fact, it has to be noted that the“repeated trials” do not have to carried out, it is only supposed thatthey could be carried out. The collective is not real but virtual.

3. THE VIRTUAL NATURE OF THE COLLECTIVE

The propensity interpretation of probability does not remove the notion of collective but subordinates it to the conditions defining the probabilistic features of the experimental arrangement generating the outcomes. We have already reminded the reader that most of modern statisticians are persuaded that the reference to a sequence cannot be based on anything else than on the notion of virtual infinite sequences of experiments. But, and this is the point, none of these statisticians has spent much time on making clear how the above reference can be realized. The statistical impact of the passage from von Mises’ and Fisher’s frequency interpretation to sequences of experiments has been analyzed only rarely, to use an euphemistic expression. It would have

been worth doing so, though, considering that the consequences are important. We will discuss this topic later. But to analyze this passage we have to refer once more to von Mises as he is one of the veryfew authors, may be the only one, who has dealt with this difficult problem. He has done so with great intellectual honesty. Hence, instead of referring to more recent authors, we must consult this one despite the fact that he worked in the first half of the past century.

Let us follow von Mises in a comment to an example linked to the theory of dispersion put forward by Lexis: “The German economist,
W. Lexis, has enriched statistics with an idea which leads to a very convenient method of comparing an observed statistical sequence of figures with a collective. I will explain Lexis’s Theory by means ofa small statistical investigation” [von Mises, R. (1961)]. von Mises considers the recurrence of the letter “*a*” throughout the first 2000 letters of *De bello gallico*. Note that it is likely that the research does not relate to the letters of *De bello gallico* but at least to all writings of Julius Cesar. Thus the first collective include all those letters. Having identified the data, von Mises firstly subdivides, according to Lexis, the observed sequence into 20 groups of 100 letters each and calculates the frequency of the letter he is interested in for each group. In thisway von Mises considers the initial part of a second collective whose terms are groups of 100 Latin letters. The categories that interest von Mises in this collective is the frequency of the “*a*”s. Once the frequency distribution of the letter has been determined through the20 groups, he calculates the average of the “*a*”s per group, which is equal to , nd then the variance, that come out to be 7,01. By doing this von Mises considered a third collective whose elements are made up of 20 groups

20 =

of 100 letters while the attributes are the variances. Before going on, it is worth noting that while the available terms of the first collective are 2000, there are 20 terms of the second one and only one term of the third collective.

Having defined the collective, von Mises imagines the distribution he would obtain if he put a great number of marbles with the 25 Latin letters into an urn — the proportion of marbles with the letter "*a*" is supposed to be equal to 0,087 — carrying out groups of 100 draws.The collective obtained in this way, the second one, would be madeup of groups of 100 draws from the hypothetical urn. Finally a third collective is supposed to be make up in order to calculate the variance that we would obtain if we had to draw sequences of 20 terms of 100 drawn from the second collective.

After having considered in the details the test of significance de- vised by Lexis and clarified by von Mises, it is worth reflecting uponits crucial points. In the first place we have to point out the prolif- eration of collectives von Mises recurs to in order to safeguard the frequency interpretation of probability. Secondly it is important to note the features of the collective that allows evaluating the hypothet- ical probabilities of the observed sample, namely the attributes of the members of the collective. These are functions — statistics or, much better, summaries — whose argument are the 2000 observed letters.The terms of the collective are made up on the basis of the 2000 let-ters that have been taken into account. The attributes are all possible variances calculated the way we have seen. Therefore the final collec- tive is an infinite sequence of groups of observation of 2000 letters, where each is, so to say, labelled by the respective variance. Finally note that this collective does not really exists, as nobody would dream of laying down the urn and the corresponding marbles with the Latin letters. Nor would it occur to anyone to carry out the above groups of draws. Therefore, it follows that: the collective that we have taken into account is not real but virtual; the collective has nothing to do neither with the size of the population nor with the one of the sample; finally, the comparison is carried out as if the observed sample hadbeen drawn at random from the virtual collective. Not even von Mises was able to interpret the probabilities involved in a test of significanceas effective (relative) frequencies.

Now that the statistical consequences of a coherent use of thenotion of relative frequency in a collective is clear, we can think ofthe consequences of such an use in the case of Dirac's photons. If,as we know from experience, the relative frequency of photons thathave been observed by a Polaroid filter was very close to 1/2, at the very most we could conclude that this could be the probability of a photon as far as the experimental arrangement is concerned. What has just been said would have to be based on a hypothesis stating thatthe photons, not only the ones taken in account in the experiment, are considered as independent and identically distributed random variables. All probabilities involved would appear as relative frequencies if the sole experiment having really carried out were indefinitely repeated.The notion of propensity is the acknowledgment of the change that has been taken place in the way of understanding probability when pass- ing from statistical mechanics to quantum mechanics. It also means accepting the impossibility to apply the frequency interpretation of probability to quantum mechanics. Therefore, in order to avoid ex- cluding the notion of probability from the more advanced scientific discipline, it was necessary to accept a notion of probability not fre- quentist but objective. This is what Popper tried to do by suggestingthe notion of propensity.

This is the historical and scientific background of propensity, at least from our point of view. But we immediately want to add that Popper's proposal is not completely convincing, as it does not takeinto account the achievements of the last century's statistical method- ology. This means that — by characterizing each time, through suitable probability distributions expressed by the sampling and the statistical model, the probabilistic features of the experimental arrangement — it will be possible to identify what Popper would call the propensity of the experimental arrangement.

But this has led us to the points we want to discuss in the next two sections. Before concluding the present section, though, it is worth stressing that from what we have said ensues that 1-predictive lawscan be objectively interpreted.

4. SaMPLINg

There are still two problems we have to deal with: the sampling and the statistical model. We dedicate this section to the first of these subjects. Let us take Popper's explanation of why he leaved the fre- quency intepretation as a starting point: "The frequency interpretation always takes probability as relative to a sequence which is assumed as given; and it works on the assumption that a probability is *a propertyof some given sequence*. But [. . .] the sequence in its turn is defined by its set of *generating conditions*; and in such a way that probability may now be said to be *a property of generating conditions*" [Popper, K. R. (1959)]. It is clear that there is quite a difference between this and the viewpoints of von Mises and Fisher. If probability is a property of generating conditions, then it can be considered as well defined ev- ery time that such conditions are specified, that is every time that an experimental arrangement is carried out. von Mises and Fisher do not accept that a probability can be assigned to anything different froma sequence. These authors do not accept that a probability can berelated to experimental arrangements. But this is the only way to allot probabilities to statistical units. We sall see that, as a consequence of this refusal, they shut out the possibility of giving an objective interpretation to the notion of statistical model.

It is known to everyone who has reflected upon the function of sampling in statistical inference that there are two ways of justifying random sampling: one that does not take the parent population into account; and another one, which tries to take into consideration this population. On the whole, the latter approves of extremely confused positions.

The first justification basically refers to Popper's notion of ex- perimental arrangement. This arrangement can be identified in the device used in order to identify the statistical unit to be drawn. To this regard, let us remind the explanation of Herzel, a scholar who has worked out very useful analyses of these problems. This expla- nation shows the substantial identification of the sampling with theexperimental arrangement. "Si ricorre a tali strumenti [tavole di nu- meri casuali o simili] per evitare le operazioni piuttosto onerose che sarebbero richieste per effettuare delle vere e proprie estrazioni da urne, come vengono eseguite nelle lotterie" (The reason for recurringto such kind of instruments [random numbers or similar] is to avoid the rather complicated operations that would be required to carry out real draws from urns as it happens in lotteries) [Herzel, A. (1994)].

Let us now go on to the second way of justifying random sam-pling that has, as we pointed out, lead to questionable if not con- tradictory assumptions. This is the justification chosen by Kendall e Stuard [Kendall, M. G. and Stuard, A. (1963)] who, when talking about the population from which the sample is drawn, divide them into the following three classes: finite and real ones, infinite and real ones — let us leave it to the authors of this classification to find similar populations — and infinite and hypothetical ones. When it became clear that the terms "infinite" and "real" were contradictory [Stuard, A. and Ord, J. K. (1987)], the sampling from infinite and real pop- ulation was substituted, without any justification, by the "sampling" from a distribution. Anyone can clearly see that the above passage does not resolve the contradiction. As a matter of fact: either the distribution concerns actual frequencies and hence there has to be a

finite population to which the distribution can be attributed to, which means returning to the previous case, or it is a distribution function, hence a mathematical function, from which it seems really difficult to draw anything.

In our opinion it seems much easier to keep for population the meaning a whole set of statistical units and to consider the sample as resulting from a suitable experimental arrangement generating se- quences of those units. If the sample is understood this way, then the size of the population is no longer influential. The size of the popula- tion has nothing to do with the possibility to perform infinitely many repetitions and hence with the (virtual) collective of which the sampleis supposed to be the initial parts. The objectivity of the sampling distribution is guaranteed by the experimental arrangement generating the sample. It would be this arrangement that would generate the col- lective in the case in which the draw would be indefinitely repeated.The generating conditions are reflected in each draw, that is in each term of the collective and not by the collective as a whole.

But for our purpose this is not the most important aspect. We are fully aware of the fact that the above statement represents a substan-tial return to the classical definition of probability seen in the light of the failure of the orthodox frequency interpretation. The classical definition can and maybe has to be understood as a kind of postu-late linking the (relative) frequency of the statistical units bearing an attribute to the probability of drawing a statistical unit bearing thatattribute. This is question we will discuss soon. For the time being we note the existence of a frequency of marbles of a given color in the urn; the existence of a frequency of individuals bearing an attribute ina finite population; the existence of a (limit of the relative) frequencyin the collective; and such a frequency exists, too, in a hypothetical infinite population. The experimental arrangement generates sequences of statistical units hence frequencies, and the probability of each sta- tistical unit is given by the 1-predictive law expressing the generating conditions. Coming back to Herzel, in the case of orthodox statistical

inferences the experimental set up "e` la presenza di un ben definito *piano di campionamento* che assegni una probabilita` ad ogni campione appartenente ad un certo insieme [. . .] che può anche essere costituito da tutti i campioni possibili" (is a well defined *sampling plane* allot-ting a probability to each sample belonging to a certain class [. . .]that can also be made up of all possible samples). Therefore, if the sampling plane is the device determining the choice of the statistical

unit, then this plane is tantamount to the conditions which generatethe sequence. At this point probability comes in if, as Herzel stateslittle later, we let "c il singolo campione: un piano di campionamento probabilistico consiste dei numeri non negativi $p(c)$, rappresentanti la probabilita` di osservare il campione c" (c denote the single sample: a probabilistic sampling plane consists in the non negative numbers $p(c)$ that represent the probability of observing the sample c) [Herzel, A. (1994)]. It follows that the probability of a sample can be understood as (the limit of) the relative frequency in the collective of the (virtual) infinite repetitions of the draw. The objectivity of the sampling distri- bution, whose (density of) probability regards statistical events — as a matter of fact they are single draws of samples from a population —, is ensured by the sampling plane, the statistical counterpart of the experimental arrangement. These considerations have lead us to the statistical model, i.e. the second point we want to discuss.

5. Statistical model

As we have seen, the necessity to speak about the probability ofa particle has compelled Popper to work out an objective concept of probability that does not need to be referred to real collectives. Quite on the contrary, in statistics things have gone another way. Followingde Finetti, both repetitions and objectivity have been abandoned. Inthe first half of last century many statisticians, urged by the ideas of de Finetti, leaned towards a subjectivism rooted in the irrationalism of the Italian relativism of the beginning of the 20th century. It has tobe added that the choice of this direction was favored by the kind of applications that were characteristic of the Italian school of statistics. In fact if, as it happens in Italy, statistical inferences are mainly atten- tive to social and economical studies, it is very easy that the criticismto the frequency dogmatism ends into the subjectivism.

As all prejudice, the frequency prejudice too, did not mince the matters. Orthodox statisticians maintained that probability was a rela- tive frequency in an hypothetical infinite population, but they did not take the trouble to analyze the consequences of their statements. They did not notice that the notion of relative frequency looses meaning pre- cisely at one of the key points of their inferences, i.e. as regards the fundamental hypotheses inference is based on: the notion of statistical model.

There are several ways of working out statistical assumptions but the most common one, for sure, consists in specifying the statistical model. As Edwards states in a celebrated book devoted to the conceptof likelihood, "A sufficient framework for drawing of inductive infer- ences is provided by the concepts of a *statistical model* and a *statistical hypothesis*. Jointly the two concepts provide a description, in proba- bility terms, of the process by which it is supposed the observations were generated. By *model* we mean that part of the description whichis not at present in question, and may be regarded as given, and by *statistical hypothesis* we mean the attribution of particular values to the unknown parameters of the model" [Edward, A. W. (1972)]. Forour purpose the distinction between model and statistical hypothesis is not important. Therefore speaking of statistical model we refer to statistical model together wit the statistical hypothesis.

First of all, we note that Edwards's point of view is the same as that of Popper as the process of Edwards is the same notion as that of experimental arrangement. Furthermore, the statistical model is the probabilistic description of the experimental arrangement. This is also our position.

Let us then go on to the statistical model and its probabilistic status. Edwards [Edward, A. W. (1972)] exemplifying the calculationof a likelihood says: "In section 1.2 it was suggested that a Normal distribution of error might be adopted as the probability model when the refractive index of a crystal was being investigated experimen-tally. That is, if μ be the true value and σ^2 the theoretical variance ofthe distribution of error, the probability of a single observation lying
in the interval $(x, x + dx)$.

In agreement with most statisticians — mainly with Fisher whosaid "Problems of specification are those in which it is required to specify the mathematical form of the distribution of the hypothetical population from which the sample is to be regarded as drawn." [Fisher,

R. A. (1950)] — we think that statistical models refer to the popu- lations or, better, to their statistical descriptions. Thus we must first consider the description of the population which we refer to or, better, an idealized description as neither infinite populations nor modalities with infinitely many attributes exist. Furthermore, the description may refer to a real population but it can also be a null hypothesis that has been put forward in order to be falsified. Obviously this is the most interesting case. Nevertheless, as Edwards clearly notes saying that (4) is the probability of a "single observation", the model is in not waythe description of a population.

Formula (4) is the probability that a statistical unit, say *X*, bearsan attribute of the interval $(x, x + dx)$ or, neglecting the differential, it is the (density of) probability of $X = x$. This is not a relative frequency. Justify (4) is tantamount to justify the leap that is performed by passing from the description of the population to the statistical model. As we have seen, physicists know quite well that they talk about the probability of a statistical unit, and they are well aware of the difficulties posed by this. The same cannot be said about statisticians, however, maybe because the consequences of the duality population- individual is really very far away from the one physicists have had to face or, maybe, because statisticians seem to have evaded the dilemma by withdrawing to subjectivism.

On the basis of what we have discussed in the above sections we shall find a way to avoid this withdrawal. This means that, once the existence of experimental arrangements is taken for granted, the link between the population, considered as a sequence of statistical units, and the statistical model can be established by postulating:

C. if in the sequence generated by the experimental arrangement, the relative frequency (density) of the statistical units bearing an at- tribute *x* of a modality is equal to $f(x)$, then the (density of) probabil- ity of drawing from the sequence a statistical unit bearing the attribute*x* is equal to $f(x)$.

We believe that the validity of this postulate is implicitly assumed at the beginning of every statistical inference, both orthodox and Bayesian. We are also aware of not having discovered anything new, as the condition **C** is essentially an extension of the classical definition of probability to which it becomes formally identical when, as every statisticians does, "random drawing" is understood as a drawing ruled by uniform distribution.

Having established the relationship between population and model

we can proceed with Edwards' example. He says [Edward, A. W.(1972)]: "hence the probability of obtaining a sequence of *n* observa- tions in the interval $(x_1, x_1 + dx_1; x_2, x_2 + dx_2; \ldots x_n, x_n + dx_n)$.

(once more the numbering of the formula is ours). This is the firststep towards the determination of the sampling distribution, and we can stop here to follow Edwards.

At this point some comments are required. First of all, the den-sity is a 1-predictive law. It is the initial 1-predictive law in which no index appears due to the assumption of the identical distribution of all observations. This law, together with independence, allows to determine . This is an initial *n*-predictive law, that is a distribution like , more exactly the particular case of that law when *n* 0.Hence the predictive algorithm is at the basis of the determination of the likelihood. This is true both for orthodox and Bayesian inferences. There is only one difference between both approaches, which is very important, beyond doubt. For Bayesian the condition **C**, and thus the related 1-predictive law, ensues from a personal judgement. This canbe based of factual knowledge but also on knowledge gained, so tosay, during sleep [Fürst, D. (1978)]. For an orthodox statistician con- dition **C** cannot be explicitly accepted. As we have seen, it is very difficult to justify by referring to collectives. This notwithstanding without such a 1-predictive law the test cannot start. The problems of specification come from this. On the contrary, taking for granted the propensity interpretation of probability, the justification of condition **C** is very natural. The 1-predictive law makes probabilistically explicit the conditions generating the sample.

6. CONCLUSIONS

The thesis put forward by followers of the frequency interpretation states that it is non allowed to speak about probability without referring to a collective. von Mises very clearly claims: "First the collective then the probability" [von Mises, R. (1961)]. Now, we think that this essentially dogmatic restriction can be dropped without giving way to subjectivism. Basically the use of collectives, and hence of relative

frequencies, has to be considered with respect to tests of significance,
i.e. to the procedure adopted in order to falsify statistical hypotheses. But we cannot expect each probability statement to be necessary given in terms of relative frequencies.

What we have done is the starting point for going away from the frequency prejudice without abandoning objectivism. In this way of setting out and dealing with the question there are two key moments. In the first place, the frank acknowledgment that an objectivist inter- pretation of sampling distributions must refer to relative frequencies. Nevertheless, it is of utmost importance to state that the reference has to avoid confusing the (finite) relative frequency, which describes the considered population, with the relative frequency of the virtual infinite repetitions. In the second place, the sampling distribution is arrived at by using a 1-predictive law. This law gives a description in probability terms of the experimental conditions in which draws are performed. The objectivity of this law is warranted by the generat- ing condition of an experimental arrangement. The objectivity of the sampling distribution ensues from the (virtual) possibility of drawing an infinite number of samples from the population. The size of the population is not influential at all.

Even though it is possible to repeat infinitely many times the draw of a sample, the acceptance of a statistical model as a description ofan actual population is assured by experience. This is the key idea atthe basis of every test of significance. Its justification is tantamountto the justification of experimental science.

To conclude, we are well aware that what we have done does neither include orthodox nor Bayesian estimation. The problem of the objectivity of estimation is much more complex. But although weacknowledge the great importance of statistical estimations, it was not the task we set ourselves.

REFERENCES

COSTANTINI, D. and GARIBALDI, U. (1997) A probabilistic foundations of elemetary par-ticle statistics. Part I, *Stud. Hist. Phil. Phys.*, 28, 483-506.

COSTANTINI, D., GARIBALDI, U. and MONARI, P. (1998) Dagli eventi singoli agli eventistatistici, *Statistica*, LVIII 191-211.

COSTANTINI, D. and GARIBALDI, U. (1998) A probabilistic foundations of elemetary par-ticle statistics. Part II, *Stud. Hist. Phil. Phys.*, 29, 37-59.

COSTANTINI, D. and GARIBALDI, U. (2000) A purely probabilistic representation for thedynamics of a gas of particles, *Foundations of Physics*, 30, 81-99.

COSTANTINI, D. and GARIBALDI, U. An axionatic approach to allelic equilibrium distribu-tions, *submitted to*, Theor. Popul. Biol.

DABONI, L. and WEDLIN, A. (1982) *Statistica. Un'introduzione all'impostazione neo-bayesiana*, UTET Torino.

DIRAC, P. A. M. (1931) *The Principles of Quantum Mechanics* , Clarendon Press Oxford.

EdwaRD, A. W. (1972) *Likelihood*, University Press Cambridge.

FISHER, R. A. (1950) On the mathematical foundations of theoretical statistics (1922)Contribution to Mathematical Statistics, *Wiley New York*, 10.309-10.368.

FÜRST, D. (1978) Tre uomini e un'urna, *Atti del convegno su: induzione, probabilità, statistica*, (Di Bacco ed.), Universita` degli studi Venezia.

GHIRARDI, G. C. (1997) *Un'occhiata alle carte di dio*, il Saggiatore Milano.

HERZEL, A. (1994) *Il campionamento statistico. Metodi statistici*, Monduzzi Bologna.

KENDALL, M. G. and STUARD, A. (1963) *The Advanced Theory of Statistics*, Vol. 1, IIed. Griffin London.

POPPER, K. R. (1959) The propensity interpretation of probability, *British J. Philosophyof Science*, 10, 26-42.

POPPER, K. R. (1967) Quantum Mechanics without "The Observer", *Quantum Theoryand Reality*, (M. Bunge d.) Springer New York.

STUARD, A. and ORD, J. K. (1987) *Kendall's Advanced Theory of Statistics*, GriffinLondon, Ch. 9.

TOLMAN, R. C. (1979) *The Principles of Statistical Mechanics*, (1938) Dover New York.

VON MISES, R. (1961) *Probability, Statistics and Truth*, (1928), MacMillan Londra, Secondeditio

v