PARALLEL DISCUSSION OF CLASSICAL AND BAYESIAN WAYS AS AN INTRODUCTION TO STATISTICAL INFERENCE

Ödön Vancsó

ABSTRACT. The purpose of this paper is to report on the conception and some results of a long-term university research project in Budapest. The study is based on an innovative idea of teaching the basic notions of classical and Bayesian inferential statistics parallel to each other to teacher students. Our research is driven by questions like: Do students understand probability and statistical methods better by focussing on subjective and objective interpretations of probability throughout the course? Do they understand classical inferential statistics better if they study Bayesian ways, too? While the course on probability and statistics has been avoided for years, the students are starting to accept the "parallel" design. There is evidence that they understand the concepts better in this way. The results also support the thesis that students' views and beliefs on mathematics decisively influence work in their later profession. Finally, the design of the course integrates reflections on philosophical problems as well, which enhances a wider picture about modern mathematics and its applications.

KEYWORDS. Bayesian statistics, favourable relation, statistical inference, confidence interval, Bayesian regions of highest posterior density (RHD).

1. INTRODUCTION

Teacher education – context of the project

Teacher students in Hungary undergo a thorough mathematical education at university. This also includes mathematically orientated courses on probability theory and statistics. The two-semester course described in this paper is optional for teacher students; it goes beyond that curriculum and is based on the parallel concept of inference. The teaching experiment has been going on since 2002. The content and the process of this course have been modified using our experience several times. This paper draws conclusions of the experience of past courses.

Bayesian statistics I–II courses have been held since the academic year 2001/02. The courses have been organized for mathematics teacher students who had studied probability and some very elementary statistics. The probability course is compulsory and not too popular because it is very theoretical. If someone is interested more deeply in probability and statistics, then he or she can choose a block of "probability and statistics" during the third and fourth year of study. This block contains four courses of 2 hours per semester each. These courses are: advanced probability theory, stochastic processes, mathematics of insurance, and introduction to mathematical statistics (this course deals with classical statistical concepts only).

Over the years, this block has been chosen by not more than 3–4 students, sometimes nobody was enrolled. It means that we could not organize a control group to teach an alternative course for comparison. Most participants of our parallel course had not taken part in this block before. The number of participants varied between 8 and 20.

The Bayesian controversy and reactions to it in the German didactics of probability

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 has also a deep impact on the way how to teach the probability part. 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 (ASA 1997).

There were three different directions in this debate.

- The Classical group is represented amongst others by Moore (1997) who argued for the classical way and considered the Bayesian approach as basically inappropriate for teaching.

- The Bayesian group is led by D. Lindley who is convinced that at universities, predominantly the Bayesian approach should be taught.

- The third and smallest group advocate an integrated way of teaching both approaches parallel to each other; see for example Migon & Gamermann (1999). This book gave us an impulse to elaborate a parallel introduction for beginners in statistics intended to be also used for secondary school teacher students.

After a long time of experience and thinking, we decided to follow the third way. We worked out our arguments as well.

Our view was also deeply influenced by the 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 (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 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". There are many more such situations than generally acknowledged.

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 next paragraph illustrates why we have gradually become supporters of the third way.

2. MATHEMATICS FROM AN ONTOLOGICAL POINT OF VIEW

The status of truth in mathematics

Ancient Greek mathematicians thought mathematics deals 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 cannot 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 19th Century, which was the reason for a long debate about the truth of Euclidean or non-Euclidean Geometry.

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 e.g. if we want to 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. 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 application. It is very important to know about the mathematical theories, because it gives legalization 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.

Didactical insights

There is one further didactical remark to be added to the discussion:

> A theory is always better understood if it is put into contrast to another one. Flexible and reflecting knowledge might enhance the individual acquisition of concepts.

School mathematics focuses on techniques and algorithms; as a consequence of this, pupils often pursue only mechanical procedures and algorithms without reflecting why the chosen method works; or why it fails sometimes. Teacher students are no exception to that. To arouse their reflection by suitable situations and pertinent discussion might help them to become better teachers and to get a deeper insight into mathematics as it really is. Generally, the image about mathematics lies very far from the reality in many schools.

For example, insights into the decimal system of numbers can be deepened if we also know other systems such as the binary system in parallel. This didactical principle of diversity has guided us also in our statistics teaching plans. These thoughts motivated us to elaborate an approach and materials for teacher students, which are suitable to explain both concepts of statistical inference without setting priorities between them. In the next section, activities and the framework of the above mentioned course are summarized.

3. CONTENT OF THE PARALLEL COURSE

Paving the way – discussing paradoxes and private conceptions

In the first semester, conditional probability and probability are discussed in the context of real problems (see the examples below as well). The approach has four different foci:

- Learning by paradoxes; clarifying which intuitions are led astray by the paradox and how they may be resolved by discussion and introducing clear concepts.

- Learning by analyzing private heuristics used in probability problems; we check in which ways these heuristics work and how and when they might lead to systematic bias.

- Discussing unusual concepts, which are more open to intuitive interpretation and may thus serve as a link between abstract concepts and the world of intuitions.

- Enabling a thorough discussion about different interpretations including the historical context to avoid unplanned transfer of ideas from the subjective to the objective corner; the confusion of ideas from both sides is a potential source for misunderstanding of abstract concepts.

There is a special "Hungarian tradition" of teaching concepts by paradoxes, which may be well seen from several books such as Székely (1986). T. Varga (1972) also used paradoxes (see the two discs problem later); 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.

We were analyzing typical situations where mistakes or misuse were committed 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 (Tversky & Kahneman 1973, or the Conjunction fallacy n. d.), or Simpson's paradox (see Malinas & Bigelow 2004, or Morrell 1999), or the Monty Hall dilemma (see the Appendix A for the latter). In teaching in class, normally the students were working in small groups on problems like Simpson's paradox; they try to understand what happens here and why it is contradictory to our expectation. Or, a story was introduced about Monty Hall and then they may think about it with the aim to present their proposed solution of the problem.

> The ensuing debate helps them to understand the situation better and to see how their colleagues think differently about the problem.

The students like these paradoxes, which are often chosen as topic in their final report in the seminar. It is important to mention they always find a topic for this final report, which in fact is optional. One student wrote a diploma thesis about Simpson's paradox later. Another student chose to analyse craps game in casinos using false dice for his thesis.

Logic and the favourable relation

The favourable relation is another important topic of the seminar; this concept was introduced by Chung (1942). We can consider this relation as a weakened form of logical implication:

- Probabilistically taken, *A* implies *B* logically means if you presume (or imagine) that *A* has (fictionally) 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* (fictionally) happens; but *B* will become more probable if *A* occurred compared

to the case when A has not occurred.

Falk & Bar-Hillel (1983) first analyzed this notion didactically and found relevant connections to the implication of the classical logic (see also Borovcnik 1992). This relation will be denoted by

- $A \uparrow B$ if and only if $P(B \mid A) > P(B)$.

- $A \downarrow B$ if and only if $P(B \mid A) < P(B)$

- $A \perp B$ if and only if $P(B \mid A) = P(B)$

The three cases are exhaustive and no two of them can occur simultaneously. The last relation is the well-known independence of events.

After introducing this relation we discuss its most important characteristics. The comparison of this relation to the logical implication is important; logical implication is only an extreme case of favourable relation expressed numerically by $P(B \mid A) = 1$, which means $A \Rightarrow B$. The logical implication follows some routine rules; for example:

- Asymmetry: $A \Rightarrow B \wedge B \Rightarrow A$ is equivalent to $A \Leftrightarrow B$; as not all statements are equivalent, the logical implication is asymmetric; i. e., there exist pairs of A and B for which it holds: $A \Rightarrow B$ *and* $\neg (B \Rightarrow A)$.

- Transitivity: $A \Rightarrow B$ and $B \Rightarrow C$ then $A \Rightarrow C$ is also true, hence the implication is transitive.

Such relations are deeply imprinted in our mind from early childhood and in primary and secondary school. It is very surprising that neither of these rules is valid for the favourable relation:

- 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 (see Figure 1, or click here to see an animated graph to demonstrate this).
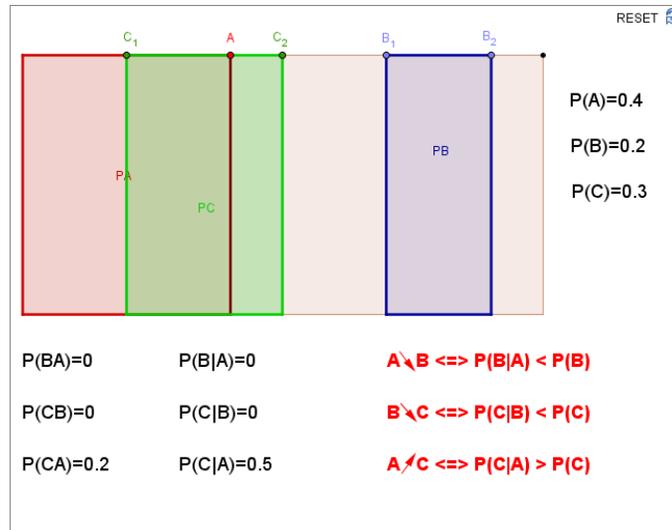
Figure 1. A counterexample to show that the favourable relation is not transitive – click to see the animation.

Advantages of the favourable relation:

- Students become more familiar in dealing with conditional probabilities and their unexpected, counterintuitive features.
- It allows an intuitive check for formal calculations.

This prepares them to understand the differences between a correct interpretation of classical inference results and the often used false interpretation (cf. Gigerenzer 1993); it also enhances the subsequent Bayesian way of inference. This relation is useful for becoming familiar with conditional probabilities and their special rules as well and to get an intuitive orientation about the effects of linking the probabilities to other events (or statements), which later are calculated formally by Bayes' theorem.

The other advantage of the relation is that it allows an intuitive check for formal calculations. It may accompany the analysis of conditional probability problems on the intuitive level. A lot of paradoxes may be clarified by using the special properties of this relation, which differentiate it from the classical implication. It is important to note that the situation here sets itself apart from a strategy that is well-known in mathematics:

When we generalize a concept, we tend to transfer our rules of the "old" concept to the "new", more general concept. In introducing the real numbers e.g., we strive to preserve the rules of counting and calculations among the more general number set as well. This tradition is broken here. That is a possible reason why we sometimes perceive a paradox with a new or more general concept; the missing transitivity in the case of the favourable relation is such an example.

Planned discussion of objective and subjective interpretations of probability

The different interpretations of the notion of probability are another topic on the agenda of the course; we analyze them using historical facts and texts as well. We clearly differentiate 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.

- 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 repeating experiences and to get 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 chance.

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

Classical and Bayesian methods in parallel

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 use, amongst others, the course elaborated by Wickmann (1991) but instead of only 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, the different 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 always use mathematical methods first and only later turn to computers for the calculations. It is worth the effort and time we invest in the conceptual analysis because

the students recognize 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.

In the classical approach, parameters are simply constants, which are unknown. For the Bayesian approach, these parameters, as unknown, have to have a prior distribution. 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 mathematical technicalities, for some nice examples the mathematics turns out to be quite easy in practice. However, for the bulk of real problems, these technicalities have to be solved by suitable software. VisualBayes program 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 might be preferred:

- If we have a unique situation we should preferably 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 Fisher, or Neyman and Pearson.

> Used information always has to be "objective":
> - How to judge that information is objective?
> - How to integrate qualitative knowledge?

In the "pure" classical approach we are not allowed to build in our "pre-knowledge" into the process of modelling. Used information always has to be "objective", which means it has to be independent from the person who models the problem with the aim to derive an estimate or to find and justify a decision. 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.

4. EXAMPLES USED IN THE COURSE

Two problems should illustrate the approach. One is from the first semester and the other one is from the second semester. Of course the first has no direct connection to Bayesian approach but we discuss it as it prepares the Bayesian way of thinking.

Different contexts yet mathematically isomorphic situations

The following three paradoxes are analyzed (see Appendix A for details):

- Prisoner dilemma (Gardner 1959)

- Monty Hall dilemma (vos Savant 1990; see also Vancsó & Wickmann 1999)

- The three discs problem (Varga 1976).

We also expand 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 convey. 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 we can introduce probability also 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 between the problems. It means that we have to translate a task into the language (text) of the other task. If this translation is perfect than we say the two tasks are isomorphic. The isomorphism between the first and second problem is quite easy to see. There are some problems in connection to the third version. We sketch the solution in the next paragraph.

These problems show another aspect as well.

In teaching probability, we focus too much on symmetry: 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 these problems there are three different options (initially with the same probability) and later on we get a piece of information which eliminates one option.

> Symmetry may be distorted by the information given.

The question is how the chances of the two remaining options have changed. Surprisingly, in all the cases the two remaining options have lost their previous symmetry and are now asymmetrical; they have not retained their same chance as we may think. It is crucial in understanding these problems that the information from the moderator or from the prison guard does not convey extra information about the first chosen box or for the prisoner himself who is asking the guard; however, it *is* favourable for the third box or the third prisoner who has not been mentioned yet. Thus, the symmetry is distorted by the information given. It is important to note that the isomorphism has always been found on our course by the students themselves, at least between the first two problems.

Isomorphism or equivalence

Isomorphism is a very precise mathematical concept: the sense in which two situations are isomorphic is heavily dependent on the characteristics, which are taken into account. From a mathematical standpoint it has to be clearly stated what is relevant; from the individual's perspective many other characteristics can count.

In saying that it is an easy task to establish a one-to-one relation between the first two situations we should also note that people would associate different values to the objects which are matched to each other: in the prisoner's dilemma "to be condemned" is an adverse consequence, but the matched object in the Monty Hall problem is to "to win the car", which is very good. Moreover, the consequences of the mathematical analysis for the situation are – despite an isomorphism – not the same, which may seem puzzling:

o In the prisoner's dilemma the probability to be condemned remains the same at 1/3 even if we are given information about one of the others who is released; but no decision or consequence arises.

o In the Monty Hall problem, the probability of winning the car also remains the same after the moderator opens another door with the goat. Here, however, the consequences are that we are unhappy because our probability of winning is 1/3 and this is now (considerably) less than not winning the car.

Connected to the Monty Hall problem is also a decision, i.e. we can change the final result by switching our decision, an option we do not have in the prisoner's dilemma. The connected values change the judgement about the situation (desirable or not desirable) despite the same probability, which has not changed in the two (isomorphic) situations. The connected

decision (no decision in prisoner's dilemma, a choice in Monty Hall) also changes the situation as a whole completely.

Hence care must be taken in order to clarify the restrictions of such an isomorphism. It is important to remark that isomorphism always is *relative* and not absolute; isomorphic in a specific sense. Such a phenomenon is often the case with mathematization of situations and might be profitably discussed in teaching. The features of the situations involved could be value-laden and emotionally linked, which might cause difficulties in the educational process and might even hinder learners to accept the concepts discussed and thus hinder the positive effects of using isomorphism in teaching. However, if put openly to the fore, issues like that could open the discussion about the mathematization process as a whole, as such processes always have to focus on some specific aspects of a situation and ignore others. It is valuable to discuss such issues so that the idea of isomorphism can be understood by students; indeed one might actually prefer to call it equivalence as this concept is less strict and may account better for the different perceptions of the situations.

Further insights and their mathematical modelling by isomorphism

A sketch of an isomorphism between the second and the third problem might illustrate matters in more 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 imaginary but without it we would not see the isomorphism.

(b) The moderator shows one empty box (one side of the disc). It eliminates one box (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.

Understand the underlying assumptions of a model

The chance for winning with strategy "change" is $2/3$ and the opposite (conservative) strategy has only a winning probability of $1/3$. These calculations are valid only under certain assumptions but this is a longer story, see Vancsó & Wickmann (1999). Here it 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. This Monty Hall problem was analysed from a psychological point of view by Krauss & Wang (2003). One of their results is the following: Players, who have played the moderator as well, are significantly better than players who have not taken this role.

We repeated this experiment with our students with a similar result. It means that changing the point of view is very important in mathematics. The favourable relation helps to understand such situations and to explain to other people how the paradox rises from a false symmetry expectation. Our students without exception understood this paradox and could to explain it to other students or friends or relatives at home. They remarked on the power of the psychological experiment. If they could not convince their "partners" then they offered a game, which illustrates the original question. Of course sometimes they could not convince the partner of "their truth".

Lotteries – the problem of the unknown number of balls

There are different lotteries in European countries; the numbers of balls in the box vary and so does the number of balls drawn at each lottery. For example, there are three different lotteries only in Hungary (Vancsó 2006); two new types were introduced in the 1990's – Hungarian players would know but not tourists:

- The oldest (A) contains 90 balls in the urn from which 5 are chosen.
- In (B), 6 balls are drawn out of 45.
- In (C) named Scandinavian, they select 7 out of 35 balls.

Consider the situation of a tourist in Hungary who does not know how many balls there are in each lottery. His question is the following: he knows the numbers of the balls drawn in one week. He has to estimate the number of balls in the box from which the numbered balls are drawn. There are two different possibilities:

- To derive a classical confidence interval for this number of balls from the data. We use the maximal number estimation in the following; there are different estimators but this one is unbiased and efficient for the total number of balls. Note, that the classical solution uses no extra information; such information could not be incorporated even if it existed.
- To calculate first the posterior distribution of this number from the combined information of the known drawing and the prior distribution on this number; from this posterior distribution that region could be derived which is known as the Bayesian region of highest density (RHD are also referred to as credible intervals or credible regions in literature, see the glossary; for technical details see e. g. Wickmann 2006).

Some aspects of students' work dealing with theoretically interesting and challenging questions will be outlined in the following. The general application of classical and Bayesian methods in parallel in the project work is contained in Appendix B, or also in an EXCEL file.

Differences between classical and Bayesian solutions

The classical solution is unique once one has decided which statistic to use – and there are optimality criteria of efficiency for example to help this choice. The Bayesian method, however, gives different results under different circumstances, i.e. under different prior information on the total number of balls in the urn. For example, if there is information that there can not be more than one hundred balls in the urn, this information changes completely the situation and the results, which may be derived. As the prior distribution on the total number of balls, a uniform distribution may be chosen on the interval from the given maximum number of the drawing to the presumed maximum number of the balls.

Of course there are other possibilities with good reasons. It may be supposed that the total number of all balls is a "special number" like: 80, 90, or 100. It may also be that it has a special character as square number like 81, 100, 121, or might consist of the same digits like 88, 99, or 111. In that case, such numbers would have a higher probability than others. These non-uniform distributions as prior distribution may be used and of course would yield a different posterior distribution on the total number and a different Bayesian RHD interval for this total number.

It is of interest 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 Bayesian RHD interval provided we use uniform prior distributions for the total number.

In the case of the oldest Hungarian lottery (A) we have had more than 2700 drawings since its introduction in the year 1957. We are able to control our result using the actual statistics of these 41 years for both methods. That means we derive a confidence interval from the data for all weeks and check how many times this interval contains the total number of balls, which is 90. For every week we also derive a Bayesian RHD interval based on a uniform prior distribution; we check how many times this interval contains the number 90. For the detailed results of this analysis see Vancsó (2004).

Extensions and other contexts

In the last few years we obtained more material, and introduced the different errors of classical inference as well. Earlier only hypothesis tests and confidence intervals were used from classical theory. We thought these were sufficient to understand the character of the two different approaches. We became more practised and used time more efficiently which led to more content.

The example about lotteries is just one out of many contexts. Other topics covered in this second semester are exit polls, the fair coin or dice problem, or testing of experts. Recently, betting in connection to sports events has become more popular. This betting situation is paradigmatic for the Bayesian approach; it was extensively used and analyzed by de Finetti who is one of the prime Bayesians. Gáspár (2006) wrote a diploma thesis on the betting context; he is now an employee of a big online betting company in Hungary. In the thesis, he uses a special technique that is typical in this betting situation exploiting prior information to estimate the initial odds for betting (later the odds are usually adapted to the stakes put by bettors).

5. EVALUATION OF THE PROJECT

There are different methods to evaluate a curricular project like ours. One important criterion is the soundness of the approach philosophically and mathematically. We have elaborated such issues in the paper. Another possibility would be the success of the students; a

further criterion is how teaching is accepted and how students feel that they understand the concepts after the course.

Success of students

The marks in examination papers have improved over the years as compared to earlier times when the students primarily had a mathematically oriented course in probability. Also, the acceptance by students increased as measured by numbers of students who chose the seminar, which was not compulsory for them.

There have been many diploma theses emerging from these seminars; some students also got well-paid positions outside the school-system. This has to be compared to the fact that probability earlier has been a subject of minor importance and has been avoided by its poor reputation to be very demanding.

One former student, who has participated in the 2004/05 seminars, is going to write a doctorial thesis with the title "Modelling in statistics". In his teaching at secondary school, he experimented with ideas and methods emerging from the "parallel" seminar. His students reached about 20% higher scores at the final examination in statistics and probability than the average of secondary schools in Budapest.

This result supports our research hypothesis that a deeper insight in theories improves later expertise in teaching. In what follows, some evidence of the outcome of the approach is given by students' opinions and concrete materials made by them during the course.

Self-reports of students

All in all, the students always found this way of teaching useful and in the end of the course, they wrote interesting essay questions and their solutions. They found the idea of involving subjectivity into mathematics very surprising. One of them wrote an essay using a famous book of Pólya about heuristics. Pólya (1954) introduced subjective probability as a measure of our conviction, and dealt with mathematical problems, which are not easy to prove but some correct consequences of the theorem are known. He shows how these facts may increase the probability of the truth of the theorem in question. He does not explicitly denote his arguments by the word favourable but uses the favourable relation over 60 pages of the book to solve and describe problems.

The students find this way of learning very useful and comment that it enables them to explore ideas more deeply, which they think is important for their later role as teachers. These statements are well documented by students' essays written after the courses in the last few years. Some extracts of their reactions are cited below. These opinions demonstrate a definite advantage of the parallel approach for their conceptual progress.

"I never thought that the degree of my certainty can be handled mathematically and be measured by probabilities. It was very impressive for me and I now understand better what happens when mathematicians work on a new theorem." (Student 2003)

"I constructed three boxes from cardboard and always hide a little plastic cat from view in one of the boxes. With this apparatus, I played Monty Hall with my friends. After many experiments, they understood what really is going on; they accepted the mathematically optimal strategy more easily when they played the moderator themselves." (Student 2006)

The effect of changing the role on the decision chosen and on the acceptance of theoretical results is confirmed by Wang & Kraus (2003). In Bristol, during an exchange programme, students from Budapest showed the Monty Hall problem; the above mentioned construction was standard part of probabilistic heuristics.

"I saw why for the first time. For some time, I felt cross with the Simpson paradox. Now, with the new concepts, it is a very good feeling. I like such "aha" experience where I understand my own brain better. I thought stochastics to be a very strange field of mathematics; or, better to say not really a part of mathematics. This course showed me how the unexpected results in statistics or probability come into being and get a better insight into probabilistic thinking." (Student in 2007)

"I understood at least that it is the very strongly imprinted causal thinking (cause and effect) and its asymmetry that makes the favourable relation so paradoxical; situations where only stochastic connections exist are symmetrical, which breaks causal rules." (Student in 2007)

These remarks express very clearly one main problem in stochastics, namely that it supposes a type of logic, which is different from the classical one. Understanding this fact, students can get more familiar in different modern topics of physics or biology e. g. quantum physics or genetics. There are remarks from students who studied physics as well. They show that their thinking is more flexible as they do not find such phenomena to be thus mystical as they have seen such paradoxical situations previously. These efforts do not belong to the main stream

of our course therefore such documents were only "collected" from personal discussions with the students and not by a specially designed questionnaire.

The interpretation of confidence intervals

Confidence intervals are open to indirect interpretations only. We all know about the difficulties of interpreting results gained by this method properly – not least from the examination papers. Some more citations of students illustrate our thesis about the positive effects of parallel teaching of the two statistical concepts – here with respect to confidence intervals:

> "I understood the confidence interval only after I had become more familiar with the Bayesian region of highest density." (Student in 2004)

> "I always interpreted 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)

This misunderstanding is common and can also be remedied by other (more classical) ways. The main point is to understand the confidence interval as a random variable and not the parameter (at least in the classical approach). From the applications, there is an urgent need in such an interpretation of an interval containing the unknown parameter with a pre-assigned probability. However, the classical approach does not provide it – contrary to what it "promises". This misleading promise prompts so many students to interpret the procedure of confidence intervals wrongly (see e. g., Gigerenzer 1993).

> "I really like the Bayesian method because I saw for the first time why people have different opinions in many cases. Because different people may have different prior distributions." (Student in 2007).

About the lottery problem, two students have initiated an interesting project. One of them analyzed the oldest lottery (5 chosen numbered balls) and found an interesting connection. Her result is shown in Table 1.

M denotes that number, which according to the prior information is the largest possible number for total of balls; and the figures in the table are the upper bound of the 0.95 Bayesian RHD intervals; the upper bound of the classical 0.95 confidence interval is shown in the first row. It demonstrates that classical and Bayesian intervals are numerically not equal in the case of "zero information" (uniform prior distribution). Note that if the maximal number of balls according to

the prior information is less than 100, Bayesian intervals are more precise than classical intervals.

Maximum number of the week		$x_5 = 55$	$x_5 = 61$	$x_5 = 79$	$x_5 = 85$
Upper bound of confidence interval		99	110	143	155
Upper bound of Bayesian RHD Prior maximal number of total balls M	100	90	93	98	99
	150	106	115	133	137
	250	112	124	158	168
	1.000	113	126	164	176
	$\to \infty$	115	128	166	179

Table 1. Comparison of classical and Bayesian (R H D) intervals for the total number of balls.

It is an interesting question whether there is such a number M for which confidence and Bayesian intervals are numerically equal. As mentioned earlier, there are different prior distributions for example assuming higher probabilities for special numbers. The uniform distribution is the best way to express the status of having no information and we see from Table 1 that the classical and Bayesian solution could be quite similar if the prior information reduces the total number by say 120, which is quite reasonable in view of our knowledge about lotteries.

Another student experimented with lottery B of 6 from 45. The results of the weeks inspected by her were 34, 35, 42 and 45 with respect to the maximum of the drawn numbers. She compares the classical 0.95 confidence interval to the Bayesian interval based on a prior uniform distribution up to M and investigates the development of the Bayesian interval when M tends to infinity If M is less than 100 and a uniform prior distribution is used, then the Bayesian RHD produces a smaller interval than the classical confidence interval.

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

Table 2. Classical and Bayesian intervals for the 6 from 45 lottery − M = prior maximal number of total balls.

It is interesting to compare the results of both methods to each other as was done in the two students' projects. There is enough data from actual drawings of the lotteries to check the success of the confidence interval method. The Bayesian RHD interval can be checked as well from case to case depending on the chosen prior distribution. According to the expressed didactical thesis about the positive impact of diversity, it may be postulated that: if somebody knows different constructions for solving a problem, he or she might understand each of the methods and the problem and the solutions 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 present both interpretations perfectly. She was very convincing and the rest of the students participating at the seminar eagerly followed and understood her. It is pity that this presentation was not video-taped for posterity.

In 2004, we posed a test on the Internet about the interpretation of confidence intervals. We asked only such people who had already studied at least 3 years of mathematics. There are only two correct answers out of 89. In contrast to that bad performance, the students of the last two seminars reached very good results with only two false answers out of 31. The question was posed in Germany as well (for the results, see Gigerenzer & Kraus 2001, p. 51).

6. CONCLUSIONS AND FUTURE PLANS

We tried to carry out such didactical principles which are general enough to serve as a basis for teaching inferential statistics. One of the important ideas is to compare and contrast new concepts with each other right from the beginning. Confidence intervals may be better understood if the Bayesian interval of highest density is also introduced and contrasted to it. Our experience supports this principle which is substantiated by students' work and interviews as well. Students found it important to understand the notion of conditional probability and manipulate it.

Bayes' theorem plays a minor role in the classical approach where it is like a foreign particle which often causes confusion as it invokes other perceptions which do not fit to the chosen framework. However, this theorem plays a central role in Bayesian inference where it is conceptually well integrated. It deals with the question how our "knowledge" develops about uncertain things if we get new information. That has been frequently a cornerstone of students' opinion. An important note is that the historical personage of Reverend Thomas Bayes himself was not a "Bayesian", he did not think about subjective probability.

- Success and feedback from students show that this parallel approach can be a good basis for teacher education at university to study inferential statistics. The students' knowledge became more reflective and conscious about the problematic issues in statistical inference. The different interpretations of probability and their foundation enhance the limitations and true interpretations of classical inferential procedures.

- Theoretical analysis and evaluation of qualitative results of the pilot projects indicate the direction of further refinements of the approach.

- Feedback from students, who have become teachers at school meanwhile, shows that their belief about mathematics and their teaching styles are different from their colleagues. While their colleagues struggle with the subject, our students who are now teachers have success in teaching probability and statistics. This is supported by qualitative personal interviews; a questionnaire is planned to measure this effect quantitatively.

- The next step could be a treatment control group comparison to provide quantitative evidence about the effectiveness of this teaching method. The growing popularity of the "parallel" seminars will make installing a control group more realistic in the near future.

- A book on the detailed ideas and results of our piloting courses is in preparation.

There are some final remarks. This author is not an expert in Bayesian statistics. The conception of the "parallel" courses for teacher students was discussed focusing on their later work. In the last two decades there have been many research studies about teacher beliefs about mathematics. For this topic, too, the experiments presented are very useful: The myth about mathematics as absolute knowledge has been challenged for the students participating in these courses. Our students understood mathematics as a result of our activities and thinking briefly "made by us". (cf. Freudenthal 1973, p. 213, or Lakatos 1976, introduction)

From this point of view, mathematics has been based on historical processes as well. While our notions are suitable to express our experience, they are not absolute. For the same situations there are different notions, such as different concepts of continuity or integrals in calculus. If there are two different approaches for solving a problem, then we can no longer claim our answer to be absolute. Such relativity of truth is at the core of modern mathematics but there are still people including mathematicians who reject these statements (the so-called Platonists).

These courses gave rich opportunities to reflect questions of the philosophy of mathematics, which is very important taking into account that a high percentage of our students will become mathematics teachers and influence the next generations of pupils at school. Advanced mathematics usually has a high priority and prestige for teacher students at our university. Statistics and probability traditionally is less popular but this "parallel" course has changed the situation a little bit in Budapest.

REFERENCES

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

Bertrand's box paradox (n. d.). *Wikipedia, The Free Encyclopedia*. Retrieved from
http://en.wikipedia.org/wiki/Bertrand%27s_box_paradox, June 4, 2009.

Borovcnik, M. (1992). *Stochastik im Wechselspiel von Intuitionen und Mathematik.* Mannheim: Bibliographisches
Institut.

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

Conjunction fallacy (n. d.). *Wikipedia, The Free Encyclopedia*. Retrieved from
http://en.wikipedia.org/wiki/Conjunction_fallacy, June 4, 2009.

Darling, D. (n. d.) Bertrand's boxes. Paradoxes. *The Internet Encyclopedia of Sciences*. Retrieved from
http://www.daviddarling.info/encyclopedia/P/paradoxes.html, June 4, 2009.

Everything2 (n. d.). Bertrand's boxes. *Everything2 is a collection of user-submitted writings*. Retrieved from
http://everything2.com/title/Bertrand%2527s%2520Box%2520Paradox, June 4, 2009.

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

Freudenthal, H. (1973). *Mathematics as educational task*. Dordrecht: D. Reidel Publishing Company.

Gardner, M. (1959). The prisoner's dilemma. 'Mathematical games' column. *Scientific American*, Oct. 1959, 180–182.

Gáspár, E. (2006). *Betting as a topic using Bayesian methods*. Budapest: Unpublished diploma thesis, ELTE University
of Budapest (Hungarian).

Gigerenzer, G. (1993). The Superego, the Ego and the Id in Statistical Reasoning. In G. Keren & C. Lewis (Eds.), *A
Handbook for Data Analysis in the Behavioral Sciences* (pp. 311–339). Hillsdale, NY: Lawrence Erlbaum.

Gigerenzer, G. & Krauss, S. (2001). Statistisches Denken oder statistische Rituale? Was sollte man unterrichten? In: M.
Borovcnik, J. Engel, & D. Wickmann (Eds.), *Anregungen zum Stochastikunterricht: Die NCTM-Standards
2000 – Klassische und Bayessche Sichtweise im Vergleich* (pp. 53–62). Hildesheim: Franzbecker.

Isaac, R. (1995). *The Pleasure of Probability*. New York: Springer.

Krauss, S. & Wang, X. T. (2003). The Psychology of Monty Hall Problem: Discovering Psychological Mechanism for
Solving a Tenacious Brain Teaser. *Journal of Experimental Psychology: General*, 132, No. 1, 3–22.

Lakatos, I. (1976). *Proofs and refutations*. Cambridge: Cambridge University Press.

Malinas, G. & Bigelow, J. (2004). Simpson's paradox. *Stanford Encyclopedia of Philosophy*. Retrieved from
http://plato.stanford.edu/entries/paradox-simpson, June 4, 2009.

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

Monty Hall (n. d.). *Wikipedia, The Free Encyclopedia*. Retrieved from http://en.wikipedia.org/wiki/Monty_Hall, June 4,
2009.

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

Morrell, C. H. (1999). Simpson's Paradox: An Example From a Longitudinal Study in South Africa. *Journal of Statistics Education,* 7(3). Retrieved from
http://www.amstat.org/PUBLICATIONS/JSE/secure/v7n3/datasets.morrell.cfm, June 4, 2009.

Mosteller, F. (1987). The Prisoner's Dilemma. Problem 13 in F. Mosteller, *Fifty Challenging Problems in Probability with Solutions* (pp. 4 and 14–15). New York: Dover.

Pólya, G. (1954). *Mathematics and plausible Reasoning Vol. II*. Princeton: Princeton University Press.

Székely, G. (1986). *Paradoxes in probability theory and mathematical statistics*. Dordrecht: Reidel-Kluwer.

Tversky, A. & Kahneman, 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. In *Beiträge zum Mathematikunterricht 2004* (pp. 621–625). Hildesheim: Franzbecker.

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

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

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

Varga T. (1973). *Lass uns spielen Mathematik,* I.–II. Leipzig-Jena-Berlin: Urania Verlag.

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

Weisstein, E. W. (n. d.). Monty Hall Problem. Retrieved from *MathWorld – A Wolfram Web Resource*
http://mathworld.wolfram.com/MontyHallProblem.html, May 25, 2009.

Weisstein, E. W. (n. d.). Prisoner's Dilemma. Retrieved from *MathWorld – A Wolfram Web Resource*
http://mathworld.wolfram.com/PrisonersDilemma.html, May 25, 2009.

Wickmann, D. (1991). *Bayes-Statistik*. Mannheim: Bibliographisches Institut.

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

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

ANNEX

For the annex of all appendices to the paper follow this link. For the EXCEL demonstration follow this direct link.

Author   : Ödön Vancsó

E-mail   : vancso@ludens.elte.hu

Address   : Institute for Mathematics, Eötvös Lóránd University, Budapest, Hungary

APPENDIX A:

SOME PROBABILITY PARADOXA

"The prisoner's dilemma"

This problem was originally formulated by Gardner (1959), and later taken up by several authors, e. g. Mosteller (1987). It was fiercely and widely discussed in the literature. For details, see following description according to Weisstein (n. d.)

"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 prisoner other than himself 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. However, he is mistaken since his chances remain 2/3."

"Monty Hall problem"

This problem became famous as it was part of a popular TV show; the discussion to follow the solution by vos Savant (1990) was signified by remarkable wrong conceptions of the presented situation and a complete misunderstanding of the presented correct solution. It is amazing that a problem as simple as this one was so much disputed among mathematicians and statisticians, not to speak of other well-educated people or novices to the subject. More details may be found from Monty Hall (Wikipedia n. d.). The following description is according to Weisstein (n. d.):

"This problem is named for its similarity to the *Let's Make a Deal* television game show hosted by Monty Hall. The problem 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"

Varga (1976) proposed a nice variant of the elder problem of Bertrand's drawers. Interesting details about the history, or the solution, may be found in Bertrand's box paradox (n. d.), Darling, D. (n. d.), or from Everything2 (n. d.). The advantage of Varga's discs lies in the circumstance that it may easily be performed as an experiment in class. There are three discs marked as in Figure 1.
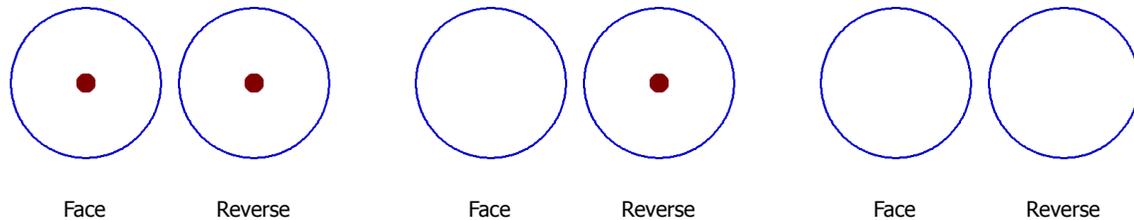

| Face | Reverse | Face | Reverse | Face | Reverse |

Figure 2. 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' (We used two different colours in our experiments). 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.

For illustrative purpose, one class experiment with this game is reported:

A teacher played this game with 10–11 years old children. He summarised his observations briefly. "Some tried to repeat the last result in their prediction each time, others used blank and spot alternately for predicting the next result. None chose the best strategy: whatever is on the face is most likely to be also on the on the reverse side). He then let one child use this strategy and the results showed that he consistently scored best over a range of fifty trials. The children began to think and to suggest reasons as to why this might be. Their thinking was intuitively supported; no one came up with a numerical solution but their answers reflect that they had started to grasp some of the relevant ideas inherent in probability."

APPENDIX B:

EXAMPLES OF STUDENTS' WORK – IN EXCEL

The following examples illustrate students' work. As is typical for the application of Bayesian methods, we had to use software; VisualBayes from Wickmann (2006), or EXCEL. In what follows, we present some graphs together with the problems and methods we used in the projects. Of course, for the paper, the layout has been enhanced.

The problem and various classical and Bayesian methods to deal with it

> In a lottery *n* out of *N*, the number *N* of balls is assumed to be unknown. We draw *n* = 6 balls without replacement from the "urn", the Lotto numbers; the numbers ordered are:
>
> $x_1 < x_2 < .. < x_6$
>
> The problem is how to extract information on the unknown number *N* of balls from the numbers drawn?

- Classical methods for finding the maximal number of balls
  Estimation of the number of balls
  Confidence interval for the number of balls

- Bayesian methods for finding the maximal number of balls
  Updating of prior distribution on the maximal number by the result of *one* week
  Cumulative – week by week – updating of uniform prior

The reader will find more details in the annex or in the EXCEL file where it is also possible to simulate the results of the week and see the influence on the Bayesian result. Here, we will show only a few graphs to illustrate the difference in results between the classical and the Bayesian approach.

Classical estimation of the number *N* of balls

There are several estimators of the unknown number, all with different properties. We refer only to a few:

Median	Mean	MLE	Extreme gaps	Mean gap
$2 \cdot \mathrm{med}\,(x_1,..,x_6)$	$2 \cdot \mathrm{mean}\,(x_1,..,x_6)\text{-}1$	$\max(x_1,..,x_6)$	$\max + \min - 1$	$\max + \dfrac{\max - 6}{6}$

Table 3. Various estimators for the number N of balls.

The data of the lottery since its start are analyzed to show the behaviour of these estimators. For classical estimators, two properties are most relevant: Whether the estimator is unbiased (or correctly centred), and whether it has a small variance, which means that in repetitions of the situation the new estimate would not differ too much from the first estimate.

From the graphs one may see the following. The median estimator has a great variance but is centred correctly; the extreme gaps estimator is better than the median estimator but with respect to the maximum likelihood estimator, it is worse. However, the MLE estimator on the other hand is not unbiased (it is only asymptotically unbiased, which means that the systematic error converges to 0 as the sample size increases to infinity).
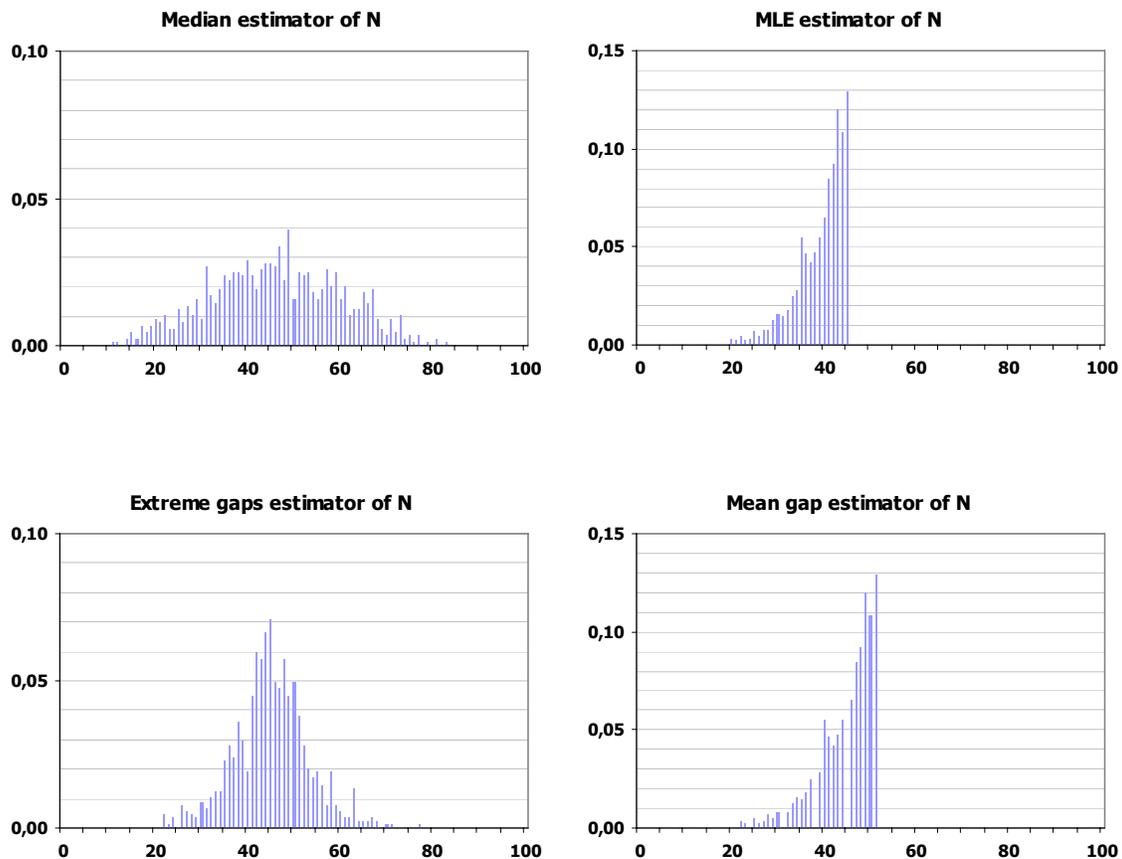

Figure 3. Repeated estimations of the unknown number of balls by various estimators.

Classical confidence intervals

This method yields intervals, which cover the unknown parameter (here the number *N* of balls) with a pre-assigned probability – supposed that it is applied in repeated cases under the *same* conditions. The graph shows the intervals for *N* week by week, calculated from the week's drawn lotto numbers. The global coverage rate is 95.7%. A disadvantage with classical confidence intervals is that it is not easy to integrate the data cumulatively from the past to give one summary confidence interval for *N*.
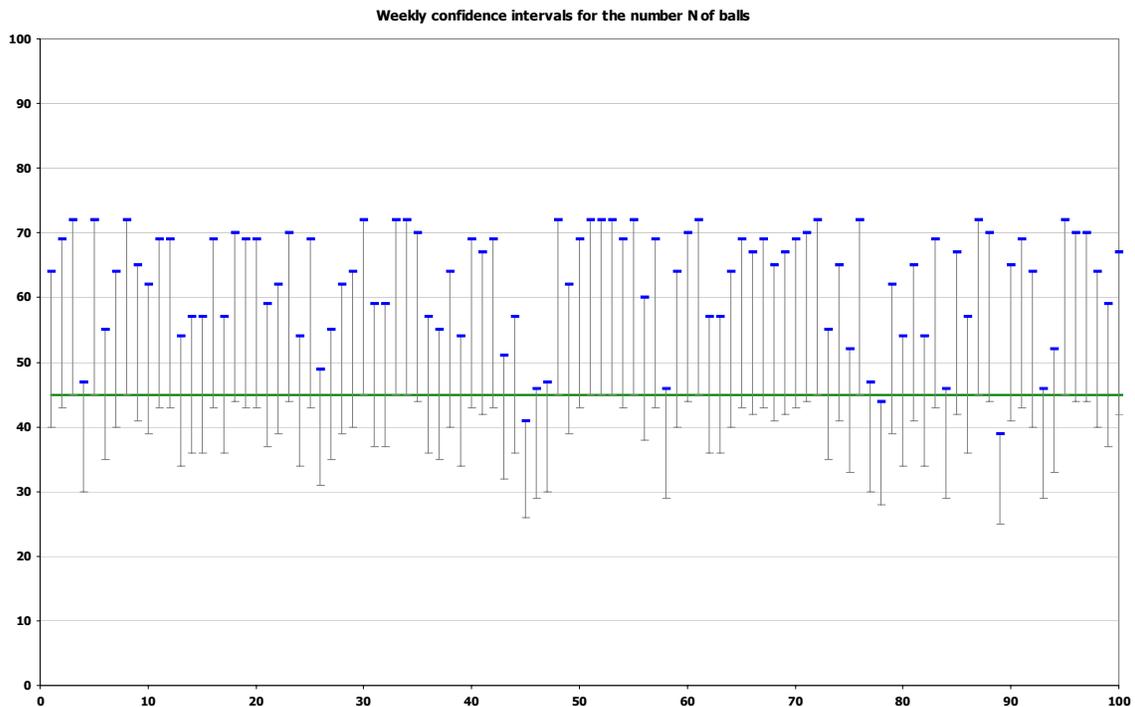

Figure 4. Classical confidence intervals on a weekly basis – the line in the middle represents the "true" value.

Bayesian methods for finding the maximal number of balls

For a Bayesian solution, it is necessary to model the prior knowledge by a distribution. Here, "complete" ignorance of this number *N* will be modelled by a uniform distribution on the interval [31, 80]. This prior is updated by the results of one week to a new posterior distribution on *N* reflecting the information of the data of one week. This new status of knowledge on the maximal numbers of balls is calculated and graphically presented.

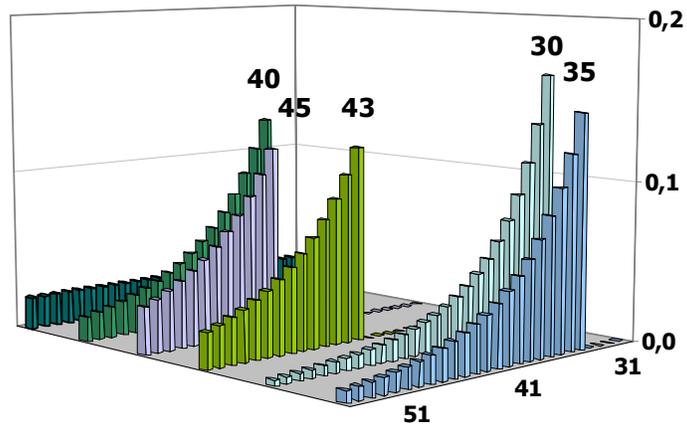**Updating of uniform prior in dependence of week's maximum -detail**

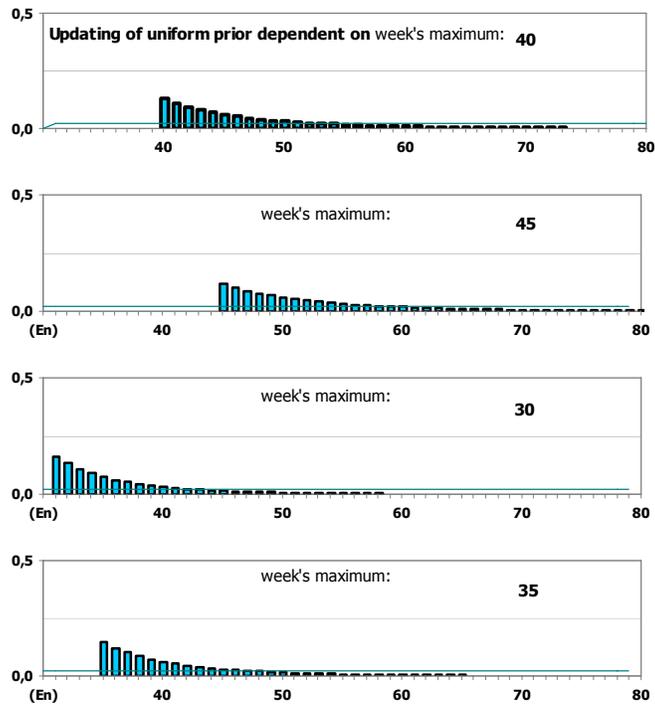Figure 5a. The posterior distribution of the number balls is dependent on the week's results.

Figure 5a. Another representation of the posterior distribution of the number balls.

Within the Bayesian framework, the posterior distribution of the number of balls after one week may serve as prior distribution for the next week. Thus, a continuous process of updating may be applied. The following graphs (Figures 6a and b) show the resulting learning process: after 9 weeks, the initial ignorance on the interval [31, 80] – modelled as uniform distribution – has been changed to a distribution, which is reduced to the values between 45 and 48, all other values have already a negligible probability at that stage. After 30 weeks, this posterior distribution expresses a status of knowledge that 45 has a high probability, while the slight "risk" still to have 46 or more balls is virtually negligible.
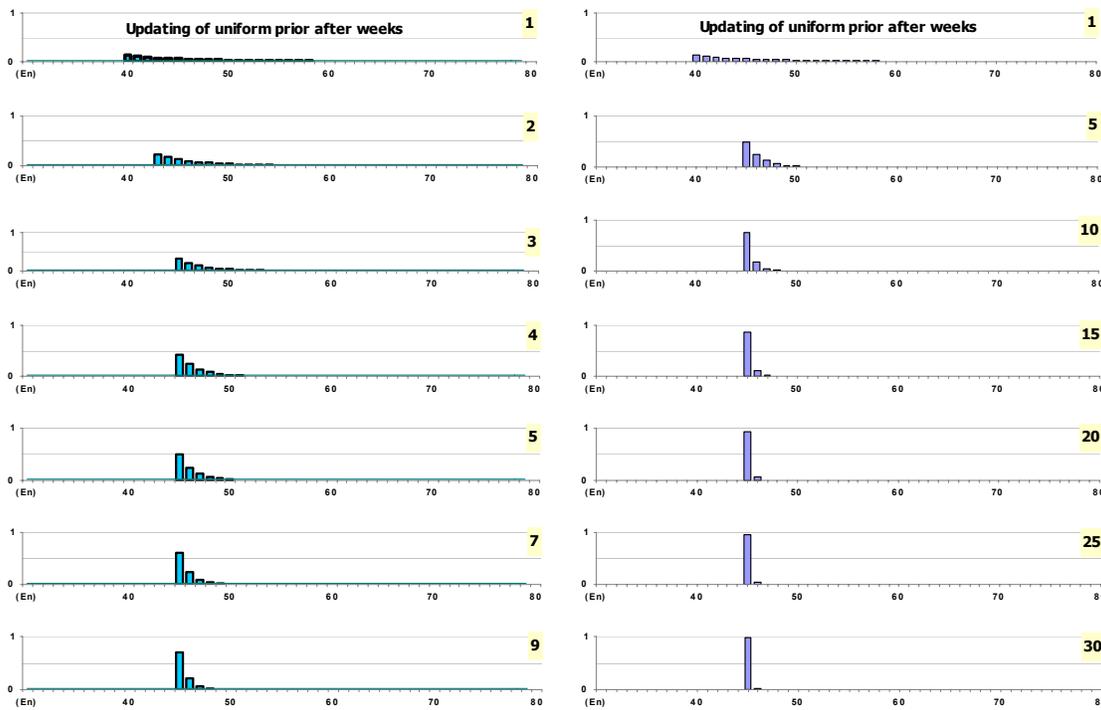

Fig. 6 a. After 9 weeks,
only 45–48 bear a non-negligible probability.

Figure 6b. After 30 weeks,
only 45 is relevant; there is a small probability for 46.

The graphs in Figure 6a and b clearly show the convergence of the posterior probability distribution with time. In fact, the true number N equals 45; we deal here with the 6 out of 45 Hungarian lotto. The repeated updating accumulates all information from the past and yields a present status of information about the unknown parameter.

Ordered numbers of successive draws

Nr	x_1	x_2	x_3	x_4	x_5	x_6
1	6	15	20	24	38	40
2	9	14	23	32	40	43
3	1	7	14	43	44	45
4	6	9	11	14	18	30
5	12	13	19	30	34	45
6	16	17	19	20	34	35
7	4	23	26	35	39	40
8	11	32	40	41	42	45
9	1	5	11	25	40	41
10	3	27	33	34	35	39
11	7	12	20	32	38	43
12	5	21	29	33	39	43
887	9	10	11	14	25	28
888	3	6	10	17	38	43
889	6	9	10	24	37	38
890	3	16	24	26	36	40

Table 4. Data of the Hungarian six-numbers lotto since its start

The whole data on the lottery since its beginning is contained in the EXCEL file where the reader may find also data on the 5 out of 90 lotto in Hungary. Furthermore, the file contains a detailed analysis of the problem by classical and Bayesian methods done as part of the project by the students. Here, we focused on a presentation of the methods used to solve the problem.