CONDITIONAL PROBABILITY, BAYES AND CLASSICAL STATISTICS—EVALUATION OF THE PLANNED SECONDARY-SCHOOL REFORM IN HUNGARY

Ödön Vancsó[1], Péter Fejes Tóth[2], and Manfred Borovcnik[3]
[1] Eötvös-Lóránd University, Centre for Didactics of Mathematics, Budapest
[2] Hungarian University of Agriculture and Life Sciences, Department Applied Statistics, Budapest
[3] University of Klagenfurt, Institute of Statistics, Klagenfurt
vancso.odon@ttk.elte.hu

Our main goal is to show that we can use classical and Bayesian statistics, including theoretical background, for introducing statistical inference in Hungarian high schools. In a first step, we will test the material we are developing with higher-level students. We introduce the concept of the project, including its theoretical background and previous pilot teaching experiments. The basis is the parallel introduction of classical and Bayesian methods of statistical inference. Long-term experience from research work and from seminars with future teachers provide evidence that this parallel approach is useful. Our framework is from the project "Guided Discovery in Mathematics Education," that can be classified in different inquiry-based education forms (see Artigue and Blomhøj, 2013).

INTRODUCTION: STATISTICS TEACHING IN HUNGARIAN SCHOOL MATHEMATICS

Tamás Varga started his "Complex Mathematics Education" in the early 1960's, and in 1978 a new curriculum for all primary school levels (grades 1–8) was introduced. It was the first time that statistics was included in Hungarian schools. Varga's idea was signified by a symbiosis between statistics and probability and a belief that teaching should begin as early as possible: the topics should be introduced in parallel, shape notions spirally, and strengthen the connections between them (Varga, 1972). The first step was experimental: active learning and learning-by-doing that included experiments, hands on activities and calculations, and drawing conclusions, with students working as independently as possible. Generalised and abstract issues resumed in later phases according to the developmental level of students' thinking. More details about Varga's approach can be found in Varga (1983) and Borovcnik (2020). This reform of 1978 included not only arithmetic, algebraic, and topological (geometric) structures but also probabilistic and statistical structures. Varga's approach of guided discovery and his holistic ideas about teaching were a great challenge to the teachers. As a result, many teachers (and parents) rejected his curricula. Teachers at the time were neither well prepared nor well supported to form the backbone of the curricular reform, though there was a broad group of teachers that tried to realise his ideas. The reform failed, and only a few elements remained in their original form. Yet, Tibor Nemetz was convinced that Varga's ideas in the field of stochastics should be continued in secondary schools. He worked on further reform that resumed elements of the original plan. Nemetz used many key ideas from Varga such as the central role of playing for learning (Varga 1970), the spiral shaping and reshaping of concepts, and the tight connection between probability and statistics. At that time, only simple calculators were used in schools. Nemetz introduced many experiments, in which the students had to estimate probabilities. The experiments had to be designed, performed, and analysed so that students got answers if they carefully followed steps. This type of task is highly suitable for introducing the law of large numbers in an experimental way and provides a good example for comparing statistical and classical probabilities with each other. Nemetz collected 13+1 experiments for students to complete in class; the results of students' realisation of the experiments were evaluated and described. This was organised in the form of a competition, and the winner was rewarded. Among these experiments were some that could be understood by combinatorial methods, whereas others required statistical methods (Nemetz, 1985).

Consider two concrete examples from Nemetz in which students must bet on one of three possible outcomes A, B, or C. In the first task, one can use a combinatorial model, but the second task is genuinely statistical; the classical notion does not help to obtain a good choice. (a) Toss five regular dice at the same time. The outcome is the sum, S, of the numbers that appear on top after the dice land. What is the most probable event? A: $S \le 9$; B: $10 \le S \le 17$; C: $S > 17$. (b) Open a minimum 200-page book to a random page and choose the 10th word from the right side! The outcome is the number, N, of letters in the word. What is the most frequent event? A: $N \le 3$; B: $N = 4$, 5, or 6; C: $N \ge 7$.

In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering & Educating Today's Learners in Statistics. Proceedings of the 11th International Conference on Teaching Statistics (ICOTS11 2022), Rosario, Argentina.* International Association for Statistical Education. iase-web.org

We are designing curriculum materials that build on the rich work from Varga and Nemetz. Our research focuses on whether it is possible to use a parallel introduction of classical and Bayesian inference in high school and how we can prepare teachers to teach such an approach.

THEORETICAL BACKGROUND OF THE PLANNED REFORM

Preliminaries to the Research Question: Parallel Teaching Classical and Bayesian Ways of Inference

Vancsó started his research on inference with a new concept of *teaching Bayesian and classical inference in parallel* in high schools and in teacher training in the early 2000's. He connected to the school experiments of Riemer (1991) and the work of Wickmann (2001) on Bayesian methods and his critique of the significance test. See also Borovcnik et al. (2001) about the forerunners of such ideas. The Bayesian way was introduced by conditional probabilities using real-life problems such as medical tests. Doubletree diagrams with counts were used to represent the problems and were included in a new textbook (at that time) in Hungary as well (see Vancsó, 2010).

As a reminder, in all classical statistics tests, there are two types of errors that one might easily recognise within the context of diagnosis. On one hand, someone may have the disease under scrutiny, but the diagnosis is *false* negative. On the other hand, a person does not have the specific disease, but the test is *false* positive. Both cases may occur, but the *main goal* is to isolate the people that have the disease (which is important if the disease is contagious). It seems to be impossible to keep both errors at a low level. Consider that the real goal is to keep the first type of error at a low level of 1% (or less). In this context, we see different probability notions at work. We use *statistical* probability for different failures of the test, *subjective* probability to estimate the size of the diseased (infected) group (or to analyse the individual case of the tested person), and *classical* (Laplacian) probability to answer the question. The answer is that the tested person in fact has the disease (is infected) if the person has a positive test—formally P(infected | +Test). From a mathematical point of view, we consider all three meanings as probability because they fulfil Kolmogorov's axioms. Vancsó (2009) includes a summary of the above-mentioned "parallel-teaching research" and an analysis of the reactions of the future teachers. They understood classical inference better after studying Bayesian inference.

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

These statements may convince the reader that a war between classical and Bayesian statisticians, as expressed in several articles in the Teacher Corner (see, e.g., Berry, 1997), is futile. However, the didactical community should seek viable ways to teach both theories in parallel; Migon and Gamerman (1999) provide an early exemplar of such a parallel approach at university level.

A newer idea is to introduce problem sequences where these three different probability roots are set in a parallel discourse (Varga &Vancsó, 2021). Different games of chance have played a key role in the emergence of probability before the theory was established. For example, Rényi (1972) wrote *Letters on Probability* with an imaginary exchange between Pascal and Fermat. Later in the 1930, de Finetti (1937) developed an axiomatic theory to justify subjective probabilities as playing a role in bets on football matches. Because of its popularity, we have chosen betting as starting point for our series of problems. See Vancsó and Varga (2020) for the structure of the tasks, including the probability aspects. We develop probability and statistics simultaneously using the concept pairs of relative frequencies and probability and mean and expected value. In both pairs, the law of large numbers, which goes back to Bernoulli (1712), connects the pairs to a complementary total. Another reason is that all different meanings for the probability notion in mathematics are equally important. Consequently, shaping an understanding of probability should use all roots. Expected value—a further central notion—relates to expected win, which is also resumed in our problem of betting.

Finally, modelling—a key in mathematics education—is also included in the design of our teaching material. De Finetti (1937, 1974) was the first to introduce subjective probability in a formal way. Carranza and Kuzniak (2008) advocate a simultaneous use of frequentist and Bayesian probability. They wrote, "Since the end of the 17th century, the 'taming of chance' (Hacking, 1990) has been following two paths: Frequentist and Bayesian" (p. 1); "there is a failure to differentiate between the two interpretations of probability, which are spontaneously combined in statistical situations where they are applied simultaneously" (p. 1). We also use a third concept, the classical (Laplacian) probability in modelling sports events by games of chance. All these sources lead us to introduce the Bayesian method in parallel with classical inference. The results of the first school experiment 2019–2020 were reported in Borovcnik, Fejes-Tóth, Jánvári, and Vancsó (2020) and are summarised in Vancsó, Borovcnik, and Fejes-Tóth (2021). Five teachers participated in the teaching experiment. Three hours was not enough time to communicate the ideas, and teachers could only focus on the classical way. Our research group will elaborate a longer preparation course for teachers to experiment this new way in schools where mathematics is taught at a higher level than usual (5–7 lessons per week). Higher level is similar to A level; it was introduced to replace the university entrance exam for such students who study further mathematics at university level. The main goal is to prepare new requirements in inferential statistics for the final exam for higher-level students before a final decision about it in 2025. Our arguments in favour of the parallel concept originate from long-term research.

Reasons Behind the Planned Concept of Parallel Teaching

In the following, we summarise the strengths of this approach. Classical notions of probability based on the notion of equally likely outcomes do not provide a foundation for later work in probability when events are not equally likely. Nemetz used such experimental tasks where the classical notion fails. Although the frequency approach to teaching probability is helpful in situations where students can perform random experiments, there are conceptual difficulties in distinguishing between the observed relative frequency and the actual probability of an outcome, which relates to an infinite sequence of experiments. This approach is accessible to less able children because it is based on the comparison of likelihoods rather than on the specification of fractions. Gigerenzer comes to a similar conclusion and advocates using counts instead of fractions. Hawkins and Kapadia (1984) favour the Bayesian way because subjective probability is “closer to the intuitions that [the students] try to apply in formal probability situations”; a focus on frequentist or classical notions “may well conflict with the children’s expectations and intuitions” (p. 372) so that the three approaches should be blended. Albert (2002) provides arguments in favour of Bayes methods: Bayesian thinking is more intuitive than the frequentist probability viewpoint and reflects common-sense thinking about uncertainty that students have before the statistics class. Students have dealt with uncertainty and use words such as “likely,” “possible,” “rare,” and “always” to reflect different degrees of uncertainty. Subjective probability is a way of assigning numbers to these different degrees of uncertainty using a scale from 0 to 1. Beliefs of a person can change as one obtains new information. As one obtains new information or data, a person’s subjective probability can change. Bayes’ rule is a recipe for determining how the (conditional) probabilities change in light of new evidence.

The Bayesian paradigm reflects learning, where one has initial beliefs about some issue; an experiment provides data; and the new beliefs blend previous opinions with information obtained in the experiment (Berry, 1997). In an estimation problem, one typically wants to be confident that a computed interval estimate contains the parameter of interest. In a testing problem, one is interested in the probability that a particular hypothesis is true. Yet, in the frequentist viewpoint towards inference, one is confident only in the *performance* of the estimate or hypothesis test in repeated sampling. This is useful in situations where one is repeatedly faced with an identical situation and performs a large number of 95% confidence intervals or hypothesis tests of level 0.05, but that does not help the applied statistician who is interested in making a one-off inference based on a single dataset. In contrast, Bayesian statements are conditional on the observed data. Because parameters are viewed as random from a Bayesian perspective, it makes sense to say that, for example, the interval (0.04, 0.54) covers an unknown proportion p with a certain probability. Likewise, one may talk about the probability that the null hypothesis applies after one sees the data. Students and even scientists make the incorrect statement that a proportion p falls into a realised 95% confidence interval with probability 0.95 (Gigerenzer, 1993, Vancsó, 2009); from a frequentist perspective, the probability is either 0 or 1. Either p is in the interval or not.

Borovcnik and Kapadia (2015) differentiated four key methods for inference, which differ according to the type of probability (frequentist or subjectivist) and the general perception of scientific truth (hypothesis, unknown statements, data as central focus of evidence): (a) Bayesian inference (BI); (b) decision-theory (DT); (c) classical inference (Neyman-Pearson; Fisher); and (d) resampling for informal inference (II). The crucial point is only between BI and non-BI. Our starting point is the parallel introduction. BI involves an interpretation of probability as a frequentist (FQT) or as a subjectivist (SJT) concept (SJT relates to the preference system of a person). Unfortunately, the key concepts to describe the properties of classical statistics are commonly misunderstood (not only by students) as probabilities for hypotheses: (a) the coverage of confidence intervals is linked to the single confidence interval even though it is a property of two random variables covering the unknown parameter; and (b) the type-I error of a test, or the Fisher p value are probabilities for (future) observations; yet, they are misperceived as probability that the null hypothesis is "true" given the data.

Our Planned Work

The curricular reform for grades 11–12 must be completed by 2025. Until then, curricular changes and requirements for the final exam must be elaborated in detail. Statistical inference is planned for the last two years of high school (grades 11 and 12, age 16–19). In the experiment, we will use *three different versions*: *A* is normal, 3+2 (extra) lessons; *B* is high-level students, 5–6 lessons per week; and *C* is special classes, 7–8 lessons per week. We plan to work cyclically. The first ideas will be tested in school experiments with specially prepared teachers. The duration of the experiments is 6, 8–10, and 12–14 experimental lessons for *A*, *B,* and *C,* respectively. Observations during pilot teaching and impact on students and teachers will be analysed. Results from analyses will result in modifying ideas for a new experimental teaching phase to be prepared. Two or three such cycles will provide enough expertise to formulate the planned curricular changes and new requirements.

SOME RESULTS OF THE PILOT EXPERIMENT

We prepared teachers to teach the experimental material for four lessons (45 minutes). For the description of the experiment, see Fejes-Tóth (2020). One conclusion was that teachers taught only the "chi-squared test" and classical inference because they needed more support for Bayesian methods. With respect to other inferential methods, Fejes-Tóth reported positive reactions from both students and involved teachers. We used Excel in these lessons; however, to fill the gap with Bayesian methods, we plan to use Cogstat (www.cogstat.org), which was developed by the cognitive psychologist Krajcsi with a focus on algorithmic thinking (Krajcsi et al., 2021). Because Cogstat is designed for university psychology students, we plan to adapt, simplify, and extend the software so that high-school students can use it. CogStat is built to choose analyses automatically and to provide optimised graphical support for enhancing methods—features that help both students and teachers to focus on conceptual and practical understanding of statistics rather than the technical details of the software. It is appropriate for students who use statistics without deeper mathematical background.

CONCLUSION

Before our next planned experiment, we will develop detailed preparatory materials for use with the experimental teachers to develop deeper understanding of both classical and the Bayesian methods. We learned from our long-term seminars with future teachers that parallel teaching can help to develop understanding of inference concepts. While developing the tasks, we consider that teachers and students are familiar with binomial and hypergeometric distributions, so we focus on tasks that can be solved using these distributions. One example that could also be analysed in a Bayesian way is: An individual suspects that the probability of rolling a six with a die (in a casino) is greater than 1/6. We test the die and roll it 100 times to confirm or reject our assumption. Formulate the problem by stating a null and an alternative hypothesis. Calculate the critical range for a 0.05 (0.01) significance level. How would you decide if the number six comes up 10, 20, or 30 times out of 100 rolls? This question can be generalised later because the die can also be analysed with the chi-squared test for the distribution of all outcomes as we had done in the pilot experiment.

Modelling plays a key role in our introduction to inferential statistics, similarly to the way we treat different notions of probability in teaching (see the loaded-dice experiment in Fejes-Tóth, 2020). The notion of probability must stand in the centre of didactical interventions, including its historic roots. For the foundation of a Bayesian view of inferential statistics, we must necessarily use conditional probabilities and Bayes theorem. We will deal with it within a context and pursue the attitude of modelling. From Borovcnik (2019a, 2019b), the medical diagnosis may be seen as a context for introducing decision-making and statistical tests, deeper discussion about these ideas, and the influence of different probability notions. We plan many opportunities for teacher reflection to enhance their understanding of foundations for our parallel concept of statistical and Bayesian inference. One important aspect is to understand different interpretations of the results in classical and Bayesian cases. In all concrete situations, we can decide which approach is more useful depending on our information, preliminaries, and the consequences of our answer or decision.

The current pandemic renewed an interest in understanding problems of diagnosis. What do diagnoses really imply and how do they connect to statistical and Bayesian inference? We agree with Borovcnik and Kapadia's (2015) statement that, "following the classical significance tests of Fisher and the statistical tests by Neyman and Pearson, and decision theory, two more approaches are considered here using qualitative scientific argument: the Bayesian approach, which is linked to a contested conception of probability, …" (p. 1). Therefore, we integrate the Bayesian approach in our planned curriculum. In reaction to their summary, that "the prime reaction of the audience was that everything beyond a frequentist approach to probability and—especially a comparative programme for statistical inference—is suitable only for a minority, as most students will not understand the inherent concepts" (p. 1), we plan to address the conceptual issues carefully and prepare detailed materials for teachers. Our long-term research with future teachers supports our design of the material and how to prepare the teachers in special seminars before the next experimental teaching phase (Vancsó, 2009). The problems in teaching classical statistical methods are—to our analyses—due to the complexity of the concepts and the reduction to classical methods as classical and Bayesian methods have developed in a twin relation and cannot be fully understood when the concepts are separated. Barnett's (1982) perspective about a comparative statistical inference supports our preference of using the various approaches and teaching them in parallel.

ACKNOWLEDGEMENT

The paper was financially supported by the project "Guided Discovery Mathematics Education—Tradition and Innovation" of the Hungarian Academy of Sciences. Using problem sequences to introduce the topic is from research in the framework of the project "Complex Mathematics Education in the XXI. Century" (heritage of Tamás Varga).

REFERENCES

Albert, J. (2002). Teaching introductory statistics from a Bayesian perspective. In B. Phillips (Ed.), *Developing a statistically literate society. Proceedings of the Sixth International Conference on Teaching Statistics (ICOTS6).* ISI/IASE. https://iase-web.org/documents/papers/icots6/3f1_albe.pdf?1402524960

Artigue, M., & Blomhøj, M. (2013). Conceptualizing inquiry-based education in mathematics. *ZDM–Mathematics Education, 45*(6), 797–810. https://doi.org/10.1007/s11858-013-0506-6

Barnett, V. (1982). *Comparative statistical inference* (2nd ed.). Wiley.

Berry, D. A. (1997). Teaching elementary Bayesian statistics with real applications in science. *The American Statistician, 51*(3), 241–246. https://doi.org/10.2307/2684895

Borovcnik, M. (2019a). Risk and decision making—Modeling and statistics in medicine: Case studies. In B. Sriraman (Ed.), *Handbook of the mathematics of the arts and sciences* (pp. 1507–1542). Springer. https://doi.org/10.1007/978-3-319-57072-3_126

Borovcnik, M. (2019b). Risk and decision making: Modeling and statistics in medicine—Fundamental aspects. In B. Sriraman (Ed.), *Handbook of the mathematics of the arts and sciences* (pp. 1471–1506). Springer. https://doi.org/10.1007/978-3-319-57072-3_62

Borovcnik, M. (2020). Many paths lead to statistical inference–Should teaching it focus on elementary approaches or reflect this multiplicity? *Teaching Mathematics and Computer Science, 18*(4), 259–293. https://doi.org/10.5485/TMCS.2020.0504

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

Borovcnik, M., Fejes-Tóth, P., Jánvári, Z., & Vancsó, Ö. (2020). Experimente zur Einführung von Ideen und Denkweisen statistischer Inferenz im Gymnasium. *Stochastik in der Schule, 40*(1), 18–27.

Borovcnik, M., & Kapadia, R. (2015). A comparative study of statistical inference from an educational point of view. In *Proceedings of the 60th World Statistics Congress* (pp. 26–31). International Statistical Institute. https://2015.isiproceedings.org/Files/CPS694-P1-S.pdf

Carranza, P., & Kuzniak, A. (2008). Duality of probability and statistics teaching in French education. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching* *statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference.* ICMI/IASE. https://www.stat.auckland.ac.nz/~iase/publications/rt08/T1P2_Carranza.pdf

de Finetti, B. (1937). La prévision: Ses lois logiques, ses sources subjectives. *Annales Institut Henri Poincaré, 7*(1)*,* 1–68.

de Finetti, B. (1974). *Theory of probability* (Vol.1). Wiley.

Fejes-Tóth, P. (2020). Inferential statistics—Research on the introduction of a new topic in the Hungarian high-school curriculum. In G. Ambrus, J. Sjuts, Ö. Vancsó, & É. Vásárhelyi (Eds.), *Komplexer Mathematikunterricht: Die Ideen von Tamás Varga in aktueller Sicht* (pp. 157–180). WTM Verlag für Wissenschafliche Texte und Medien. https://doi.org/10.37626/GA9783959871648.0.10

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: Methodological issues* (pp. 311–339). Erlbaum.

Hawkins, A. S., & Kapadia, R. (1984). Children's conceptions of probability: A psychological and pedagogical review. *Educational Studies in Mathematics, 15*(4), 349–377. https://doi.org/10.1007/BF00311112

Krajcsi, A., Csapodi, C., & Stettner, E. (2021). Algotaurus: An educational computer programming game for beginners. *Interactive Learning Environments, 29*(4), 634–647. https://doi.org/10.1080/10494820.2019.1593862

Migon, H. S., & Gamerman, D. (1999). Statistical inference: An integrated approach. Chapman and Hall.

Nemetz, T. (1985). *Mathematics for higher level. Grade 11*. National Schoolbook Publisher.

Rényi, A. (1972). *Letters on probability.* Wayne State University Press.

Riemer, W. (1991). *Stochastische Probleme aus elementarer Sicht*. BI Wissenschaftsverlag.

Vancsó, Ö. (2009). Parallel discussion of classical and Bayesian ways as an introduction to statistical inference. *International Electronic Journal of Mathematics Education, 4*(3), 181–212. https://doi.org/10.29333/iejme/242

Vancsó, Ö. (Ed.). (2010). *Mathematics for all. Textbook for Grade 10.* National Textbook Publishers.

Vancsó, Ö., Borovcnik, M., & Fejes-Tóth, P. (2021). A complex concept about statistical inference and a planned school experiment based on it. In *Proceedings of the 63rd World Statistics Congress*. International Statistical Institute. https://www.isi-web.org/files/docs/papers-and-abstracts/154-day3-ips078-a-complex-concept-about-statis.pdf

Vancsó, Ö., & Varga, E. (2020). Kannst du das System schlagen? Wetten im Sport als Kontext für die Kombination der Ideen der klassischen, statistischen und subjektivistischen Wahrscheinlichkeit. In G. Ambrus, J. Sjuts, Ö. Vancsó, & É. Vásárhelyi (Eds.), *Komplexer Mathematikunterricht* (pp. 367–390). WTM Verlag für Wissenschaftliche Texte und Medien. https://doi.org/10.37626/GA9783959871648.0

Varga, T. (1970). Probability through games. A sample of three games. In International Commission of Mathematical Instruction (Ed.), *New trends in mathematics teaching* (Vol. 2, pp. 424–440). United Nations Educational, Scientific, and Cultural Organizations.

Varga, T. (1972). Logic and probability in the lower grades. *Educational Studies in Mathematics, 4*(3), 346–357. https://doi.org/10.1007/BF00302583

Varga, T. (1983). Statistics in the curriculum for everybody—How young children and how their teachers react. In D. R. Grey, P. Holmes, V. Barnett, & G. M Constable (Eds.), *Proceedings of the First International Conference on Teaching Statistics* (Vol. 1, pp. 71–80). Teaching Statistics Trust. https://iase-web.org/Conference_Proceedings.php?p=ICOTS_1_1982

Varga, E., & Vancsó, Ö. (2021, July 18–21). *Problem sequences for developing two basic notions: Probability and expected value in Hungarian secondary schools* [Paper presentation]. 14th International Congress on Mathematical Education. Shanghai, China.

Wickmann, D. (2001). Inferenzstatistik ohne Signifikanztest. In M. Borovcnik, J., Engel, & D. Wickmann (Eds.), *Anregungen zum Stochastikunterricht* (pp. 133–138). Franzbecker.