A Parallel Discussion of Classical and Bayesian Ways as Introduction to Statistical Inference – Teacher Training in Hungary

Vancsó, Ödön
Eötvös Lóránd University, Faculty of Sciences Institute for Mathematics
Budapest, Hungary
vancso@ludens.elte.hu

Summary

The purpose of this paper is to report on a research project at Eötvös Lóránd University in Budapest. The research is based on a new idea of teaching the basic notions of the classical and the Bayesian inference statistics parallel to each other to teacher students. The course, which is based on the above mentioned concept, lasts for two semesters and is optional for teacher students. The experiment has been going on since 2002. The content and the process of this course have been modified using my experience several times. This paper is meant to be dealing with the conclusions of the last course. I am also going to present some examples used during the course to illustrate its aims. The main result and conclusion can be summed up as follows: teaching both methods parallel to each other is very effective and helps to have a better insight in statistics first of all by clarifying the basic notions and skills. This result is supported by the remarks and notes of students taking part in the course. They had been studying probability theory before they started the above mentioned course and a lot of them confirmed that they understood the very notion of probability better and could apply the theory of probability more self-confidently in real situations. Finally, the results support my opinion about the fact that the view of the teacher on mathematics plays a decisive role in their later profession. This course is suitable to reflect a lot of philosophical problems as well, which is very important in letting students have a picture about modern mathematics and its applications.

Introduction

The key idea in this paper is that the two different ways of inference statistics should be taught together in the school. There was an interesting and hot debate about the classical and Bayesian ways of inference in teaching statistics at university in the teacher's corner of the American Statistician in 1997 (ASA, 1997). That was when I started to work and research on this topic and my own deliberations and results can be read in a Hungarian text with an English summary (Vancsó, 2005). There were three different directions in this above mentioned debate. The first is represented among others by Moore (Moore, 1997) who argued for the classical way and considered the Bayesian approach as basically inappropriate for teaching. The second group is the Bayesians led by D. Lindley who is sure of the fact that the Bayesian approach should be taught at universities. The third and smallest group is the representatives of the integrated way who are thinking and arguing about the parallel teaching of both approaches, see for example Migon & Gamermann (1999). This book gave me an impulse to work out a similar introduction for beginners in statistics available for secondary school teacher students, too.

After a long time of experience and thinking I decided to follow the third way. I worked out my arguments as well (Vancsó, 2005). My view was also deeply influenced by the cooperation and collaboration with D. Wickmann and M. Borovcnik in the working group "Stochastik in der Schule" of the German Society of Didactics of Mathematics. A booklet was published about this work (Borovcnik, Engel, & Wickmann, 2001). Wickmann exhaustively discussed both the philosophical and epistemological background of the confrontation between the first two groups in an article (Wickmann, 1998). I agreed with him in several points but I do not think that one of the ways is “right” and the other is “wrong”. He argues that classical statistics introduced in the usual way is the wrong approach, because the interpretation of statistical results is generally false and in some cases is not at all appropriate to be used. Opposed to the classical philosophy, the Bayesian point of view states that the classical inferential approach is wrong because therein probability has to be used only in objective “chance machine” interpretation and not in the more general cases when one is suffering from a lack of (subjective or objective) information. The next paragraph is to illustrate why I have gradually become a supporter of the third way.

What does mathematics mean from an ontological point of view?

Ancient Greek mathematicians thought mathematics to deal with absolute truth. If the axioms are true (they considered them to be true) and the deductions are made in a correct logical way then the resulting theorems are true. Accordingly, every proven theorem has to be (absolutely) true. Modern mathematics, to the contrary, says nothing about truth. If an axiom system is built up, axioms can be completely arbitrary. There is no requirement e.g. about its connection to the real world or an imagined truth. Theorems are the consequences of axioms therefore they have no connection to truth either. They can be deduced from axioms by using standard logical operations. In this modern axiomatic view, a mathematical theory can not be true or false – this is a wrongly posed question; it is only possible to investigate whether the system of axioms is either relatively consistent or contradictory. That was not so clear during the XIX. century, which was the reason for a long debate about the truth of the Euclidean or non-Euclidean Geometry.

Today it causes no problem to see that there are different geometries depending on different systems of axioms and the only question might be in which situation they can be used e.g. if we want describe a real situation. However, that is a question about the application and not about the theory itself. It is well known that all different geometries are relatively consistent. I think sometimes we forget about the development of modern mathematics and return to the Greek basis and believe in our theorems as absolutely true statements and that is one reason for the fight between classical and Bayesian statisticians. In my point of view the situation in statistics is similar to the Geometry debate of the XIX. century. It is a false dichotomy to take either classical or Bayesian statistics. Both of them are consistent theories, the choice between them comes up only in the application.

There is still one more didactical remark left: A theory is always better understood if it is opposed to another one. For example, the insight in the idea of the decimal system of numbers can be deepened if we know other systems for example the dyadic system. This didactical principle has guided me in my statistics teaching plans. These thoughts motivated me to work out a concept and materials for teacher students, which are suitable to show both concepts of inference statistics without wanting to set up priorities between them. In the next part I would like to summarize what is done in the frames of the above mentioned seminars.

Content of the seminars in two parts

In the first semester conditional probability and probability are discussed using real problems connected to them (see the examples below as well). There is a special “Hungarian way” of teaching by paradoxes, which may be well seen from several books (e.g., Székely, 1986). We were analyzing typical situations where mistakes or misuse were taken by using a familiar way of thinking; see the many fallacies in statistics starting with a lot of elementary cases such as Linda’s fallacy (Tverski & Kahnemann, 1973) or Simpson’s paradox or the Monty Hall dilemma. In this part, the favourable relation introduced by Chung in 1942 is discussed with the students. We can look at this relation as a weakened form of the implication. If *A* implies *B* that means if *A* has happened, then the probability of the fact that *B* will happen is 1 (true). In case of the favourable relation this is not the case but it is true that *B* will become more probable if *A* occurred as if it had not occurred. R. Falk and M. Bar-Hillel first analyzed this notion didactically and found relevant connections to the implication of the classical logic (see Falk & Bar-Hillel, 1983 and also Borovcnik, 1992).

Several different misconceptions can also be connected to this favourable relation. It is very useful to discuss the characteristics of this relation because it is based on conditional probabilities and gives a deeper insight to them. One possible dissolving of a lot of paradoxes can be reached by using the special opposite properties of this relation to the classical implication. The analysis of this relation helps the students to exercise the rules of conditional probabilities and see such interesting cases which seem to be surprising at first sight. It is useful for becoming familiar with conditional probabilities and their special rules as well and get an intuitive orientation about the effects of conditioning the probabilities on other events (or statements), which later are calculated formally by Bayes' theorem.

In this part, the different interpretations of the notion of probability are given and analyzed using historical facts and texts as well. We clearly differentiated between the so-called "objective" probability notion and the subjective or 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). To the contrary, the "subjective" probability notion is connected to our current level of knowledge about the things not only in probability situations and may therefore be applied to a broader spectrum of problems.

For example if I say "the chance of failing this test is 60%" this is a subjective probability because there is no chance related to repeating experiences and to get relative frequencies. It is a unique case, as tomorrow I will write a test. Based on some information about the difficulty level of the test and my preparation efforts, I try to estimate the chance.

The course in the first semester has its own goals as well but it is an important prerequisite for the second semester to inference statistics where different probability notions and conditional probability and its rules are regularly used e. g. the Bayes' theorem both in discrete and continuous distribution cases. In the second semester, such kinds of real problems are introduced which can suitably be analyzed from both points of view. In that part we used among others the course elaborated by Wickmann (1991) but instead of only criticizing the classical method we were building up both constructions and solving problems using the classical and the Bayesian method in parallel. In the end we were discussing the different solutions and their interpretations.

In that part of the course the different mathematical techniques are playing a greater role. The numerical solution of a problem takes sometimes several weeks regarding the two methods together, which occasionally requires totally different mathematical tools for each of the approaches. It has to be noted that we first always use pure mathematical methods and only later computers. The waste of time is getting back because the students see several connections between stochastics and other topics of mathematics. This might reduce the outstanding and singular role of stochastics within mathematics and strengthen the self confidence of students in teaching probability and statistics later.

The next paragraph illustrates a suitable example. The main observation regarding the usage of the methods made by me is the following “rule of thumb”: if we have any special information or pre-knowlegde about a unique situation we should rather use the Bayesian approach. It can be true without extra information before but a unique case only. In the so-called production line (moving-band) situation we rather tend to use the classical methods followed Fisher, or Neyman and Pearson. More notes will be made in the next paragraph and in the conclusions section.

Examples for problems

Two problems should illustrate the approach. One is from the first semester and the other one is from the second semester.

1. Three different contexts but mathematically isomorphic situations:

The following three paradoxes are analyzed: Prisoner dilemma (Gardner, M., 1959), Monty Hall dilemma (vos Savant, 1991; see also Vancsó & Wickmann, 1999) and the three discs problem (Varga, 1976). We elaborate on also on issues like why they seem to be so different regarding the inherent level of difficulty. It is important to see how the new information can influence the probability of an event. The point of these examples is to illustrate what a new piece of information means and what it does not. These problems serve as an excellent opportunity to analyze conditional probabilities and to use Bayes’ theorem; they amount to an ideal preparation for the Bayesian approach. We can analyze it using only objective probability and of course in a broader sense as well. For the key ideas and how they could be applied here, see Vancsó & Wickmann (1999).

A very interesting task is to formulate the isomorphism. The isomorphism between the first and second problem is quite easy to see. There are some problems in connection to the third version. These problems show another aspect as well, namely that we focus on symmetry all-too much: in teaching probability there are a lot of cases where everything is symmetrical and equiprobable e. g. coins, dices etc. This fact misleads us because there is a crucial asymmetry in these cases. In all problems there are three different opportunities (maybe with the same probability) and later on we get a piece of information which eliminates one opportunity. The question is how the chances of the two remaining possibilities have changed. Surprisingly, in all the cases the two remaining possibilities have lost their previous symmetry and are asymmetrical now; they have not got the same chance as we could think. It is important to note that the isomorphism was always found by the students themselves at least between the first two problems.

A sketch of an isomorphism between the second and the third problem might illustrate matters more in detail. In both cases there are three possible outcomes: where the car is hidden among the three closed boxes or which disc was chosen from the three different ones. There is a moderator but with a different task in the two situations. In the second he *knows* what we have to find out i.e. where the car is hidden and after our first choice he shows us an empty box from the remaining two (and he is able to do that because he really knows where the car is). Thereafter we have to decide to retain our first choice or to change our decision. The question is what has to be done and why. In the third situation the moderator has chosen a disc and shows us the colour of one side of the disc and we have to bet on the colour of the other side. This excludes one disc and the question is: do the two remaining discs have same chances or not. The “translation” is the following:

(a) The moderator chooses one disc that corresponds to choosing one box for the present in the Monty Hall dilemma. Then we choose one disc (one box). The first step is just imagined but without it we won't see the isomorphism.
(b) The moderator shows us one empty box (one side of the disc). It closes out one box (one disc).
(c) We either decide to remain at the first choice (box or disc) or change. It has to be slightly modified in case of the disc problem. "Change" in this situation means if we choose the opposite colour as the colour of the side of the disc shown to us and "retaining the choice" means here if we bet on the same colour as it was shown to us.

The chance for winning with strategy "change" is $2/3$ and the opposite (conservative) strategy has only the probability of $1/3$. These calculations are valid only under certain assumptions but this is a longer story, see again Vancsó & Wickmann (1999). Here should only be noted that the current modelling of the situation comprises also that the moderator *always* makes us the offer of a choice, which is not always sensible in the second situation where the moderator could "tease" or "help" us also.

2. Lotteries

There are different lotteries in European countries, which means that the two numbers (how many balls are in the box and how many are chosen from it) are different. For example there are three different lotteries only in Hungary. The oldest one contains 90 balls and there are 5 chosen from the urn. In the 1990's two new types of lotteries were introduced. In the first, 6 balls are drawn out of 45; in the one named Scandinavian, they select 7 out of 35 balls (Vancsó, 2006). Consider the following situation: somebody, who does not know how many balls there are in the oldest lottery, arrives in Hungary. His question is the following: he knows the result of one lottery (the chosen numbers) and he has to estimate how many balls are in the box from which the numbered balls are taken. There are two possibilities to apply: the classical confidence interval or the Bayesian region of highest density, which is referred to as RHD in the following.

In order to produce a confidence interval from the results of a special week's results gives a possibility. We are using the maximal number estimation (there are different statistics of course but this one is an unbiased and efficient estimator). The classical solution uses no extra information (this could not be involved even if it existed) whereas the Bayesian way gives different results under different circumstances. For example, if I have the information about the ball numbers that there can not be more than one hundred of them, it is not equal to the situation without this piece of information. As prior distribution, the uniform distribution may be chosen on the interval from the chosen maximum number to the imagined maximum number of the balls. It is interesting to compare the classical and the Bayesian solution in the case of uniform prior distributions up to n, where n tends to infinity.

There is a purely mathematical question related to this as well: in what situation the following statement is true: The 0.95-confidence interval is numerically the same as the 0.95-RHD provided we use uniform prior distributions. In the case of the oldest Hungarian lottery we have had more than 2100 experiments since its introduction in the year 1957. We are able to control our result using the statistics of these 41 years. It is possible in both cases. I reported the results in detail elsewhere (Vancsó, 2004). Now I summarize only the most important remarks in the next paragraph.

Students' notes and remarks

I would like to show some students' opinions and concrete materials made by them. All in all the students always found this way of teaching useful and in the end of the course they gave interesting essay questions and their solutions. They are of the opinion that this way of learning is very useful and they can see more deeply behind the things later as teachers. These statements are well documented by student's essays written after the courses in last years. Some citations based on students' interviews are:

> "I understood the confidence interval first after I had become more familiar with the Bayesian region of highest density (RHD)." (student in 2004)
>
> "I always interpret the classical result wrongly because I thought the confidence interval contained the estimated parameter with the given probability which is usually selected as a high figure. I have understood at last what it really means." (student in 2005)
>
> "I really like the Bayesian method because I saw for the first time why people have different opinions in many cases. Because of the partners have different prior distributions." (student in 2007)

These opinions express an important advantage of the parallel approach.

About the lottery problem, two students have initiated an interesting development. One of them, Rita Deme-Farkas analyzed the oldest lottery (5 chosen number balls) and found an interesting connection. Her result is shown in following table:

Maximum number of the week		$x_5 = 55$	$x_5 = 61$	$x_5 = 79$	$x_5 = 85$
Upper side Confidence interval		99	110	143	155
Upper side Bayesian RHD	$M = 100$	90	93	98	99
	$M = 150$	106	115	133	137
	$M = 250$	112	124	158	168
	$M = 1\,000$	113	126	164	176
	$M \to \infty$	115	128	166	179

M denotes the largest possible number for the balls in the lottery and the figures in the table are the upper bound of the 0.95-RHD and of the classical 0.95-confidence-interval in the first row. It shows that the classical and the Bayesian intervals are not numerically equal in the case of "zero information" (uniform prior distribution). An important result is if the possible maximal number of balls is less than 100, then RHD gives much more precise results. It is an interesting question whether there is such a number M for which the confidence interval and the RHD are numerically equal. At about 120 as maximal number the same numerical results might occur.

The second work is performed by Anna Szabó who experimented with the lottery of six balls from 45. The results of the weeks inspected by her were 34, 35, 42 and 45 as maximum of the drawn numbers. In these cases she found the following 0.95-confidence intervals and 0.95-RHD if we use the relation M tends to infinity:

Confidence interval	RHD $M \to \infty$
[34, 55]	[34, 59]
[35, 56]	[35, 61]
[42, 68]	[42, 73]
[45, 73]	[45, 79]

In this case, the following remark seems to be true. If M is less than 100 and a uniform prior distribution is used, then the RHD produces a smaller interval than the classical confidence interval.

It is interesting to analyze both methods and compare the results to each other (see above). I think if somebody knows different constructions for solving a problem, they understand each of them better. This semester Hana Burján (a student who had studied engineering and economics too and now she would like to be a mathematics teacher) held a presentation about an estimation problem solving it by both methods and could perfectly present both interpretations. She stated that she has understood finally what confidence intervals really mean and could contrast it perfectly to the notion of Bayesian RHD. She was very suggestive and the rest of students participating at the seminar understood her. It is pity that this presentation was not videotaped to support the following idea.

I tried to carry out such didactical principles which are general enough to serve as a basis for teaching inference statistics. One of the important ideas is to contrast and oppose new concepts to each other right from the beginning. Confidence intervals can be better understood if the Bayesian interval of highest density is also built up and contrasted to the confidence interval. My experience supported this principle which I try to substantiate by students' works and interviews as well. Students found it important to understand the notion of conditional probability and manipulate with it. Bayes' theorem, which plays central role only in the Bayesian inference statistics (although it used in classical probability theory as well but in a restricted sense) shows, how our "knowledge" develops about uncertain things. That has been frequently a cornerstone of students' opinion. An important note is that T. Bayes himself was not "Bayesian", he did not think about subjective probability.

Conclusions

- I think the work which has been done shows that this concept can be a basis for teacher students to study inference statistics.
- The reactions of students and the results of the experiments confirm that this idea is suitable to be worked out in a more detailed form.
- The next step could be a well-prepared statistical analysis in order to prove the effectiveness of this teaching method.
- I am planning to write a book on the concept and the first results of our piloting courses.

References

ASA, 1997, *The American Statistician* 51, No. 3. *Teacher's corner.*

Borovcnik, M. 1992, *Stochastik im Wechselspiel von Intuitionen und Mathematik,* BI Verlag Mannheim (German).

Borovcnik M., Engel J., & Wickmann, D. (ed.) 2001, *Anregungen zum Stochastikunterricht,* Franzbecker, Hildesheim (German)

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

Gardner, M. 1959, The prisoner's dilemma. *Scientific American*, Mathematical games.

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

Moore D. S.: 1997, Bayes for Beginners? Some reasons to hesitate. *The American Statistician* 51, (3), 254-261 and 272-274.

Székely, G. 1986, *Paradoxes in probability theory and mathematical statistics*, Reidel-Kluwer, Dordrecht-Boston Mass.

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

Vancsó, Ö. 2004, Schätzung der größten Lottozahl aus dem Maximum der gezogenen Zahlen, *Beiträge zum Mathematikunterricht 2004*, Franzbecker, Hildesheim, 621-625.

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

Vancsó, Ö. 2006, Glücksspiele in Ungarn, *Mathematikinformation* Nr. 45, September, 20-30. (German).

Vancsó, Ö. & Wickmann, D., 1999, Das Drei-Türen Problem in Bayesscher Sicht, in: *Beiträge zum Mathematikunterricht 1999*, 551-555, Franzbecker, Hildesheim, Berlin (German).

Varga, T. 1976, *Matematika 1*, Akadémiai Kiadó, Budapest (German).

vos Savant, M. 1990, The Monty Hall dilemma. *Parade Magazine*, Ask Marilyn column, 9. September 1990.

Wickmann, D., 1991, *Bayes-Statistik*, BI Verlag Mannheim (German).

Wickmann, D., 1998 Zur Begriffsbildung im Stochastikunterricht. *Journal für Mathematik-Didaktik* 19 (1), 46-80 (German).

Appendix

"The prisoner's dilemma":

In this problem, three prisoners *A*, *B*, and *C* with apparently equally good records have applied for parole, and the parole board has decided to release two of them, but not all three. A warder knows, which two are to be released, and one of the prisoners (*A*) asks the warder for the name of the one (of the other two) prisoners who is to be released. While his chances of being released before asking are 2/3, he thinks his chances after asking and being told "*B* will be released" are reduced to 1/2, since now either *A* and *B* or *B* and *C* are to be released.

Monty Hall problem

This problem is named for its similarity to the *Let's Make a Deal* television game show hosted by Monty Hall. It is stated as follows: Assume that a room is equipped with three doors. Behind two are goats, and behind the third is a shiny new car. You are asked to pick a door, and will win whatever is behind it. Let's say you pick door 1. Before the door is opened, however, someone who knows what is behind the doors (Monty Hall) opens one *of the other* two doors, revealing a goat, and asks you if you wish to change your selection to the third door (i.e., the door which neither you picked nor he opened). The Monty Hall problem is deciding what to do: change your choice or retain it.

Three discs problem

This problem was first by Tamás Varga. There are three discs marked as in Figure 1.

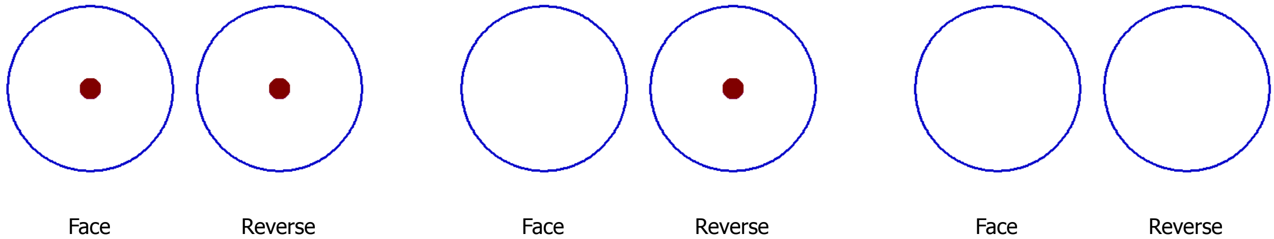

Fig. 1: Varga's discs

One of these discs is held up to the children; only one side is shown to them and they are asked to guess what is on the reverse ‚spot or blank' (I used two different colours in my example). After a series of random guessing and getting the other side of the disc shown to see whether they had made the right guess, the children were asked to devise and write down a strategy for guessing, which they would apply each time subsequently. Only for illustration, one experiment is cited about this game:

A teacher plays this game with 10-11 years old children. He reported his observation briefly. “Some tried repeating the last situation each time, others used blank and spot alternately. None chose the best strategy (whatever is on the face is most likely to be the same on the back). He then let one child use this and the results showed that he consistently scored best over a range of fifty tries. The children began to think and to suggest reasons as to why this might be. The thinking was intuitive, no one came up with a numerical solution but their answers showed that they had begun to grasp some of the ideas inherent in probability.”