From Frequency to Probability. Some Questions posed by the new French senior high school Curricula

Bernard PARZYSZ
IUFM Orléans-Tours & Équipe DIDIREM (université Paris-7)
72, Faubourg de Bourgogne
45000 Orléans (France)
bernard.parzysz@orleans-tours.iufm.fr

The domain of randomness made a late entry in the teaching of mathematics, even if the idea of probability appeared as early as 1654, in the famous correspondence between Pascal and Fermat. Moreover, even if the relation between frequencies and such a notion had already been sensed, it is only with Jakob Bernoulli (1713) that the link between a probability and the frequency of an event in the repetition of random experiences was clearly established by the law of large numbers. The viewpoint on the articulation between the two domains has evolved a lot over the last two centuries. To some extent, their teaching at secondary level mirrors this historical genesis since, to take the example of France :

- on the one hand they appeared only in the late '60 in senior high school curricula;
- on the other hand the links established between probability and statistics have fluctuated a great deal in these curricula over the last thirty years.

This reveals underlying epistemological and didactical difficulties that I will try to emphasize through an analysis of the new French curricula (Parzysz 1997).

1. Overview of the evolution of the teaching of statistics and probability in France

The real appearance of probability in senior high school curricula coincides with the beginning of the so called 'modern math' period. The times were indeed favourable, since a greater importance was given to set theory and axiomatic approaches; an introduction to probability theory derived from Kolmogorov's set of axioms could easily be integrated into such a scheme, and it made it possible to show the students that mathematics could be applied efficiently to various domains of human activity. The probabilistic viewpoint chosen was of the 'Laplacian' type, i.e. it was based on Laplace's first principle, which postulates an equal probability for each of the elementary events and defines the probability of an event A by the formula

$$P(A) = \frac{\text{number of cases producing A}}{\text{number of possible cases}}.$$

Such a reduction to equal probability, natural in analysing games of chance, was quite in keeping with the set theory underlying every domain of mathematics teaching, and the calculation of a probability merely amounted to using combinatory techniques. Of course, even if there was a chapter dealing with statistical data, its link with probability was virtually non-existent, since the aim was precisely to emphasize a theoretical object *before* applying it as a tool for studying statistical populations.

The backward surge from this kind of approach, which took place at the beginning of the '80, brought about a shrink back of the teaching of probability, while organizing and managing statistical data appeared at every level of junior high school (from 1986 on). In 1991 the two domains were brought together at last, since a 'frequentist' approach of probability was clearly recommended (the emblematic experience being the repeated tossing up of a drawing pin). Finally, since 2000, the new statistics curriculum is grounded on the idea of observing the fluctuation of samples by simulating the repetition of a random experiment and observing the stabilization of the frequency distribution of the possible outcomes, the notion of probability being introduced later as a 'theoretical frequency'.

So, two different approaches of the concept of probability follow each other in the teaching of randomness in France:

- a 'Laplacian' approach, taking only into account spaces in which the elementary events have the same probability of appearance *a priori*, probabilities being calculated by counting.

- a 'frequentist' approach, in which probability is obtained *a posteriori* as the 'limit' of the frequency of the occurrences of an event linked with a random experiment which is repeated many times.

These two approaches appear in fact as complementary and contradictory:

- *contradictory*, because the first one is situated at first in a mathematical model, whereas the second one has a concrete substratum for a starting point;

-*complementary*, because Bernoulli's theorem (i.e. the so-called 'law of large numbers') ensures that there is a connection between the two viewpoints by guaranteeing that -within the theory- the frequencies converge towards the probability.

2. Towards a dialectic process

From a didactical viewpoint, a Laplacian approach presents both advantages and drawbacks. Setting itself in an ideal context of equal probability, it avoids the interpretation of 'reality' into a theoretical model, but by so doing it does not establish any link with the real observation of phenomena, and is likely to appear as an unwarranted and sterile game to the students. Moreover, it implies a necessity for finding equal probabilities "somewhere".

This constraint makes it necessary to consider only random experiments for which an equal probability can reasonably be postulated, e.g. games of chance or balls drawn out of a box. So, the model is given at the beginning and not established from experience. But such an approach may reinforce the 'uniformist' conception which is deeply rooted in most of the students, according to which, in random situations, when several eventualities are possible in the most complete uncertainty, they consider all of them as having the same probability (Pratt 2000). Moreover, the theory cannot give an answer by itself when several alternative models are proposed.

So, a frequentist approach seems necessary for conceiving a teaching of randomness which, on the one hand, will give all its meaning to the concept of probability (without being limited to equal probabilities) and, on the other hand, will take the students' (mis)conceptions into account (by relying on a concrete substratum, allowing a confrontation of these conceptions with experience). But it requires one to carry out a sufficiently great number of experiments within a limited amount of time, hence the compulsory use of computer-aided simulations. Recently, Saenz Castro has conceived and implemented a learning curriculum based on this kind of approach, in which the units are centred on a cognitive conflict linked with the students' initial conceptions (Saenz Castro 1998), and Pratt used the simulation of the repeated tossing of a loaded die (a same number is written on several sides), in order to thwart the uniformist conception and make the law of large numbers appear as a tool for solving problems (Pratt, to be published).

The researches mentioned above have another point in common: they are all restricted to situations based on equal probabilities (even if it is sometimes hidden, as in the case of the loaded dice). This means that they may make it possible to discover the probabilities (= 'simple' rational numbers) assigned by the teacher to each elementary event when introducing the model managing the situation into the machine. However, to follow the idea through to its conclusion, it would then later be necessary to consider a random experiment (of the 'drawing pin' type) for which no probability could be assigned *a priori* to the events in which one is interested. This is precisely what Ventsel advocated 30 years ago :

'If practice shows that, when the number of experiments increases, the frequency of the event has a tendency to stabilize, rigorously approaching a constant number, it is natural to assume that this number is the probability of the event. It is clear that this hypothesis can only be verified for

events whose probabilities are likely to be calculated directly. (…) It is quite natural to assume that, for events which cannot be reduced to a system of cases, the same law remains true…'

(Ventsel 1973, p. 24)

One may see in this statement the outlines of a modelizing process in three steps, aiming to validate among students the idea that, for any event A linked with a given random experiment, *there exists* a theoretical value P(A) towards which the frequency f_n(A) of the occurrences of A converges (in a sense to be defined) when the experiment is repeated n times 'in the same conditions':

First step : this convergence is empirically observed when considering only events A for which a probability can be assigned *a priori*. In practice, such an event will be derived from a random experiment, to which a space provided with a uniform law can be canonically associated.

Second step : then, for events A to which one cannot assign an *a priori* probability (i.e. related with an experiment to which no 'uniform' space can be associated), the repetition of the experiment enables one to notice the 'stabilization of the frequencies'.

Third step : by basing oneself on the analogy between the first two steps , one decides to 'take the plunge' and to assume that one can, in this case as well, associate a 'theoretical frequency' P(A) with the event A, for which the observed frequency f_n(A) (where n is the greatest number of repetitions available) is an approximate value. There still remains to determine the nature and the quality of this approximation.

3. Some didactical consequences

Of course, this process will pose some didactical problems. It necessarily requires a previous familiarization with the domain (at junior high school level), in order to bring the students to be able to identify a random situation, to describe it, to problematize it and to keep only the relevant elements (Henry 1999). Moreover, this approach requires *effective* repetitions of the same random experiment; one can see at once that limitations related with the constraints of time immediately arise. This is why, in classrooms, one is compelled to make do with computer simulations, which poses the question of the students accepting such a substitution. This stage must not be considered as self-evident; it too has to be subjected to a didactical process.

Indeed, the computer may well appear as a 'black box' to the students, having no apparent relation with a concrete random situation; therefore, it seems necessary to put up the idea of this black box being an acceptable substitute, and this requires :

- a confrontation between several random situations and their simulations (with calculators and/or computers), in order to validate the random generator;
- endeavouring, when carrying out a simulation, to keep as close as possible to the concrete process.

Anyway, some biases are unavoidable, such as the fact that the 'random' generator is not based on randomness, or the fact that the simulation process implies putting into the machine the probabilities which will be approached later with the frequencies observed. One may begin with the study of very simple random experiments, for which the students are able to provide an unambiguous theoretical model; for instance, a box containing balls of two colours seems a fundamental support for this purpose (Coutiñho 2000). And, when possible, they can be asked to find by themselves a way to carry out the process on their machines. The simulation will enable them to check empirically the law of large numbers, namely: the larger the number of repetitions of the experiment, the closer one gets to the model. This method will also be useful to choose from several alternative models.

Another delicate moment is that of the conceptual shift necessary to accept the existence of a theoretical limit for frequency; actually, contrary to the previous examples, in this case one does not have an *a priori* value of this limit, which would in some way ascertain its existence. In an objectivist conception, the point is to accept that such a limit exists, even if one is not able to exhibit it precisely (which is a common situation in physics anyway). It is only with studying various 'canonical' examples that this idea will be accepted by the students.

The notion of ‘theoretical frequency’ allows one to transfer, in a ‘natural’ way, the properties of frequency distributions to probability distributions. (Henry 1994). It will be necessary to emphasize clearly the fact that one is now working within a theory, and that -as in geometry- the problems will be solved only with the help of definitions and theorems belonging to this theory. But this construction of the theoretical domain will only make sense if it leads to show its efficiency for solving concrete problems, i.e. by using models in order to answer ‘real’ questions (as for instance in inferential statistics).

It is this kind of process which has apparently started up in the French senior high school curricula implemented since 2000, but nevertheless some points would need to be improved through a revision. In particular, only situations based on equal chances are recommended, which notably restrains the scope of the concept of probability and leaves the students halfway through their construction of this concept. Moreover, even if the distinction between reality and theoretical model is established without ambiguity, modelization itself does not clearly appear as an aim of the teaching. To end with, through lack of time, the process does not reach the use of theoretical models for solving concrete problems.

It is to be hoped that some research will now take in interest in the problems raised above, and that the junior high school curricula, which are currently in hand, introduce early enough -as is the case with some of our neighbours- a substantial acquaintance with randomness, in order to familiarize students very early with this domain and to enable them to undertake and bring to completion the processes of probabilistic modelization at senior high school in good conditions. Moreover, this would enable them to understand the status of maths as a tool for modelization, and the importance of turning a theoretical look on problems taken from reality.

REFERENCES

COUTIÑHO, C. (2000). Introduction aux situations aléatoires dès le collège : de la modélisation à la simulation d’expériences de Bernoulli dans l’environnement informatique Cabri-Géomètre 2. Thèse de doctorat. Université Joeph-Fourier, Grenoble.

HENRY, M. (1994). L’enseignement du calcul des probabilités dans le second degré, in Repères-IREM 14, pp. 69-104.

HENRY, M. (1999). L’introduction des probabilités au lycée : un processus de modélisation comparable à celui de la géométrie, in Repères-IREM 36, pp. 15-34.

PARZYSZ, B. (1997). L’enseignement de la statistique et des probabilités dans l'enseignement secondaire, d'hier à aujourd'hui, in Enseigner les probabilités au lycée, pp. 17-38. Ed. Commission inter-IREM Statistique et Probabilités.

PRATT, D. (2000). Making sense of the total of two dice, in Journal for Research in Mathematics Education 31, pp. 602-625.

PRATT, D. (to be published). The emergence of probabilistic knowledge, in Proceedings of the 3rd Conference of European Researchers in Mathematics Education (Bellaria, Italy).

SAENZ CASTRO, C. (1998). Teaching probability for conceptual change, in Educational Studies in Mathematics 35, pp. 233-254.

VENTSEL, H. (1973). Théorie des probabilités. Ed. Mir, Moscow.

RESUME

L’enseignement des probabilités et de la statistique en France, qui ne remonte guère à plus de 30 ans dans l’enseignement secondaire, a préconisé deux approches très différentes du concept de probabilité : ‘laplacienne’ puis ‘fréquentiste’. Cette dualité pose en particulier le problème du statut de la simulation informatique d’expériences aléatoires, et, au-delà, celle de la prise en compte de la modélisation dans l’enseignement des mathématiques, questions qui débouchent sur certaines questions épistémologiques et didactiques susceptibles de susciter de nouvelles recherches.