Teaching Statistical Inference from Multiple Perspectives Integrating Diverging Schools of Inference

Ödön Vancsó, Eötvös Loránd University Budapest, Hungary
e-mail: vancso@cs.elte.hu, vancso.odon@gmail.com

Abstract

Carranza & Kuzniak (2008) have analysed the negative impact of reducing probability to a purely frequentist notion on the students' perception of the methods to learn. In the controversy between classical and Bayesian statisticians, to avoid a subjectivist conception of probability was the last 'argument' to help after inconsistencies in classical statistical inference have shown to be irreparable. However, both schools have their relative merits and flaws as seen from the foundational perspective. In the famous discussion in the American Statistician (1997), Moore finished with the argument that classical inferential statistics is easier to teach (and understand). Our conclusion conflicts with such views: Knowing the Bayesian way of thinking *enhances* the comprehension of classical methods. The ongoing debate on the difficulties and its implications – at secondary level inferential statistics nearly vanished from curricula worldwide – seems to undermine Moore's judgement. We develop the different views in parallel – so that students can understand them better. To support this sophisticated task we integrate reflections on philosophical issues, use paradoxes, and make extensive use of computers. We build on encouraging feedback from our past courses.

Key words. Statistical inference, Bayesian HPD regions, Bayesian statistics, conditional probability, confidence interval, favourable relation.

1. Introduction

This paper is based on a two-semester seminar at university (Bayes Statistics I and II, presented by the author at Eötvös University Budapest) since 2001; the seminar has been repeated yearly with modified content. More details about background and content of these special courses may be found in Vancsó (2005; 2009; 2010). This paper summarizes the main ideas and results. Official curricula in Hungary contain statistics since the year 2000. Earlier only elements of probability theory had been taught. T. Nemetz investigated ways to introduce inferential statistics intuitively. The main idea was forecasting specific events in the context of probability games (see Nemetz, 1989; Varga, 1976). The character of the suggested approach towards inferential statistics may be seen from Vancsó (2005):

"The key idea of our approach towards probability and statistics is that the two different ways of inferential statistics should be taught together at school-level, which also has a deep impact on the way how to teach the probability part [see also an analysis of conditional probability in Borovcnik & Kapadia, 2013]. There was an interesting and intense debate about classical and Bayesian ways of inference in teaching statistics at university in the teachers' corner of the *American Statistician* in 1997."

The discussion was led by Berry (1997), Albert (1997), Moore (1997), and Witmer et al. (1997). There were three different positions in this debate.

- The Classical group represented by Moore (1997) argued for the classical way and considered the Bayesian approach as too difficult for teaching.
- The Bayesian group led by D. Lindley who is convinced that – at universities – predominantly the Bayesian approach should be taught.
- The third and smallest group (that did not join the ASA discussion but formed its ideas at the same time) advocates an integrated way of teaching both approaches parallel to each other; see Migon & Gamermann (1999).

After a long time of experimenting, reading, and thinking, we decided to follow this third way in Hungary. The background of our long-term project is elaborated in Vancsó (2009):

"Our view was also deeply influenced by the collaboration with D. Wickmann and M. Borovcnik in the working group 'Stochastik in der Schule' [...] (Borovcnik, Engel, & Wickmann, 2001). Wickmann (1998) discussed both the philosophical and epistemological background of the confrontation between the first two groups. He argues that classical statistics introduced in the usual way is the wrong approach, because the frequentist interpretation of statistical results is generally false and in some cases is not at all appropriate (Wickmann 1998, pp. 57–60).

The Bayesian point of view 'attacks' the classical inferential approach because therein probability is reduced to solely an objective 'chance machine' interpretation: probability may only be interpreted by relative frequencies in independently repeated identical random experiments. In the more general cases of a unique situation, probability may not be used in classical theory, according to the Bayesian critique. Taking this argument seriously, any application of probability to a situation perceived as unique (not repeatable) would be 'forbidden'. [...] For teaching, this has unfortunate consequences: either to fudge such situations, or to exclude them. Neither is a good situation when one wants students to understand the concepts they apply. The [following section] illustrates why we have gradually become supporters of the third way."

2. Philosophy of mathematics as a background in statistics teaching

To clarify the status of mathematical concepts, it is helpful to take a philosophical perspective. Today, it causes no problem to acknowledge that there are different geometries depending on different systems of axioms; the only question might be in which situation they can be used if we want to describe a real situation. However, that is a question about modelling the situation or about the application of the theory and not about the theory itself. It is well-known that all different geometries are relatively consistent from a modern axiomatic viewpoint. We sometimes forget about the development of modern mathematics and return to the Greek basis and believe in our theorems as absolutely true statements. This step backwards might be one reason for the intense debate between classical and Bayesian statisticians; at times, the scientific dispute has adopted the character of a religious war.

The conflict in statistics seems to be quite comparable to the geometry debate of the 19th century. It is a false dichotomy to take either classical or Bayesian statistics. Both of them are sustained by consistent theories; the *choice* between these 'schools' comes up only in the case of application. It is very important to know about the mathematical theories in the background because it grants legitimacy not only to an objectivist interpretation of probability but also to the so-called "subjective probability"; the latter concept, too, is embedded within a consistent theory. Without such a theoretical justification, the Bayesian approach would not have any ground and could be dismissed as an interesting but non-scientific way of thinking. To address the different interpretations of probability is an old problem. Carranza & Kuzniak (2008) analyse the impact of focussing the theory completely on an objectivist concept of probability while the examples require a subjectivist concept to solve the posed problems; the tension between these extreme poles causes a great confusion on the part of the learners. This is a main reason why we decided for developing both concepts in parallel.

3. Seminars following the third way

The content and structure of the seminars was influenced strongly by our Hungarian roots which comprise an approach towards teaching concepts by paradoxes, which may be well seen from Székely (1986). T. Varga (1976) also used paradoxes with primary school pupils like "the long run" paradox, which seems to be in conflict with the tendency for searching for patterns in the emergence of random sequences after 10 heads in coin tossing, tail seems to be "more probable" for many.

This way of teaching can be seen clearly in the case of the favourable relation which plays a central role of the conception of the first seminar. This seminar deals with the probabilistic foundation and addresses the intuitive background of the Bayesian way of thinking including the favourable relation and its connections to the logical implication. In order to argue how a different logic is used in stochastic thinking contrasted to classical logic we briefly analyse a special relation between events, named “favourable relation”, which was introduced by Chung (1942) and used by Falk and Bar-Hillel (1983) in order to enhance some probabilistic and statistical paradoxes.

Case	*Sign*	*Description*
(1) $P(B\mid A) > P(B)$	$A \uparrow B$	A favours B or A influences B positively.
(2) $P(B\mid A) = P(B)$	$A \perp B$	Events A and B are stochastically independent.
(3) $P(B\mid A) < P(B)$	$A \downarrow B$	A does not favour B or A influences B negatively.

This relation can be considered as a weakened form of implication (Vancsó, 2009):

- Probabilistically taken, A implies B logically means if you presume (or imagine) that A has happened, then the probability that B will happen is 1 (true).
- Connected to this is the so-called favourable relation: A favours B does not mean that B is true if A is supposed to happen; but B will become *more probable if A* has occurred compared to the case when A has not occurred.

Some features discriminate the favourable relation from logic implication, e.g.:

- *Symmetry and asymmetry*: $A \Rightarrow B \wedge B \Rightarrow A$ then $A \Leftrightarrow B$; as not all statements are equivalent, implication is *not* symmetric. For two non-equivalent events, it holds $(A \Rightarrow B) \Rightarrow \neg(B \Rightarrow A)$, i. e., the logical implication is *asymmetric*.
- *Transitivity*: $A \Rightarrow B$ *and* $B \Rightarrow C$ then $A \Rightarrow C$, hence implication is transitive.

Such relations are deeply imprinted in our mind from early childhood and reinforced by teaching. It seems ‘paradoxical’ that the favourable relation violates them:

- *It* holds $A \uparrow B$ then $B \uparrow A$; the symmetry is true for all three versions of influence; i. e., the favourable relation is symmetric.
- *For* the transitivity, there is no general rule; sometimes it is true that $A \uparrow B$ and $B \uparrow C$ implies $A \uparrow C$ but sometimes this does not hold.

Advantages of the favourable relation are: (i) Students become familiar with conditional probabilities and their counterintuitive features; (ii) it allows an intuitive check of calculations; (iii) a lot of paradoxes may be clarified by using special properties of this relation, which differentiate it from the classical implication, see Vancsó (2009). Other rules of implication were compared to the favourable relation in Borovcnik (1992).

Ex.: If $A \Rightarrow B$ and $A \Rightarrow C$ then it holds $A \Rightarrow (B \cap C)$. The favourable relation violates such a rule; three different cases are represented in Figure 1, in case i. $\overline{A \cup B \cup C}$ has 4 points, in ii. 3, in iii. 2.

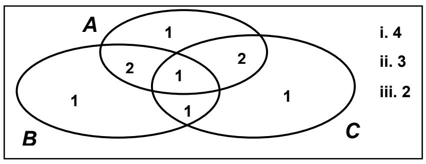

Figure 1.

All cases	Case i.	Case ii.	Case iii.	It holds
$P(B\mid A) = 1/2$		$P(B) < 1/2$		$A \uparrow B$
$P(C\mid A) = 1/2$		$P(C) < 1/2$		$A \uparrow C$
$P(B \cap C\mid A) = 1/6$	*and* $A \uparrow B \cap C$ $P(B \cap C) < 1/6$	*but* $A \perp B \cap C$ $P(B \cap C) = 1/6$	*but* $A \downarrow B \cap C$ $P(B \cap C) > 1/6$	

We were analysing typical situations where mistakes are caused by a wide-spread naive way of thinking; see the many fallacies in statistics such as Linda's fallacy (Tversky & Kahneman, 1973), or Simpson's paradox (Malinas & Bigelow, 2004; Morrell, 1999), or the Monty Hall dilemma (see the appendix of Vancsó, 2009). In teaching in class, the students were working in small groups on such puzzles or paradoxes trying to understand what happens and why it is contradictory to their expectation. Or, a story was introduced about Monty Hall and they may think about it with the aim to present their proposed solution of the problem. See Vancsó (2010) for further examples and more detailed analyses of comparing these relations.

4. Deliberate discussion on objective and subjective interpretations of probability

The different interpretations of the notion of probability are another topic on the agenda of the course; we analyse them using historical facts and texts as well. We clearly differentiate between the so-called 'objective' probability notion and the subjectivist view on probability.

- The objective term of probability can be used only in situations where a real 'machine' of chance exists, more abstractly formulated, a probability experiment exists, which can be repeated under the same circumstances; in those cases the relative frequencies show a special kind of stabilization.
- On the contrary, the 'subjective' probability notion is connected to our current level of knowledge about aspects not only in probability situations and may therefore be applied to a broader spectrum of problems.

For example, if we say 'the chance of failing this test is 60%' this is a subjective probability because there is no chance related to repeated experiments to get the relative frequencies. It is a *unique* case as tomorrow we will write a test. Based on information about the difficulty level of the test and our preparation efforts, we try to estimate the chances to pass or fail.

The course in the first semester has its own goals as well but it is an important prerequisite for the second course on inferential statistics where different probability notions and conditional probability and related rules are used e. g. by Bayes' theorem, which is discussed both for discrete and continuous distributions.

5. Classical and Bayesian methods of the Bayes II course in parallel

In the second semester, real problems are introduced which can be suitably analysed from both points of view. In that part we use, amongst others, the course elaborated by Wickmann (1991) but instead of criticizing the classical method we are building up both constructions and solving problems using the classical and the Bayesian method in parallel. At the end we discuss the different solutions and their interpretations.

In this part of the course, mathematical techniques gain momentum. The numerical solution of a problem sometimes takes several weeks using the two methods together, which occasionally requires totally different mathematical tools for each of the approaches. It should be noted that we use mathematical methods first in a simpler embedding and only later turn to computers for the calculations in more complicated examples. It is worth the effort we invest in the conceptual analysis because the students recognize several connections between stochastics and other topics of mathematics. This might reduce the outstanding role of stochastics within mathematics and strengthen the self-confidence of students in teaching probability and statistics later.

In the classical approach, parameters are simply constants, which are unknown. For the Bayesian approach, these parameters have to have a prior distribution as they are unknown. With the help of Bayes' theorem, this prior distribution is updated when data become known from a random sample. While the process of applying the theorem involves technicalities – only in a few examples the mathematics turns out to be quite easy. Thus, these technicalities have to be countered by suitable software. VisualBayes is an easy tool for this purpose (Wickmann, 2006); it can be used not only on PCs but on graphical calculators as well, as it is based on the computer algebra system Derive.

Regarding the usage of the methods from the two schools, the following 'rule of thumb' might help for orientation, which method to prefer:

- If we have a unique situation we should use the Bayesian approach; in this case we have to express our special information or pre-knowledge about the parameters by a suitable prior distribution.
- In the so-called production line (moving-band) situation we tend to use the classical methods following the ideas of Fisher or the work of Neyman and Pearson.
- In the pure classical approach we are not allowed to use our pre-knowledge for the process of modelling. Information always has to be "objective", which means it has to be independent from the person who models the problem. Information – at least potentially – has to be open for scrutiny by a repeated experiment from which one could check the assumed probabilities or probability distributions by the relative frequencies of the performed experiment. However, there is often no such experiment for the parameters of a distribution, which is chosen to model a variable, which is to be investigated.

6. Conclusions

The seminars have been evaluated continuously by written essays and interviews after its completion. The students appreciate this way of learning as very useful and they expect to be able to deal with the subtleties of the concepts successfully in their future career as teachers. Two citations from the many interviews may illustrate the impact of the approach:

"I understood the method of confidence intervals first after I had become more familiar with the Bayesian region of highest density [RHD]." and "I really like the Bayesian method because I saw for the first time why the people have different opinions in many cases. Because of the partners have different prior distributions."

These opinions express an essential advantage of the parallel approach. For more detailed feedback see Vancsó (2009). Finally we summarize the approach presented here and its potential to 'dissolve' the educational conflicts related to inferential statistics:

- Instead of teaching either a classical or (exclusively) a subjectivist approach to inferential statistics, one could follow the options of a blended approach or a parallel approach. The disadvantages of purely objectivist curricula have been dramatically shown by Carranza and Kuzniak (2008). A purely subjectivist approach might have similar drawbacks. A blending of the concepts from the different schools seems hopeless because of the irreconcilable character of the statements of the different schools from a philosophical point of view. To let the approaches stand as they are but to teach them *in parallel* and to learn about their character as well as their relative merits might be the more promising variant to solve the controversy in the foundations of inferential statistics for teaching.
- The reactions of students and the results of the experiments show that this approach is suitable to be worked out in a more detailed form. The next step will be a well-prepared teaching experiment with an accompanied statistical analysis of both the consequences on the beliefs of prospective teachers and the improvement in understanding the concepts of statistical inference and the stability in solving related problems.

References

Albert, J. (1997). Teaching Bayes' rule: a data-oriented approach. *The American Statistician*, 51(3), 247-253.

Berry, D. A. (1997). Teaching elementary Bayesian statistics with real applications in science. *The American Statistician*, 51(3), 241-246.

Borovcnik, M. (1992). *Stochastik im Wechselspiel von Intuitionen und Mathematik* (Interplay between intuitions and mathematics). Mannheim/Leipzig/Wien/ Zürich: Bibliographisches Institut.

Borovcnik M., Engel J., & Wickmann D. (Eds.) (2001). *Anregungen zum Stochastikunterricht* (New ideas for teaching stochastics). Hildesheim, Berlin: Franzbecker.

Borovcnik, M., & Kapadia, R. (2013). From puzzles and paradoxes to concepts in probability. In E. J. Chernoff, & B. Sriraman (Eds.), *Probabilistic thinking: presenting plural perspectives. Advances in Mathematics Education.* New York: Springer.

Carranza, P. & Kuzniak, A. (2008). Duality of probability and statistics teaching in French education. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education.* Monterrey: International Commission on Mathematical Instruction and International Association for Statistical Education. www.stat.auckland.ac.nz/~iase/publications/rt08/T1P2_Carranza.pdf.

Chung, K. L. (1942). On mutually favourable events. *Annals of Mathematical Statistics*, 13, 338-349.

Falk, R. & Bar-Hillel, M. (1983). Probabilistic Dependence between Events. *Two-Year-College-Mathematics Journal*, 14, 240-247.

Gardner, M. (1959). *The Scientific American book of mathematical puzzles & diversions*. New York: Simon & Schuster.

Malinas, G. & Bigelow, J. (2004). Simpson's paradox. *Stanford Encyclopedia of Philosophy*. Online: http://plato.stanford.edu/entries/paradox-simpson.

Migon, H. S. & Gamermann, D. (1999). *Statistical Inference – An integrated approach*. Arnold: London-New York-Sydney-Auckland.

Moore D. S. (1997). Bayes for beginners? Some reasons to hesitate. *The American Statistician*, 51(3), 254-261.

Morrell, C. H. (1999). Simpson's Paradox: An Example From a Longitudinal Study in South Africa. *Journal of Statistics Education,* 7(3). Online: http://www.amstat.org/PUBLICATIONS/JSE/secure/v7n3/datasets.morrell.cfm.

Nemetz, T. (1989) Valószínűségszámítás és statisztika speciális matematika osztályok számára (Probability and statistics for special classes in mathematics) Nemzeti Tankönyvkiadó, Budapest (National Schoolbook Publisher).

Tversky, A. & Kahnemann, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207-232.

Vancsó, Ö. (2005). *Classical and Bayesian statistics in the didactics of mathematics*. PhD thesis, University Debrecen (Hungarian with English summary).

Vancsó, Ö. (2009). Parallel discussion of classical and Bayesian ways as an introduction to statistical inference. *International Electronic Journal in Mathematics Education*, 4(3), 291-322. Online: www.iejme.com/032009/main.htm.

Vancsó, Ö. (2010). Mathematical logic and statistical or stochastical ways of thinking – an educational point of view. In C. Reading (Ed.), *Proceedings of the 8th International Conference on Teaching Statistics*. Online: iase-web.org/Conference_Proceedings.php?p=ICOTS_8_2010.

Varga, T. (1976). *Matematika 1* (Mathematics 1). Budapest: Academic Publisher (Akadémiai Kiadó).

Székely, G. J. (1986). *Paradoxes in probability and mathematical statistics*. Dordrecht/Boston: D. Reidel.

Wickmann, D. (1991). *Bayes-Statistik* (Bayes statistics). Mannheim/Leipzig/Wien/ Zürich: Bibliographisches Institut.

Wickmann, D. (1998). Zur Begriffsbildung im Stochastikunterricht (On the construction of notions in the teaching of stocastics). *Journal für Mathematik-Didaktik*, 19 (1), 46-80.

Wickmann, D. (2006). VisualBayes. Hildesheim, Berlin: Franzbecker.

Witmer, J., Short, T. H., Lindley, D. V. Freedman, D. A., & Scheaffer, R. L. (1997). Teacher's corner. Discussion of papers by Berry, Albert, and Moore, with replies from the authors. *The American Statistician*, 51(3), 262-274.