EVOLUTION OF THE TEACHING OF STATISTICS AND PROBABILITY IN FRANCE AT SECONDARY SCHOOL LEVEL

Bernard Parzysz, IUFM de Lorraine/Université de Metz, France

It is only about 1965 that statistics and probability appeared in French high school curricula. The aim was to give students statistical tools for their studies, and probability seemed a necessary prerequisite for acquiring these tools, giving them technical means to study the normal distribution (presented as a limit of the binomial distribution). Since then, four periods can be distinguished in the teaching of this particular domain.

FIRST PERIOD: 1970-1981.

In 1970 the situation changes, with the so-called “modern maths period”. Only the descriptive part of statistics remains, probability dominates. This is an example of didactical transposition, the reason put forward being the necessity of following the evolution of scientific knowledge. There are also other reasons for teaching it: its increasing place in scientific research and its usefulness in everyday life. But, even more than that, students will be able to apply their knowledge about sets, algebraic structures and relations between sets to probability, as well as combinatory formulas (which appear as a prerequisite for probability). As is the case for geometry at that time, probability appears as an axiomatic theory, the properties of probability sets being deduced from a small number of axioms of Kolmogorovian style. This approach can be called “Laplacian”, since Pierre-Simon de Laplace (1749-1827) indicated that, having ensured that all the possible outcomes have the same chance of happening, one has the well-known formula: probability of an event = (number of favourable outcomes) ÷ (number of possible outcomes). For statistics as well, their main usefulness is to enable students to use their knowledge of sets. The aim was to show that “modern maths” can be applied to many domains and, thus, legitimate the generalised teaching of set theory.

The link between probability and statistics is weak, since the main purpose is to study a theory. At the most, you can find a remark about frequency “appearing as an approach of the notion of probability”, or about the formal analogy between probabilistic and statistical parameters, but no more. In this approach probability has its source in reality, but maths now provides a “clean” theory, cleared of its statistical impurities.

SECOND PERIOD: 1981-1986.

The weak link between probability and statistics is now broken: statistics is taught in fifth and lower sixth (ages 15-17), since probability is taught only in upper sixth (age 17-18), with the exception of double statistics which is taught in upper sixth (but only to non scientific students). Apart from this, the guidelines have changed little. But—and this is general for maths—a drift towards formal teaching, often senseless for students, had been noted during the “modern maths” period. So the recommendations proscribe any formalisation, especially with non-scientific students. So in statistics, students will not have to collect and process statistical data themselves; teachers are advised to give them ready-to-use data, “to avoid a tiresome accumulation of experiments”. Didactically, there is a real danger: it seems sensible not to have students spending too much time on various statistical enquiries, but the risk is great for teachers—who are generally pushed for time-to skip this step completely, and have their students working only with “clean” data.

Students need to undertake a statistical process from beginning to end, because:

- with ready-to-use data, you have a “black box effect” like the “random” key (or the “sin” key, or the “ln” key...) of a pocket calculator. What do the numbers really mean ? If they come from measures, how were these measures obtained ? etc., all questions which may be important for answering your problems sensibly. But when you get the data yourself, you can answer such questions easily.
- in a table, data have often been gathered in classes, and the question of choosing “suitable” classes, and a “suitable” number of classes is already solved. However, this is an important point of a statistical process, and it would be detrimental if students were not made aware of this problem by having “experienced” it.

The same concern with the meaning of what is taught led the authors of the curriculum to advise that the study of the theory of probability (which is presented in the same way as before) can be prepared through “using data coming from games, pseudo-random sequences given by a pocket computer, or series of experimental measures or observations”.

The contents are more differentiated than previously. Until 1986, the curriculum for non scientific sections was just a subset of the “scientific” curriculum ; but from then on:

- the “literary” curriculum is oriented towards historical aspects, including the study of historical problems, for instance, the 1654 letters between Pascal and Fermat;
- the “economical” curriculum is oriented towards economical and social science problems, being intended to prepare the students to study that subject at university.

THIRD PERIOD :1986-1990.

From 1986 statistics is taught through junior and senior high school. The general guideline for maths is “to link observations of reality to representations: sketches, tables, diagrams”, “to link these representations to mathematical activity and to concepts”, and the way of reaching such goals is “to build mathematics on problems met in several disciplines, and, in return, to use mathematical knowledge in various subjects”. So the process of modelisation is now part of the teaching of maths, which no longer consists in just studying models. It seems obvious that statistics fits well with this project, since:

- you can start from a situation observed in any discipline (geography, biology etc.), about which you want to answer some questions
- then you collect numerical data about this situation, in order to answer the questions
- you process these data in suitable ways (“representations”), leading to define some useful concepts: variable, function, frequency, mean...
- you try to find answers to the initial questions, and finally confront them with the initial situation.

One would expect this principle to apply to probability as well, but it is not the case. There is still no link between probability and statistics, and the approach is still of the Laplacian type. And combinatorics remains a prerequisite for the study of probability.

FOURTH PERIOD: 1990-....

Studying statistics through junior high school has normally led the students to “a certain familiarity with randomness”. Thus the fifth form offers an opportunity to synthesise simple statistics, “which constitutes an important element for the formation of all the students” (the fifth form is not differentiated in “literature”, “economics” and “science” sections).

The use of “rough” data is now recommended, contrary to 1981: “statistics do not come ready-made; they proceed from successive reasoned choices”; teachers are advised to have their students undertake a statistic study from its beginning to its end “at least once”. This is a consequence of the fact that scientific calculators are now widely spread among students, since they are available at low cost. Moreover, the sense of such work is put to the fore: “at each step of the treatment (...), signification is gained at the expense of losing part of the information”.

(Let us remark that collecting data is not modified by that technological change; what is new is how easy it is to process them, saving a time that can be used precisely to collect real data, in sufficient number.)

The major change is that the new curriculum poses the question of modelisation in probability. In opposition to what was the case before, the theoretical probabilistic model is no longer the starting point, but it is constructed as an answer to some questions, in order to solve a certain (wide) category of problems. Students now have -with the help of the teacher- to build tools enabling them to study “real” situations, in which statistical data are collected, organised and then processed. They no longer have to apply general tools, which are defined *a priori*, to varied situations. This new point of view on the teaching of probability implies the taking into account of an essential fact: the domain within which the model can be “reasonably” used. Here is a simple example: let us suppose that we are interested in the link between outside temperature and the daily consumption of a fuel heater. We begin with collecting data, in order to obtain a kind of curve giving the consumption as a function of the temperature. If we use it to estimate the consumption for an outside temperature of +5°C, we shall get a “reasonable” answer. But if we want to use it for a temperature of +25°, we shall surely find that the heater ...produces fuel ! (quite unfortunately, this does not correspond to reality).

Which way to probability ? If students and teachers have to construct a theoretical model of probability together, on what basis will that be ? Two major ways are possible:

The first is to start from the frequency of a population character, and then identify this frequency with its “theoretical value” (probability). This approach is recommended for the “economic” section: “the starting point can be the census of a population”. However, in this case, the distinction between frequency and probability is far from obvious. Making the shift from statistics to probability explicit—which is necessary to make students aware of a change from reality to a theoretical model—requires the determination of a random experience: extracting an element from the population at random. But this part of the modelising process is generally omitted so there is a real danger of reinforcing a “cardinalist” conception of probability among students (i.e., a probability is necessarily a ratio between two numbers of elements). An effect of such a conception can be observed when, to solve a probabilist problem (especially those using conditional probability), some students shift from theory to an “ideal reality”. So if the probability of a given event E is 0.148, they will suppose a population of 1000 individuals, and consider the subpopulation of the individuals realising E to be 148 individuals.

Another possibility is to ground the access to the notion of probability in a repeated random experience, observing that the frequencies of the appearance of an event get closer and closer to one another as the number of experiences increases (a computer-assisted simulation can help here). This type of approach is recommended for all three sections (literary, economic and scientific): “the introduction of the notion of probability will be based on the study of statistical series, obtained from the repetition of a random experiment, by highlighting the properties of frequencies and the relative stability of the frequency of a given event when this experiment is repeated a great number of times”. This can be called a “Bernoullian” approach, since Jacques Bernoulli (1654-1705) was the first to give a version of the so-called “Bernoulli’s law” which is alluded to in this quotation. Let us recall here a simple form of this law:

Being given a sequence (X_n) of independent random variables following the same law (let m be the mean of this law), the random variable $Yn = \frac{X_1 + X_2 + \ldots + X_n}{n}$ converges in probability towards m.

This new approach requires answering an important question: probability appears now as a limit of frequencies, but in what sense? Obviously, students cannot be given the above definition, since they do not know anything about the theory of probability. Moreover, this type of limit is quite different from the one they are used to in calculus. In a concrete way, you may start with a simple random experiment, let us say the tossing of a coin, and either realise or simulate it (but beware of the “black box effect”). Then you observe the frequency of heads (f_n) that you obtain in a sequence of n tosses, for n = 100, 200, 500, 1000, etc. You can see that, as the number n increases, the variation of f_n becomes less and less “irregular”. But, nevertheless, there is no logical necessity for (1) f_n having a limit value, and, even if you admit the existence of such a value, (2) this value being equal to 1/2. This becomes clearer when you toss a drawing pin: you still have two possible results, but symmetry cannot help you find the value of a possible limit.

This shows clearly the shift from reality to model. If, you must ascribe a probability to the outcome “the drawing pin falls on its head”, a “natural” way is:

1. to admit that you can do it: the existence of a limit can be guessed from various types of random experiments undertaken by the students, such as the ones mentioned above;
2. to give it a numerical value, and this value—although you can get approximations of it—will *not* be given by the experiment (even if it is repeated a very large number of times): different series of 1000 tosses will not give the same result, even if they are not far from each other. For instance, if you find 0, 347, 0, 354 and 0,331 for 3 series, which value will you take ? the “middle” one, their arithmetical mean, or ... ? You have to *decide* it on the basis of your experimental series. Having done that, you have already begun to construct a model you can use to solve problems.

It is almost the same with the tossing of a coin: if you obtain 511 heads on 1000 tosses, you can *admit* that the coin is “well balanced”, and then construct a model in which:

- the coin never falls on its edge
- “getting a head” and “getting a tail” have the same probability 0.5.

A consequence of this approach is that Laplace’s formula appears only as a particular case, and that, subsequently, combinatorics (often an obstacle to the learning of probability) has now become a mere tool for calculating probabilities (instead of being a compulsory way for the access to the theory of probability). Another consequence is that statistics has now become the founding stone on which the notion of probability is built.

CONCLUSION

As can be seen from this short study, the teaching of probability in France from 1970 on has been based on two successive points of view:

- first a Laplacian approach, from which a subjective notion of probability, determined by non experimental considerations, derives

- then a Bernoullian (or frequentist) approach, from which an objective notion of probability, as a limit of frequencies observed experimentally, derives.

Hence, a good frequentation of statistics has now become necessary. This is what the official curricula recommend; but it supposes quite a change in the high school teachers' state of mind, since many of them consider statistics useless, or not belonging to mathematics, and as a consequence do not teach them, or do it only if they can "spare time". Much has to be done with preservice teachers, to make them aware that:

- being able to interpret "soundly" the statistical data given by the media (and to criticise what the media say about them) is part of the formation of citizens, who are more and more confronted with information of all sorts
- probability is a theory grounded on phenomena belonging to the real world, which makes use of mathematical results and enables to solve a great variety of problems.

This is our challenge for the coming years.

REFERENCES

Bernoulli, J. (1713). *Ars conjectandi*. Traduction de N. Meusnier. Ed. IREM de Rouen (1987)

Henry, M. (1994). *L'enseignement du calcul des probabilités dans le second degré: perspectives historiques, épistémologiques et didactiques*, in Repères-IREM 14.

Laplace, P-S de (1814). *Essai philosophique sur les probabilités.* Réédité par Christian Bourgois, Paris (1986).

Parzysz, B. (1997). *Les probabilités et la statistique dans le secondaire, d'hier à aujourd'hui*, in Enseigner les probabilités au lycée. Ed. IREM de Reims.

Pichard, J-F. (1997). *Les probabilités, des origines au tournant du XVIIIe siècle*, in Enseigner les probabilités au lycée. Ed. IREM de Reims.

Renyi, A. (1966). *Calcul des probabilités*. Ed. Dunod, Paris.