The Role of Statistical Inference in Teaching and Achievement of Students

Ramesh Kapadia, Alps Adria University Klagenfurt, Austria
e-mail: ramesh.kapadia@virgin.net

Abstract

The methods of teaching statistical inference vary and too often, insufficient links are made to the achievement of students. Many of the underlying ideas are counter-intuitive, as illustrated by the well-known examples of Kahneman and Tversky, and Gigerenzer. Recent testing of able students will be discussed. Simulations can also play a part, as demonstrated with an example from real-life of breathalysers. The ideas of decision-making are not particularly deep and complex, but the application of ideas in practice proves to be difficult. For example it is not simple to use ideas of utility in conjunction with probability, where the differences are quite subtle and need practice to be well understood. Students find it hard to apply the expected utility theorem and ideas of the probability premium in insurance. They are also perplexed by Arrow's theorem showing that simple and intuitive conditions for voting lead to contradictions, undermining the idea of a fair ballot. We will make reference to students' reactions in recently presented courses to undergraduate students at Klagenfurt University.

Key words. Bayesian statistics, statistical inference, simulations, pedagogy, decision theory, social choice.

1. Introduction

The objectives of this paper are to discuss the sources of difficulties in comprehending statistical inference as supported by various research studies, to assess the validity of the research through empirical studies with undergraduate students, and to draw conclusions linking theoretical analysis and experimental results. Our major hypotheses about the main factors influencing the difficulty of statistical concepts in solving problems of statistical inference are that:

1. Probability is a counter-intuitive idea especially with regards to private conceptions, as encouraged in subjective approaches, in contrast to equal likelihood and frequentist approaches.
2. Language is a major factor when students solve probability problems.
3. Context is a further major and essential factor.
4. The data format might have a huge impact on "success" in probability problems.

2. Conditional probability and language

We start with some core ideas and difficulties about conditional probability and inference. These ideas are important in teaching probability and too often forgotten. It should be remembered that the teaching of probability is relatively new in the school curriculum: in most countries it was only introduced towards the end of the last century, about twenty years ago. While there has been discussion of key ideas, not least through the International Conference on Teaching Statistics, there has been insufficient analysis of the underlying pedagogy. There has also been ambivalence with regards the role of probability and its links to statistics over the last few decades. Probability is too often been seen as another area of mathematics, with its 'simple' rules, while statistics is very strongly promoted in terms of its relevance to the real world. This approach is changing with the increasing emphasis on risk and there is a greater awareness of the importance of teaching probability, as espoused in the latest CERME 7 (see Nikiforidou & Page, 2011).

There are important links between probability and everyday language. The main philosophical positions on the meaning of probability *must* underpin the pedagogy of probability, yet is too easily or quickly neglected in most teaching approaches. Moreover, some linguistic issues have arisen in our own research, not least because of the international, multi-linguistic approach adopted. To foster international co-operation, we have used the same test in versions originally devised in Spanish by P. Huerta and later translated into German and English by M. Borovcnik.
There are many words in normal language which are used relating to probability and risk. This is true for all areas of mathematics, including number, algebra, and geometry. Primary teachers are fully aware of this and there are many research papers published on how the precision of mathematics is often out of line with the vagueness of language. As one famous example, a phrase from the Bible is 'to go forth and multiply'. The term 'multiply is used to mean 'increase', but multiplication does not inevitably lead to an increase in secondary school mathematics, once fractions or decimals have been introduced. We note that there are relatively few research papers on such issues arising out of the use of ordinary words in probability. Moreover, we note that unlike a concept like 'distance' which can be directly linked to its everyday use, probability requires the use of a model (Borovcnik & Kapadia, 2011).
We follow the approach of Oakes (1986) and Barnett (1973) with regards to the philosophical approaches to probability. A variety of locutions is used in everyday speech to express uncertain propositions such as:

- The next throw of dice is unlikely to be six.
- A farm labourer would probably vote for a left-wing party.
- Smoking reduces the chances of being healthy.
- The Chinese will surely become the largest economy later this century.

These propositions use probabilistic words to describe (in a loose sense) outcomes in some kind of chance set-up. They are the way that most children and adults think about chance events; note that aspects such as hope, belief and emotion are involved, which are seen as having no place in a mathematical approach. Probabilistic terms can also be used to reflect some (personal) measure of belief in a proposition between certainty and falsehood. Shafer (1976) has named these two approaches as aleatory and epistemic probabilities. Some authors have maintained this distinction (Carnap, 1950); others maintain that all statements are fundamentally aleatory (Reichenbach, 1949); some assert that all statements are fundamentally epistemic (de Finetti, 1974).
The majority assert that only aleatory probabilities are 'true' probabilities and that epistemic uses should be avoided (von Mises, 1928), thus ignoring the use of probabilistic words in everyday language and rejecting subjectivist probability. It is a relatively straightforward deduction to derive Bayes theorem; controversy arises over the situations to which it applies. In particular it is controversial when applied to test a hypothesis H, with given data D, i.e. $P(H \mid D)$, which is termed inverse probability – the focus of heated debate between schools of probability for over 100 years. The application of the theorem is not easy, especially in decision theory.
The axioms of probability are not in dispute. The contention is over the meaning and scope of probability statements. There are five main positions taken by the various schools of probability: classical; relative frequency; fiducial; logical; personal.
The original approach is classical, derived from games of chance, linked to equal possibilities, based on the postulate of insufficient reason. In 1774, Laplace used this principle to calculate the probability that the sun will rise tomorrow, given that it had risen daily for at least 5000 years. However the principle of insufficient reason is rather flaky. It also leads to paradoxes. If ignorance of a variable leads to a rectangular distribution, a non-linear transformation of the variable does not lead to a rectangular distribution, but the principle of insufficient reason should apply equally to the transformed variable. This leads to a paradox or contradiction of the principle directly. This approach is no longer taken seriously by the various competing schools of probability; yet it is still taught in secondary schools across the world. This could be one reason

that many pupils accept the classical approach from their teachers overtly but have private doubts which are never fully resolved.

The approach via relative frequencies was developed in detail by von Mises (1928), whereby a probability is the limiting value of a relative frequencies. He notes that "just as the subject matter of geometry is the study of space phenomena, so probability theory deals with mass phenomena and repetitive events" (p. vii). These aggregates (such as a sequence of coin tosses) are seen as the proper domain of probability theory and are called 'collectives'; von Mises denies that probabilities can be assigned to other propositions such as what will happen to the Chinese economy in the future.

Thus von Mises places probability firmly in the aleatory tradition. This clearly restricts the use of the word probability in usual discourse, with a very particular and rather narrow meaning. This is similar, for example, to the mathematical use of the word function, and its everyday use. Hence probability applies only to collectives which are aggregates with two empirical properties. First 'the relative frequencies of certain attributes become more and more stable, as the number of observations is increased' (p.12), i.e., there is a finite limit as the number of trials increases to infinity, and the limit is defined to be the probability of the selected attribute or event *A*.

A second more subtle requirement is that the elements possessing attribute *A* should be randomly distributed in the long run of observations. Whilst the first assumption is often discussed in the school situation, the second requirement is rarely introduced but again is equally important and can lead to 'private doubts' amongst pupils which are left unresolved. Hence, the probability is a property of a collective and not of an element in the collective. However, the approach cannot have a predictive value and it makes no sense to talk of the probability of the 4th toss of a coin being a head. Even the probability that in 50 tosses there will be 20 tails is undefined. There is a further problem in terms of what constitutes an appropriate collective, such as when an insurance company has to decide if a man will die in the next year: should they look at death rates of men, of men and women, of men of a certain age, etc? In practice insurance companies are much more sophisticated than merely looking at all men and women of all ages as a single collective; otherwise they would go bust.

Fisher developed fiducial probability. He tried to reconcile aleatory and epistemic approaches by fiducial probability, which he never really defined properly. He discussed statements that particular unknown parameters lie within specified limits.

For logical probability, probability is the logic of partial entailment, of which the traditional logical calculus is a special case. Thus probability is a logical relation between propositions and hence strictly normative. Jeffreys developed the principle of indifference in conjunction with the simplicity postulate but most modern Bayesians now prefer to link probability to personal uncertainty, the final approach.

Amongst others, Savage gives an account of subjectivist probability, which "refers to the opinion of a person as reflected by his real or potential behaviour. This person is idealized; unlike you or me he never makes mistakes [… nor ...] makes such a combination of bets that he is sure to lose no matter what ever happens… the probability is basically a probability measure in the usual sense of modern mathematics. [...] The extra-mathematical thing, the thing of crucial importance is that the probability is entirely determined in a certain way by potential behaviour […] such that it is the odds that thou wouldst barely be willing to offer for *A* against not *A*" (Savage et al., 1962, pp. 11).

The Bayesian approach is prescriptive with regards to the revision of probability in the light of new information. This is essential for a coherent approach. Coherence is seen as the key axiom of subjectivist probability; in English language coherence has a positive connotation, which seems to support subjectivist approaches. The major schools of probability also differ in their approach to inference (Oakes 1986, pp. 110-148). From a pedagogical point of view, the approaches of the schools of probability described above lead to some specific aspects which should be borne in mind in planning teaching.

1. First, the history of the subject shows that probability has to begin with intuitive notions of equal likelihood despite the apparent paradoxes and constraints.
2. Second, a subjective degree of belief also needs to be explored, particularly where personal hopes and wishes may be significant. There are many private conceptions of pupils which should be explored – as does happen in primary school, for example with the notion of conservation, as espoused by Piaget.
3. Finally, the idea of long-run frequency and stability, as well the notion of randomness need to be introduced. Whilst such an approach may well be part of an overall curriculum, it is important to remember that it should not be rushed; otherwise the ideas will be developed on 'the shifting sand' of 'pupils' misconceptions'. Indeed some blame the recent financial crisis partly on a poor understanding of probability and risk amongst those in charge.

There are also continuing studies following Tversky (2011) whose collaboration with Kahneman led to ground-breaking work on heuristics and biases, as well as the Nobel prize (in economics). Yet probability does play a crucial role in everyday life so it is essential that school education addresses the underlying issues coherently.

Our research has been on conditional probability, even though our test does not mention the word in order to avoid complications which may arise from use of the word in test questions. Instead we focus on numerical and algebraic questions relating to proportion, which is used as a proxy for probability – following the classical or frequentist position adopted in schools. This is because all pupils are introduced to the notion of proportion at school from a relatively early age and meet it many times in the mathematics curriculum. It is much later that probability is introduced and sometimes probability is not studied extensively in school.

Our test was originally devised for use in Spain (Huerta & Cerdán, 2010). Collaborative work led to its translation into German and English. It has not always been easy to translate ideas and this has led to some discussion between the authors. As an example, we discuss modo subjuntivo. We found that from the point of view of the Romanic languages modo subjuntivo is more natural and easier. Experts note that this mode is the more natural way for expressing the question in Spanish and that, probably, this will be the same for all the Romanic languages. However, for English and German, it seems to be less accepted and less well understood. Thus the translation from Spanish to English and German was into indicative form.

Another important element when considering conditional probability is that even a change in order of words can have very important changes in the probability being calculated. Also much care needs to be taken over the use of the two words of 'and' and 'or'. The two words are given very precise meanings in mathematics and probability, which is far from what happens in everyday language. This was another reason for having some tests which relate to probability but avoid the use of the term.

For example, consider three events: 'having TB, given a positive test result'; 'having TB and getting a positive test result'; 'getting a positive result when one has TB'. In language these may be taken to be very similar or almost identical events; in probability the first is a conditional probability and the second is a composite probability (of both events taking place); the third one is a conditional probability which is inverse of the first one (linked by Bayes formula). Such subtleties are crucial but too rarely explored in the traditional classroom, except in the context of answering test or examination questions correctly. Another example is 'having TB' or 'positive test'. It is not easy to distinguish between the two terms logically. Moreover, the word 'or' is always used in mathematics to include both events as well as either one of two; in language the word is often thought to exclude the possibility of both events. We have tried hard to avoid such ambiguities but this was not always possible.

3. Context and data format

In our test we also explored how similar problems may lead to students experiencing difficulties in finding solutions. Sometimes the same basic information was used in terms of actual probabilities but the context was changed. Sometimes the format of the data was changed, such as using the word percentages or proportions instead of probabilities. Sometimes whole numbers rather than percentages were used, following on from the work of Gigerenzer. More direct examples are clear from the actual test which will be used for the presentation.

4. Results

Two courses have been presented to students on the didactics of probability and on decision theory respectively. The course on didactics was for students who were planning to become teachers and had already had formal courses on probability: five students attended. The second course was at a higher mathematical level and aimed at students in their final years of study: this was for six students. Hence the students were certainly well above average in mathematical ability. A key element to remember is that the courses were presented in English, which is a second language to them.

A startling finding from both courses was that most students were generally not able to apply Bayes theorem to these questions at the start of the course, despite prior exposure (in a conventional mathematical notation). There were two students (out of six on the second course) who could do the questions with good facility, but none of the students on the first course could solve the problems, indeed there were three who made mistakes on all the questions; for the second course, two were not able to solve any of the problems. Overall, of 11 students, five could not solve any problems, while only two displayed good facility. At the end of the first course, four of the five students improved their facility considerably; follow-up work to test long-term retention is planned later this year.

In terms of pedagogy, the approach was to stress various aspects to help students improve in tackling such problems. These included reminding students to look at questions carefully as well as offering different techniques such as tree diagrams or contingency tables, or whole numbers, to solve problems. This was done by working through an example (relating to breath tests for alcohol amongst motorists) in a systematic way in depth and detail. As an example this was found to be motivating and relevant by all the students, but this may not be the case in some countries or cultures (such as Saudi Arabia, where alcohol is mainly banned) or with students of different ages. Much time was spent in language and specially how a slight change in order or words makes a significant change in terms of the underlying perception and ensuing calculation. For example, we considered the following questions: Find the probability that a driver who is sober gives a positive reaction/ a driver is sober and gives a positive reaction/ a driver gives a positive reaction but is sober/ a driver who gives a positive reaction and is sober.

In terms of techniques, tree diagrams are very powerful but need care with regards to labeling (where there is variation between countries) and completion of all branches. Contingency tables are also valuable, but disguise the process of the two events in time. Gigerenzer (2002) strongly advocates the use of whole numbers, which certainly are easier than percentages and fractions but hide the underlying probability and uncertainty. All these aspects were discussed with students. For those on the first course, three of the five students showed substantial improvement when tested at the end of the course, one showed some improvement, and one was still confused. Results for the second cohort who are also studying more complex ideas of decision making and social choice will be presented at the conference.

5 Decision-making and inference

Our course in decision-making begins with paradoxes such St Petersburg and the one due to Allais. Students are asked to pose the problems to friends leading to valuable discussions in class. This motivates the need to study utility in the context of making decisions. We believe that this provides a more concrete introduction to inference, since a decision has to be made. Thus, though there are difficulties in assigning prob-

abilities, the need for a decision consolidates the approach for students who thereby develop a better feel for inference: that is our hypothesis which is in the process of being tested by the approach taken in the course. A stress has been placed on Bayes formula as well as the underlying ideas of coherence. Students also value the intuitive approach of using subjective probabilities in situations where a frequentist approach may be unsuitable. Early signs are that students are positive and motivated by the approach. A key difference is that there are opportunities to discuss ideas, have different opinions and develop deeper understanding of inference.
Personal probability can be applied to any decision-making situation and subjective probabilities are revised according to new information using Bayes formula connecting prior to posterior probabilities. A Bayesian can express his *opinion* of the probability that a coin will land heads, but not his *uncertainty* that it will land heads. For the latter would lead to making a statement of his uncertainty about uncertainty and to a logical regress. Thus, for Bayesians there is no objective or true probability in any situation of uncertainty. Of course, no means are provided for giving a prior probability or comparing prior probabilities of the same event by several people. Most strikingly, de Finetti states that “PROBABILITY DOES NOT EXIST”.

6. Conclusions

The results so far have confirmed that Bayes theorem and the applications to inference and decision theory are not straightforward even for mathematics students with extended exposure to these ideas. Whilst there is no independent test data, these students are certainly in the top quartile (or higher with regards to mathematical ability. Yet they find the ideas underlying decision-making quite difficult to use in practice. This confirms that the underlying intuitions are quite deep and hard to shift.

References

Barnett, V. (1973). *Comparative statistical inference*. New York: Wiley.

Borovcnik, M., & Kapadia, R. (2011). Modelling in probability and statistics–key ideas and innovative examples. In J. Maaß, J. & J. O’Donoghue (Eds.), *Real-World Problems for Secondary School Students–Case Studies* (pp. 1-44). Rotterdam: Sense.

Carnap, R. (1950). *Logical foundation of probability*. University of Chicago Press.

Finetti, B. de (1974). *Theory of probability*, New York: Wiley.

Gigerenzer, G. (2002). *Calculated risks: how to know and when*. New York: Simon & Schuster.

Huerta, M. P. & Cerdán, F. (2010). El cálculo de probabilidades en la formación del profesorado de matemáticas de secundaria. In M. Moreno, J. Carrillo, J., & A. Estrada (Eds.), Española de Investigación en Educación Matemática XIV (pp. 353-364). Lleida: Sociedad Española de Investigación en Educación Matemática.

Laplace (1774). Mémoire sur la probabilité des causes par les événements. *Mémoires de mathématique et de physique présentés à l'Académie royale des sciences, pars divers savans*, T. VI, 621-656.

Nikiforidou, Z. & Page, J. (2011): Risk taking and probabilistic thinking in preschoolers. In D. Pratt, *Working group on Stochastic thinking. Proceedings of CERME 7*. Online: http://www.cerme7.univ.rzeszow.pl/index.php?id=wg5.

Reichenbach, H. (1949). *The theory of probability*. Berkeley: University of California Press.

Mises, R. von (1928 / 1951). *Probability, statistics and truth*. Allen & Unwin.

Oakes, M. (1986). *Statistical Inference. A commentary for the social and behavioural sciences*. New York: Wiley.

Savage, L. et al. (1962). *The foundations of statistical inference*. London: Methuen.

Shafer, G. (1976). *A mathematical theory of evidence*. Princeton, N.J.: Princeton University Press.

Tversky, A. (2011) Thinking fast, thinking slow. London: Allen Lane.