TEACHING DECISION MAKING AND STATISTICAL THINKING WITH NATURAL FREQUENCIES

Laura Martignon and Christoph Wassner
Max-Planck-Institute for Human Development
Germany

Our comparative studies investigate the influence of different representations (i.e. formulas or graphical models and numeric formats) on the understanding of "big ideas" in stochastics (e.g. characteristics of probability, conditional probability, distribution, significance). We know from previous work (e.g. Sedlmeier & Gigerenzer, 2001) that special tree-representations combined with frequency-formats increase the performance in understanding dramatically. Another aspect of the experiments is the utility of different presenting-modes (e.g. static vs. dynamic, imitation vs. learning by doing). The pupils of age 15-19 receive a computer-based training with different representations resp. modes on basic probability tasks. The effects of the training are measured by subsequent tests. Thus we obtain insight, if they succeed easily in using the learned representations and if they benefit from it. The first results support the assumption that groups of pupils trained with frequency-representations have a better understanding of key-problems in stochastics.

INTRODUCTION

Probability theory has a right and a left hand. On the right is the rigorous foundational work using the tools of measure theory. The left hand thinks "probabilistically", reduces problems to gambling situations, coin tossing, etc. (Leo Breiman, 1968, preface iii).

In some fields of Mathematics an adequate representational environment has been used for centuries. One important example is the differential calculus. In fact, a crucial moment in the history of the differential and integral calculus, and consequently in the history of analysis and physics, was René Descartes' inception of what we call today the cartesian product, or also, the coordinate system. Descartes' first use of the cartesian product was in the melting of algebra and geometry into what is called analytic geometry that allowed, for instance, representing the conics in terms of simple algebraic expressions (an ellipse is given by $ax^2 + by^2 = c$). Later on, this representational environment made concepts like "velocity" and "space as a function of time" transparent and understandable. It was, in fact, the adequate representational environement for Newton's dynamical approach to his formulation of the laws of mechanics. Another, even older example of a successful process of search for the adequate representation: the representation of numbers as linear combinations of a base. Imagine how much more difficult multiplication is using Roman numerals (for those who have not worked extensively with them) than using Arabic base-ten notation.

Probability theory is a relatively young field in mathematics that is still searching for its adequate representational environment, that is, a combination of visual aids with notational tools that will connect the two hands described by Breiman above. Note that, while the right hand described above reduces probability to a Kolmogorov measure, the left hand counts events, going back (in history) to Laplacian quotients. The bridge between these two worlds is a formal edifice of theorems, of which the law of large numbers is the basic cornerstone. In essence, the representational environment is still basically reduced to the notations of measure theory and Venn diagrams (with the eventual addition of contingency tables and probability-labeled trees) for visualization of probabilistic concepts. This combination has many shortcomings. The phenomenon that is changing the representational problems of probability theory is the advent of the computer, which is finally providing visualizations that successfully link its two hands. At the same time, the tendency is now to go back to a more pre-set-theoretical treatment, and use Laplacian probabilities.

The aim of the paper is to show that an extensive use of "*natural frequencies*" combined with the adequate use of the computer for graphically exhibiting the bridge between frequencies and probabilities may provide the adequate representational environment for teaching probability in schools improving students' transfer of formal concepts into problem solving and also allowing "learning by doing". First results confirming these conjecture – namely improvements obtained by interventions in German Gymnasia – will be described and analyzed.

THE SITUATION OF TEACHING STOCHASTICS IN GERMAN SCHOOL

Recent questionnaire-surveys (Wassner & Martignon, not published.) show that German school students consider probability theory an extremely appealing branch of Mathematics. Yet, the same surveys reveal that in spite of its appeal, probability theory is considered more difficult than other mathematical fields. The problem seems to be the transfer of abstract concepts to problem solving, where the wordings of problems describe real-life situations. Whereas the probability-problems they encounter in daily life are concrete and numerical, the instruments are presented to the students in a highly formalized language. Often mnemonic rules are taught, like the "path rules" and the "rule of total probability", which are easily forgotten. Another aspect is that task contexts are seldom motivating or extracted from students' experience, being, in general, quite artificial.

THE PROBLEM OF REPRESENTATION

As already stated above, the way in which mathematical ideas are represented is fundamental to how people can understand and use those ideas. As Gigerenzer and Hoffrage (1995) have pointed out, the adequate representation makes the computation and often produces the solutions of the problems treated. When students gain access to good representations they have a set of tools that significantly expand their capacity to think the right way. "Representations must be treated as essential elements in supporting students' understanding of mathematical concepts and relationships; in communicating mathematical approaches, arguments, and understandings to one's self and to others; recognizing connections among related mathematical concepts; and in applying mathematics to realistic problem situations through modeling. New forms of representation associated with electronic technology create a need for even greater instructional attention to representation" (NCTM, 2000 p.66).

What we want to address here is the need of good representations for Bayesian inference tasks, assessing conditional probabilities and understanding the meaning of significance. If these basic tools become well enrooted in students' minds, they will easily remember and adopt them to everyday problems.

The term representation applies to processes and products that are observable externally as well as to those that occur "internally", in the minds of people doing mathematics. Humans – and students in particular - have primary intuitions about probability and we believe, in contrast to the heuristics and biases (Kahneman & Tversky, 1982) position, that these intuitions are "correct" in a normative sense. Thus our question is: What are these correct intuitions and to which representational environments do they apply? Our position is, that they are the result of genetic selection and learning processes and interaction with the natural environment. The mind works with representations of the environment to which it is well adapted (Cosmides & Tooby, 1996). The mind, Gigerenzer claims, is adapted to a representation via "natural frequencies", which is based on simple counting and is the same used by our ancestors, when they had to make inferences under uncertainty. The phylogenetic selection process is, of course, not a matter of some hundred years. Thus, one didactical recommendation is to use natural frequencies as often as possible to gain students' insight in the problems.

EXAMPLE: BAYESIAN INFERENCES

Perhaps the most striking examples of cognitive illusions with conditional probabilities are tasks solvable with the Bayes theorem. The use for medical risk communication or judgment at court give an impression of the relevance of this rule for the daily life. With respect to this, it is astonishing how small the attention to that rule and to related problems at German school is. Not to mention that the rule gives the basis for another approach to statistics that has already found a lot of followers in the statistical methods research.

The predominance of frequency representations on typical bayesian inference tasks in the context of medical diagnosis appeared very clearly within training-studies. With the help of frequency-tree representations long-term performance-rates of over 90% were attainable compared to rates of 20% in groups getting a formula-based training (Sedlmeier & Gigerenzer, 2001). A study which used a *frequency tree* in the complete version (see Figure 1) and detailed frequency information (e.g. "10 out of 1000 women of age 40 have cancer" etc.) for a typical bayesian task ("what is the probability of positive test result given the women has cancer?") achieved a performance-rate of nearly 80% which was given to people *without* training.

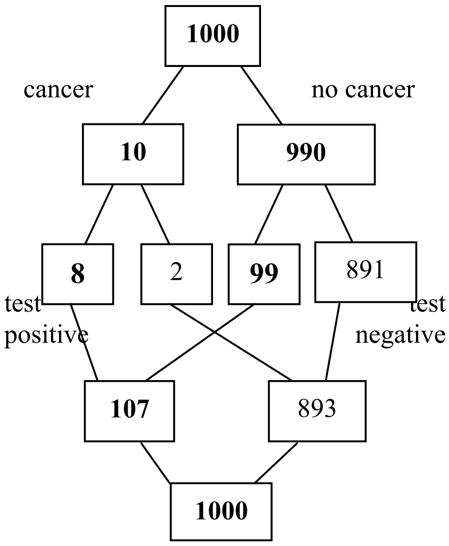

Figure 1. Frequency Tree, Complete Version.

We attribute this unusually high performance-rates from a didactical viewpoint in addition to the advantages and reasons described above to the close conformity of the frequency tree representation with the inductive way of information processing in the bayesian inference tasks. First we divide any sample of the relevant totality by a given base-rate (base-rate neglect will not be possible any more). We build up the first level of the tree, the first branching, from which we have a clearly separated subset of the totality - in the binary cue situation one subset with the condition and one without. The next step is the including of the information given for the conditional events, which leads to the second level of the tree. We reach the segmentation required by the task structure (conjunctive subsets). Henceforth we add the subsets regarding the “new” condition (which is in the 2x2 situation easy to identify) and get supersets, one superset with the “new” condition and at least one without (equivalent to the rule of total probabilty or second path rule). We reach the third level of the tree. Now it is possible to read off the solution simply by combining the relevant nodes. Combining means just calculating a Laplacean probability, that is dividing the number in the subset-node by the number in the corresponding superset-node.

Mathematically we would derive the Bayes formula from a definition of conditional probability (Laplacean probability) and do exactly the same but the other way round, in a deductive way. The mind is able to do both, but the more intuitive and plausible way is the inductive one. That´s what we know from learning theory. Just to mention that the fourth level, built by adding third level nodes, leads to the proof that we haven´t changed totality which could be misunderstood and is equivalent to the important so-called "first path rule".

Another advantage of this representation is that you can derive any bayesian “inversion” possible with the given information. The obvious advantage over the 2x2-table is that this tree-representation keeps the inductive information-processing-structure even in the general case, which means with more than 2 dichotomous cues (e.g. Monty-Hall-dilema, Krauss & Wang) or more levels (e.g. Simpson-paradox).

SOME EMPIRICAL RESULTS

In a study with school-students of secondary level (age 16-18) we compared frequency-representation-training with probability-representation-training. Within a computer-based training the students learned step-by-step to solve typical bayesian tasks with the help of a certain representation. The effects of the training were measured by subsequent tests. It was also required from the participants, to give statements about their solutions and to answer transfer questions. Thus we obtained insight, if they were able to succeed easily in using the learned representations and if they were able to benefit from it.

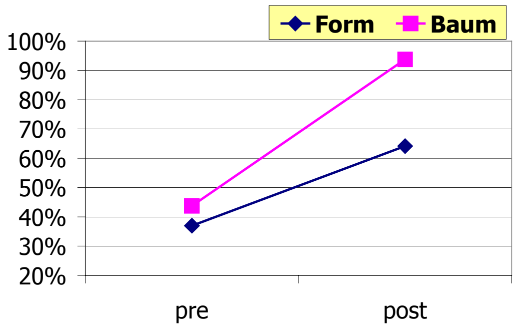

Figure 2 Performance Rates (Pre-post) of Students Trained with Frequency-Tree ("Baum") and Probabilty-Formats ("Form").

One result (see Figure 2) showed clearly that students trained with frequency-trees like in Figure 1 perform much better than students trained with probability-labeled representations. Note that in spite of the fact that the students had been taught at school in this topic beforehand, they didn´t perform well in the pretest. With frequency-trees they were able to reach performance-rates of 95% and were also much better in answering transfer questions. The empirical results are consistent with the theoretical result that the frequency format can be handled by an algorithm that is computationally simpler than that required by the probability format.

REFERENCES

Breiman, L. (1968). *Probability*. London: Addison-Wesley.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58*, 1-73.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychological Review, 102*, 684-704.

Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 509-520). Cambridge: Cambridge University Press.

Krauss, S., & Wang, X.T. (subm.). *The psychology of the Monty Hall problem: overcoming difficulties in solving a tenacious brain teaser.*

NCTM - National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.

Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General, 130*, 380-400.