Bayes' and Fisher's Conceptions of Statistic in the Context of Empirical Paradigm

Vladimir Reznikov, Nowosibirsk

The conceptions of Bayes and Neyman-Pearson are considered in methodological literature as irreconcilable opponents as for their goals, tasks, and methods of solving statistical problems [1-3]. This article demonstrates that in some respects important for the statistical practice the methodological principles of the conceptions of Bayes (BC) and Fisher (FC), the last is ideologically close to Neyman-Pearson, reveal the similarity in positions. As an alternative to these directions of investigations an empirical conception is proposed. Let us start with some critical arguments against Bayes and Fisher's conceptions.

Arguments against Bayes' conception are as follows.

1. The proponents of Bayes' conception are accused of subjectivism which becomes apparent in the assigning of a priori probabilities. What is especially criticized is the idea of using the uniform distribution in the case of the lack of knowledge about the described state of affairs.

2. The followers of the Bayesian paradigm maintain that the calculation of probabilities based on Bayes theorem amounts to the inductive logic. Their opponents argue that methods of induction along with Bayes theorem do not constitute logic.

3. The most important task of the inductive logic is to determine the degree of hypotheses confirmation. The solution of this problem proves to be sensitive to the choice of probabilistic measure. In certain situations every measure leads to paradoxical solutions. In addition, very often there are no conclusive proofs in favor of a measure for formal description of a concrete situation.

4. Hypotheses verification does not originate from this paradigm. It is an adaptation of the classical statistics.

5. The notion of confidence probability for the estimation of the distribution parameter, which is constant, is not correct because the constant falls on the interval with the probability either one or zero. All intermediate probabilities between zero and one which are used in Bayesian methodology do not make any sense.

6. The adequacy of the Bayesian paradigm for scientific applications is also criticized. Scientists are not Bayesians.

In its turn the arguments against Fisher's conception are as follows.

1. The inability of using a priori information came under criticism.

2. The verification of hypotheses is the most important task of Fisher's conception. The methodology of the hypotheses verification came under criticism.

3. The followers of the Bayesian paradigm point out that in the framework of Fisher's conception there is no unique confidence interval. Confidence probability is determined for the multitude of intervals. In particular, when solving the problem of parameters estimation, it is not known on which interval falls the required parameter.

4. The objective character of Fisher's paradigm is denied. In objectivist conceptions probabilities are assigned to the events that have not yet happened. For this reason the probability of event depends on the place and time this event happens.

5. The adequacy of the Fisherian conception for scientific applications is also criticized.

In spite of differences between the approaches (BC) and (FC) they are methodologically similar.

1. Both lines of inquiry are model oriented. In both approaches the problem of model building, which is crucial for special sciences, is ignored. (BC) as well as (FC) assume that a model is given and what is necessary is to specify its unknown parameters.

For example, the problem of the estimation of distributions parameters belongs to this class of tasks. From the very beginning, much more information is assumed known (model) than to be additionally obtained (model parameters).

In science, like in any other domain of rational thinking, behaving, and acting, it was always conventional to move from solving simple tasks to investigating complex problems. With this in mind, it seems unreasonable to build a statistical distribution with the mean and dispersion as parameters in the beginning, in order to determine the mean, dispersion, and other statistical characteristics.

2. In both approaches, basic features of models, like independence, are either considered given a priori or presupposed on the basis of intuitive considerations. It is assumed that if the investigations are conducted in controllable conditions and the experiments are independent, the results of the experiments will also be independent.

3. Solutions of many problems, like estimating the distribution parameters or building regressive models, lead to the same or similar quantitative results. Of course, the interpretation of the solutions in both cases will be different.

4. The parameters estimation criteria (consistency, unbiasedness and efficiency) are used in both approaches.

5. The used methods of parameters estimation are not robust because the estimations obtained much unbiased even in the case of a small deviation of statistical characteristics of data from model ones.

6. Each approach criticizes the methodological principles of the concurrent one. For instance, the groundlessness of the conception of hypotheses confirmation in Bayes' approach caused by the problem of measures multiplicity is criticized. At the same time, a satisfactory solution of this problem cannot be found in the framework of Fisher's approach as well. In its turn, the methodology of hypotheses verification in Fisher's analysis is criticized from the Bayesian point of view. The latter, however, does not propose any ways of improving the verification of hypotheses.

The empirical conception of statistics developed by Alimov is an alternative to BC and FC methodology. The main task of the empirical metrological conception (EMC) consists in constructing a semi-empirical model of data. The semi-empirical model of data is effective for prognosis, if the semi-empirical parameters of the model are robust. Theoretical characteristics are the representation of the stable semi-empirical characteristics. General basic characteristics of the data model, like uniformity and independence, are neither considered given a priori nor presupposed on the basis of intuitive considerations (for example, experiments are conducted in controllable conditions and independently from one another; therefore the model of independent observation is accepted). Instead, they are obtained on the basis of the formal examination of independence, uniformity, and other characteristics. If the formal independence is obtained, there are reasons for cautious claim that the experiments were truly conducted in independent conditions.

EMC criticizes Fisher's methodology of hypotheses verification. The point is that Fisher's methodology presupposes that the distribution of the criterion statistic is known up to small probabilities. The main principle of the hypotheses verification consists in the refutation of the hypothesis, even if it is true, once in the only conducted experiment a low-probable event happens.

This method is convincing neither for acceptance nor for refutation of the hypothesis. A convincing way of the hypothesis refutation consists in the discovering of an essential difference between the frequency of real event and its probability, provided that the hypothesis is true. This method is based on frequency and requires many experiments to be conducted.

In order to demonstrate that the appeal to small probabilities is not realistic, let us show that for the convincing estimation of the probability about 10**(-n) we need about 10**(n+1) experiments.

Suppose that we examine the frequency of an event $\varpi_m(A)$, where m is the quantity of the conducted experiments. Denote by s_k the number of test in which the event A happened k-th time. When passing from $m=s_k-1$ to $m=s_k$ the frequency changes from

$\varpi_{(sk-1)} = (k-1)/(s_k-1)$ to $\varpi_{sk} = k/s_k$.

The relative frequency leap in this transition is equal to

$\delta(k)= (\varpi_{sk}(A)- \varpi_{sk-1}(A))/ \varpi_{sk}(A)=(s_k-k)/(s_k-1)k$

For rare events s_k>>k and therefore $\delta(k)$ is approximately equal 1/k.

Therefore, only when k>10 the relative increase in frequency will be less than 10%.

A weakness of the contemporary estimation methods consists in that they do not provide stable estimations. The most powerful estimation method is the method of maximum likelihood (MMP). It starts from the assumption that the results obtained are independent and the maximum value is assigned to their combined density. Method MMP is exact, if data completely fit with its various requirements. In case of a small deviation of the data actual characteristics from the necessary model characteristics, the obtained estimations very often have large bias. As a rule, in the practice of using the method MMP the change in combined density value for different data is not examined in order to determine on which data the combined density will be actually maximal. How the results will change when the combined density of the obtained results is not maximal is not examined as well.

The applied criteria of the quality of estimation have no pragmatic value. The popular criterion of the quality, which is consistency, does not allow one to estimate how the estimation obtained from the end sample differs from the optimal estimation derived from the presupposition that there is an infinite sample of data.

The methodology of building confidence intervals in BC and FC is also criticized in the empirical conception. In BC one speaks of the probability of a constant's fall on an interval. In FC one finds a multiplicity of intervals, and it is not clear how to use them. The empirical conception suggests a constructive procedure for building a unique "confidence interval". The probability of the falling on the interval is not determined. To illustrate this approach let us consider the estimation of the average value. Data are divided on groups with the equal level. For each group the average value is calculated. Minimal and maximal average values are the edges of the actual " confidence interval". Then we make the obtained interval more accurate on the basis of new data with more level and more number of groups. If we find an interval with the length not exceeding the permissible error, we can speak of the stable estimation of the semi-empirical average value and of the building of the confidence interval for it.

To determine the values estimations that are essentially connected to the values whose estimations prove to be stable the methodological principles representing the development of the methodology of using mathematics in natural sciences (first suggested probably by Comte) are used. Comte argued: let a variable X take values x_1, x_2,… x_n, and these values can be measured. Another variable Y is not measurable. The role of mathematics consists in describing the dependency F between the variables X and Y. If the variable X is measured with permissible error, the solution of the equations set F(Y)=X allows one to determine the required values of the variable Y.

It is assumed in the statistical literature that the theorem of the law of large numbers has a special empirical significance. In conclusion the special epistemological and heuristic significance of the theorem of the law of large numbers which is often considered as a corner stone of statistics is criticized.

In the simplest case the theorem of the law of large numbers is stated as follows.

Theorem. Let μ be the number of occurrences of the event A in n independent tests, and p be the probability of the occurrence of A in each test. Then, whatever $\varepsilon > 0$,

$$\lim_{n\to\infty} P\{|\mu/n-p(A)| < \varepsilon\}=1 \quad (1)$$

The problem of the status of the theorem of the law of large numbers appears. Is it probability-theoretical? That is to say, does it allow one to calculate the probabilities of events on the basis of elementary probabilities or does this theorem allow one to reveal and verify primary probabilistic characteristics, for example, to determine the stability of frequencies, and thus belong to mathematical statistics?

In the theorem of the law of large numbers the theoretical value "the probability of success" is given a priori, and thus the theorem cannot be used in order to determine this already known probability. The conclusion of the theorem speaks of the event's probability, which consists in the difference between the probability and frequency of the event A. The event in the applied probabilistic theory is a result of any experiment. The probability of event is a theoretical value. That is why the probability of event cannot be a result of a real test.

In the empirical interpretation the probability of event is determined if and only if the corresponding frequency is stable. In the empirical approach the external probability in the expression (1) should be replaced by frequency. The conclusion then takes the following form:

$\omega_n\{|\mu/n\text{-}p(A)| < \varepsilon\}=1$ (2).

In the expression (2) the symbol ω denotes frequency and the symbol n designates a large number.

Furthermore, the conclusion of the theorem rewritten with the help of the expression (2) speaks of the frequency of the deviation of the absolute difference between the probability of the event A and the frequency of the same event, provided this difference is less than ε. The heuristic significance of the theorem conclusion for determining the stability of the frequency in its empirical interpretation is in doubt. Indeed, if the probability of the event A is close to zero, the conclusion of the theorem speaks of the frequency of the frequency of the event A. Supposing that the theorem has a heuristic significance, it should consist in assuming to be possible to determine the stability of the frequency of the event A from the frequency of the event A.

It is obvious that the methodological significance of this interpretation of the theorem is overestimated. First, to determine the frequency of the deviation of the frequency of the event A from the probability is not possible without determining before the frequency of the event A. Second, in order to use the frequency of the frequency of event for determining the stability of the event, it is necessary before to make certain of the stability of the frequency from the frequency of event. To push the analogy further, in order to make certain of the stability of the frequency from the frequency of the event A it is necessary to use more complex construction. Namely, one needs the frequency of the frequency of the event B, which is in its turn the frequency of the event A. There is a logical circle here.

References

Giere, R.N. 1997 "Scientific Inference: Two Points of View", *Philosophy of Science* 64 (Proceedings), 180-184.

Howson, C. 1997 "Error Probabilities in Error", *Philosophy of Science* 64 (Proceedings), 185-194.

Mayo D.G. 1997 Error Statistics and Learning From Error: Making a Virtue of Neccesary. *Philosophy of Science* 64 (Proceedings), pp.195-211.