18/4 (2020), 259-293
DOI: 10.5485/TMCS.2020.0504
tmcs@science.unideb.hu
http://tmcs.math.unideb.hu

Teaching Mathematics and Computer Science

Many paths lead to statistical inference – Should teaching it focus on elementary approaches or reflect this multiplicity?

MANFRED BOROVCNIK

Abstract. For statistics education, a key question is how to design learning paths to statistical inference that are elementary enough that the learners can understand the concepts and that are rich enough to develop the full complexity of statistical inference later on. There are two ways to approach this problem: One is to restrict the complexity. Informal Inference considers a reduced situation and refers to resampling methods, which may be completely outsourced to computing power. The other is to find informal ways to explore situations of statistical inference, also supported with the graphing and simulating facilities of computers. The latter orientates towards the full complexity of statistical inference though it tries to reduce it for the early learning encounters. We argue for the informal-ways approach as it connects to Bayesian methods of inference and allows for a full concept of probability in comparison to the Informal Inference, which reduces probability to a mere frequentist concept and – based on this – restricts inference to a few special cases. We also develop a didactic framework for our analysis, which includes the approach of Tamás Varga.

Key words and phrases: Informal Inference, Statistical inference, Bayes inference, Resampling, Varga's approach, Elementarisation.

MSC Subject Classification: 97K10, 97K70, 97K50, 97D20.

Varga's approach towards the didactics of mathematics

In the attempt to investigate Varga's legacy for today's didactics of mathematics (e.g., Varga, 1970; 1972; 1982; 1983), a long-term project has been established at the mathematics didactics centre of ELTE, the Eötvös-Lórand Tudomanyegyetem Budapest, together with the Hungarian Academy of Sciences with the title "Recent Complex Mathematics Education Research Project". This title refers to one specific feature of Varga's approach: Rather than to reduce the complexity of curricula, resources, and mathematics, Varga intends to allow for a reasonable and necessary level of complexity. This complexity also refers to the mathematics taught and the interrelations between mathematical concepts: To facilitate an access to this complexity, Varga develops well-designed systems of tasks that involve the learners into a clear-cut problem situation – often embedded in games with analogies to develop by the learners with support of their teacher in a guided-discovery experiment.

Varga (1983) provides a system of tasks typical of his approach to the didactics of mathematics; he directs these tasks to 9-year-old children, thus designing an ambitious programme that links probability with statistical inference from the outset. In the introduction to this essay, Varga (pp. 71) notes that for small children

> "in their experiences about random phenomena their involvement is greatly increased by first predicting the outcome then performing the experiment, finally comparing the results with their prediction. In some cases a further step is an attempt to explain the pattern of results by some mini-theory. During their games they are motivated to devise strategies which increase their chances, and then they get some feeling for game theory. The mini-theories demand combinatorial considerations. Yet since the order is hypothesis → experience → theory, I feel it appropriate to subsume such learning sequences under the term statistics."

Varga takes the 9-year olds into the midst of statistical inference going far beyond the usual suggestions to introduce descriptive statistics in the primary school though Varga admits that he aims only at statistical inference at a heuristic level. He explains that it is essential to combine a suitable language for designating situations of uncertainty with qualitative degrees of probability with a body language, with signs given by shaping the hands and fingers to certain forms.

Varga continues with his famous example about "figure[ing] out and record[ing] the behaviour of a randomly tossed coin." It would not be Varga if he did not mention how to facilitate the protocol of the results by a square grid of 8 rows with 16 entries for each of the 128 "coin tosses" out of the mind of the children. In the ensuing analysis, he introduces also the "expected" number of results. He focuses on the number and length of runs, to which – according to well-known psychological investigations there are strongly biased conceptions among children and adults: Too many and too short runs are introduced into the mind protocol of “coin tossing” as compared to real coin tossing.

The next task in this series deals with criteria when to judge a coin as biased. Varga fabricates loaded “coins” by filling one side (or both sides) with coins, wrapping the cardboard, and marking one side with 0 and the other with 1 so that it can smoothly be thrown like a simple coin. After 30 tosses, the result is “17 zeros and (only) 13 ones”. The history of the sequence – as always with randomness everything is possible – shows that the 1s had a majority of 4 to 1 in the last five tosses. “Is the coin loaded or fair? Or should we go on with the tosses, before taking position?” (Varga, 1983, p. 75).

Interestingly, Varga not only asks the children to judge the “coin” but also leaves open the possibility of continuing to toss if one feels that the data is insufficient for a decision. The children in his experiment “voted for going ahead, doing 30 more tosses.” Here, Varga interrupts and asks them about their decision rule before they continue to toss the coin. In a classroom discussion, it turned out that up to 25 1s, the coin “would be maintained by them to be suspicious” while from 26 onwards “they could still accept it as being fair.” In particular, from 27, the majority voted that the coin is acceptable. Some votes were [still] cast to 28 and 29 [votes as dividing line for a fair coin].” Varga brought 9-year olds into a lively discussion about when an observation is rare enough that one could reject the null hypothesis that the coin is fair. Of course, this conclusion is only implicit in Varga’s classroom experiment, but the children’s reactions to it were quite encouraging, as they were able to find a very good rule for a decision after the first experiments.

Varga goes ahead with a variant of rolling dice with an investigation of the waiting time for the first six. The task is to bet on the exact waiting time. The children could wager whether the first six occurs on the first roll, or on the second, etc. Most of the children voted for the fourth roll. Their decision coincides with the median of the random waiting time – less or equal to four has a probability of larger equal ½ as has the opposite event that the waiting time is larger equal to four. The median coincides with a probability of roughly ½ under the assumption that the coin is fair. In their answering pattern of using a fifty-fifty criterion, the children decide similarly to the previous task.

After two experiments related to Benford’s law, Varga concludes his task system with an experiment that he calls “from perception testing to hypothesis testing”. In this, he varies the Fisher’s Lady-Tasting-Tea experiment (see Batanero & Borovcnik, 2016) with five trials only and chewing gums of different brands. Thus, Varga ends up with the classical significance test and *p* values for 9-year olds – of course only at a heuristic level (p. 80):

"From the point of view of teaching statistics, the reactions of children are more interesting than technical details. Batches of five trials were usually administered. Children felt that the following qualification is fair:

No failures in five trials: excellent One failure in five trials: acceptable"

In his typical way of a Socratic dialogue, Varga continued with the children:

"How many in ten trials do you think would be acceptable? […] The answer was again almost immediate: [two]. The teacher tried to make clear that 'acceptable' means hardly by pure chance, and asked them, if they felt that two or less in ten trials could be the result of guesswork just as easily as one or none in five. They did. He decided to postpone confronting them with contrary experimental evidence at a later date."

This was then done by shaking boxes containing five, ten, or twenty two-colour counters. Without any calculations, the children recognised that their reasoning through proportionality – which they applied intuitively – was misleading them in situations like these. There are other task systems of Varga, such as the Three Discs (Varga, 1970). Gosztonyi (2017, pp. 1735) characterises his approach by the following statements (bullets by the present author):

- "In the situations described by Varga, teacher and students are in permanent dialogue: Varga gives several examples how to guide these dialogues in order to help the development of mathematical notions, but to give also important autonomy to the students in this process. […]
- In Varga's conception, ordered series of problems play a crucial role in the construction of long-term teaching processes. […]
- A similar process is described for different materials. [Varga's] handbook indicates the analogies between the corresponding phases of these different activities, and explains their differences which make them problems of different nature in the eyes of students […]. Students have to recognize progressively the links and the analogies between these different problems: that is what will lead to a progressive generalization of methods and solutions.
- The series contains activities with different materials: the students build towers with coloured cubes; thread beads […]. One organizing principle is the variety of experiences, apparently fare from each other, and stimulating a diversity of senses. But there is also certain progressiveness in their order, namely in the level of abstraction: starting from the manipulation of physical objects, through drawing and until the manipulation of symbols."

We may summarise Varga's approach in the following way:

- Reflections on the nature of mathematics. Mathematics is in permanent development, a view that connects to Fischer's (1984) Open Mathematics. Mathematics grounds on intuitions and experiences, which brings out the role of intuitions and connects Varga's ideas to Fischbein (1987). Mathematics and its teaching have a genuine dialogic nature embedded in social activities, which anticipates Glasersfeld (1991) but bases constructivism on a dialogue.
- Didactical approach. Varga focuses on heuristic methods of teaching and a limited use of formal language. He embeds the educational exercises in creative, playful activities often originating from games where the learners clearly understand the situation and the task. He makes extensive use of tools and manipulatives that facilitate a creative approach by the learners so that their thinking is not restricted by language and formalism they do not understand. As teaching strategy, Varga counts on guided – yet active discovery. This continues to be a balancing act between guiding the learners and letting them enough room for their own creativity. Yet, a substantial part of subtle guiding is required in Varga's approach as he aims at understanding the present mathematics and not constructing one's private conceptions.
- Pedagogical and psychological background. Varga connects to Piaget's "abstraction réfléchissante" and the stages of development. Yet, based on relations between scientific and everyday concepts, he introduces the mediating role of semiotic forms and abstraction and generalisation from Russian pedagogy (e.g., Vygotsky, 1967; 1978; Davydov, 1967). The Building Towers (Varga, 1982) show the increasing level of abstraction from a building-blocks task ending up with tree diagrams to represent the combinatorial solutions at a very general level.
- Complexity. Varga focuses on complexity in any way perceivable. Complexity of curricula, resources, and practices. Complexity of the mathematics that the children have to learn, which comprises the coherence of concepts, their interrelations, and the connectedness of various domains of mathematics. Last but not least, complexity of pedagogy including task systems and methodical interventions such as the Socratic dialogue, which is

especially “dangerous” for the teachers if they lack the mathematical background or if they lack the awareness about the pupil’s current status of knowledge and strategies. This is a definite rejection of plain simplification; the simplification lies in the specially chosen approach, which allows corresponding abstractions and generalisations to higher and wider levels.

- Key role of task systems. Tasks (or games with tasks) should allow for a creative approach by the learners, in which they should ripe in mathematical notions by developing their solution strategies and finally by comparing different solution strategies. The tasks should be organised to task systems (an idea later resumed by Steinbring, 1991) that allow for observing differences in problems, and recognising analogies between tasks that may stabilise the thinking processes so that flexible interrelations between the concepts emerge, which enable a progressive generalisation of methods and solutions.
- Demand on teachers. Varga expects that teachers should have high-level mathematical knowledge and combine it with pedagogical creativity. Specific challenges lie in the design of interventions that allow for autonomous yet guided discovery for the children. A paradox per se. However, Varga has developed quite many guiding works for teachers, which they might transfer also to other areas of teaching mathematics.

Todays’ efforts in the *Complex Mathematics Project* focus on the development of guidelines for the Socratic Method for various topics of mathematics.

Didactic approaches described by didactic triangles

To evaluate the approach of Varga, it may be helpful to give an overview on several approaches towards the didactics of mathematics in general before we investigate trends towards teaching probability and statistics. A didactic triangle forms the frame for describing the constitutive elements of the various approaches. This triangle connects the vertices of theory (the mathematics to be learned), reality (to which the mathematical concepts are to be applied and from which the tasks originate, which should be solved), and the subject (which should be enabled to think in mathematical terms about the real problem). From the discussion of the approaches, it becomes clear that the role of intuitions turns out to be even more important than it already is in other parts of mathematics and mathematics learning.

For the sake of analysis, we separate the vertices. Yet, the discussion makes it clear how they are interrelated and that they get their full didactic potential only within the network of interdependencies. The vertex T of theory relates to the logic foundation and development of the theory. The vertex R of reality concerns the real phenomena, irrespective of the fact whether a thinking subject locates these phenomena or not, or whether this subject investigates these phenomena by any kind of theory. Finally, the vertex S of the subject, again for the while isolated, and not determined whether it is seen from a cognitive-psychological perspective or in a sociologic embedding. The subject corner embodies its emotional and rational components. The interplay between these vertices is of interest for didactical considerations. For probability and statistics, the reader may find several sources that relate to this interplay (see, e.g., Kapadia & Borovcnik, 1991).

We investigate the trends of the didactics of mathematics to see how the vertices of the triangle and their interrelations shape the various approaches. There are overlaps in time and in the approach of several authors who followed or developed the one or the other approach. The intention is to get a concise description of the general development (see Figure 1).

The vertices in the triangle and their mutual relationships change in the course of development. Within the position of New Mathematics, the assumption was that if only the theory corner is self-contained and built up without gaps, this regulates all understanding and all applications. In the following wave of applications, one acknowledges already a substantial feedback from applications back to an understanding of the theoretical concepts, which are no longer isolated from each other, but are only “brought to life” by the applications.

In Freudenthal’s Phenomenology (Freudenthal, 1983), phenomena become the starting point for the acquisition of concepts; concepts of the theoretical corner serve and are developed to order and structure phenomena of the reality corner. In Fischer’s Open Mathematics (Fischer, 1984), the focus is on the model-forming subject who uses theoretical concepts and develops them if necessary to create models of reality. Fischbein’s Interplay of Intuitions and Mathematics (Fischbein, 1987) sees the individual acquisition of concepts as characterised by a peculiar, alternating influence of primary ideas of the subject and theoretical inputs from the theory corner, intensified or even initiated by events, images or situations from the real corner, resulting in secondary intuitions that dynamically influence (boost or hinder) the further process.

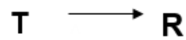
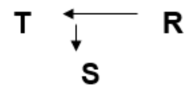
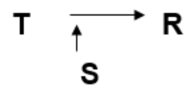
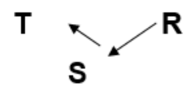
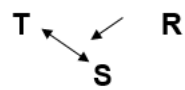
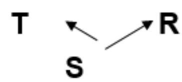
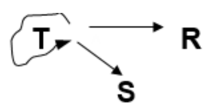
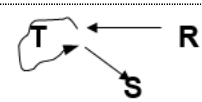
<table>
<tr><td>• 1960s New Mathematics
All one needs is logic and investigation of T.</td><td>
</td></tr>
<tr><td>• 1970s Applications.
The subject S just has to learn how to apply T to R.</td><td>
</td></tr>
<tr><td>• 1975s Freudenthal's Didactical Phenomenology.
Learn from phenomena from R by developing T.</td><td>
</td></tr>
<tr><td>• 1980 Fischer's Open Mathematics.
Develop T in order to solve problems in R.</td><td>
</td></tr>
<tr><td>• 1985s Varga' Rich Task Systems.
Rediscover T by problems from R embedded in games.</td><td>
</td></tr>
<tr><td>• 1985s Fischbein's Intuitions and Theory. Develop an interplay between concepts from T and intuitions from S.</td><td>
</td></tr>
<tr><td>• 1995s Constructivism.
S constructs concepts from T to explore R.</td><td>
</td></tr>
<tr><td>• 2000s Simulation.
Simulate T to find patterns in "R" to solve problems in R.</td><td>
</td></tr>
<tr><td>• 2010s Machine Learning.
Establish patterns from experience in R to replace T.</td><td>
</td></tr>
</table>

Figure 1. Trends in the didactics of mathematics and the didactic triangle

Varga's approach (see Gosztonyi, 2015a; b; 2019) focuses on the input of real situations – often embedded in games – and the learner tries to solve a posed problem to develop a partial view on a mathematical concept. In order to reach at such a view in a way in line with a traditional mathematical concept, the learner has to be creative on his own side but is sensibly dependent on the support by the teacher. Varga follows a form of Socratic dialogue to guide the discovery of the mathematical terms by the learners. Insofar, the triangle opens to the teacher and teaching intervention as the dialogue between teacher and learner gets a dynamic role in concept acquisition.

Constructivism (Glasersfeld, 1991), on the other hand, places the process of confronting an individual (or a group of people) with real-life situations at the starting point of active concept construction; subjective areas of experience play a decisive role in how the concepts are actually constructed and understood (see Bauersfeld, 1988). Newer approaches to the didactics of mathematics only varied the existing ones especially refining the position of constructivism: define the role of the teacher better, optimise the design of tasks, and include new media. We see less reference to investigate how such task systems are interwoven and how one should design task systems in order to develop the mathematical concepts from a variety of perspectives, as Varga intended with his task systems.

We discuss two developments that play a revolutionary role in the didactics of stochastics. The first is the simulation approach, which examines a stochastic situation by simulating the assumptions behind it. The mathematical references play a subordinate role; after all, one may solve the problems by simulation alone so that it becomes obsolete to learn the theory. However, critical voices assume that teaching must build up a profound understanding of the stochastic assumptions behind the simulation algorithms. This means that one can no longer consider the vertex of theory in isolation and that the subject must actively participate in the exploration of the vertex of theory.

Even more difficult is the position of the subject in the machine-learning approach. Learning through experience certainly shapes human learning. If one neglects any other types of learning, it becomes superfluous to include the theory corner to build up sensible concepts and analyse their connections. It becomes also obsolete to acquire theoretical concepts for an overview of the problem, which should provide a structural insight. Theory in fact becomes void when algorithms that simply reflect experience remain the only source of new knowledge in the machine-learning paradigm: Everything needed is to expose a self-learning system to multiple situations and to observe the consequences of the system’s reactions and adapt the algorithm accordingly by an evaluation of the responses. A typical example relates to chess computers, which outperform the best chess players due to knowledge of previous chess games. This makes chess uninteresting. Yet, mathematics will still be important to understand and evaluate the structure of the system’s responses.

Trends in the didactics of stochastics

The development of a discipline is not linear, but with hindsight, it is still possible to classify the achievements and to identify certain trends and desiderata. The didactics of stochastics can be summarised concisely as follows: One should incorporate stochastic concepts into the curriculum earlier, and as suggested in the proposals at the beginning of the 1970s, one should get by with stochastics with less effort and scope. We describe these trends in detail.

Trend towards philosophical clarification of the concepts

Central to this trend is the nature of the concepts: What is probability? What do statistical statements mean? In this context, the references to epistemology, to the theory of science or to the historical emergence of the concepts become significant. Key topics and works in this tradition are:

- Meta knowledge about the meaning of statistical statements (Heitele, 1975; Steinbring, 1991).
- Bayesian controversy and its impact on the meaning of probability (Barnett, 1982; Wickmann, 1990; Vancsó, 2006).
- Learning from history (Kapadia & Borovcnik, 1991; MacKenzie, 1981).

Advantages and disadvantages of this trend are very close together. While the relationship between probability and the evolution of relative frequencies is central to an adequate understanding of stochastic terms, for didactical purpose, there are always *two* urgent questions: How can we perceive the concepts? How can we use the concepts? We may see the full complexity of probability from profound analyses such as Batanero, Chernoff, Engel, Lee, & Sánchez (2016). The concepts are not only sophisticated but remain virtual as Spiegelhalter (2014) states. Clarification will remain superficial and may face emotions and intuitions so that it can lead to never-ending discussions and confuse the learners more than it would clarify the concepts for them. At least, it is not convincing for learners to invest too much effort on learning it if it is not already clear for them that the concepts are *useful to solve problems*. This means that the didactic priority may be to let the learners experience first that the concepts are useful. Especially for stochastic instruction, there seem to be strong emotional barriers against the form in which one formulates the concepts within a theory, so that showing their utility becomes even more important than in other sub-disciplines of mathematics. Varga (1970; 1982;

1983) embeds playful activities in his task sequences, which engages the learners towards solving a goal; he postpones the clarification issue to a time when the students have already acquired their own experience with the concepts.

Trend towards applications

Following the general trend towards applications and modelling in the didactics of mathematics as a reaction against the New Math, applications shifted into the focus of stochastic teaching from the early 1990s. Even in epistemologically and scientific-theoretically oriented analyses, the role of applications has been widely discussed, for example, the application problem (primarily in physics) from a historical perspective by Steinbring (1991). In detail, the focus on applications resulted in a reorientation of the priorities in the (proposed and actual) curricula. The applications also connected probability and statistical inference more tightly and the potential for statistical inference changed the view on probability from an objectivist-frequentist concept to something more general including Bayesian aspects. Later, Vancsó (2006) attempted to reunite the classical and the Bayesian theory of probability for didactical purpose.

As an alternative to statistical inference, the 1990s witnessed a didactic revival of Exploratory Data Analysis (EDA, Borovcnik & Ossimitz, 1987; Biehler, 2007). There are interesting parallels between the interactive style of EDA and a project-like approach in teaching. In the Anglo-American culture, projects have always had a different status and play an important role in statistics education, see Hawkins (1991) or the multiplicity of presentations at ICOTS 4 on project-based teaching organised by Bentley (NOC, 1994, pp. 17-31).

The driving force behind was to apply mathematics with the purpose of making decisions under uncertainty. The advantages are obvious as applications can lead to a deeper understanding of the underlying mathematics. The disadvantages lie in the rolling out of frequently used (but trivial) methods of descriptive statistics so that too little time was left for teaching methods of inferential statistics (also important for applications, but complicated). Further disadvantages of applications in general are the strong dichotomisation of “too easy – too difficult” and the (still) lacking didactic approaches to teach difficult techniques and terms (for non-mathematicians) in a way that makes their significance and limitations for a situation clear and thus guides the interpretation of results.

Missing trend towards intuitions

The discussion of stochastics didactics is still characterised by the keyword "stochastic thinking" (see Borovcnik, 2018), which goes back to Heitele (1975). Yet, no one has really explained what stochastic thinking actually is, but one argues that it is different from other ways of thinking, such as logical thinking or the tracing of causal chains. Fischbein (1987) regards intuitions or intuitive concepts as essential for an understanding of probability. The experiments of Varga (1983) are suitable for making this peculiar way of thinking tangible within the framework of stochastics.

It is not surprising that in other areas of mathematics didactics, schemata, or concept images, etc. are use to describe the process of acquiring concepts (Piaget & Inhelder, 1951; Tall & Vinner, 1981; or Bender & Schreiber, 1985, for geometry); intuitions, on the other hand, play a marginal role and are rather seen as something to be eliminated. There is a broad consensus in mathematics didactics about the operative acquisition of concepts, which, on the basis of concrete materials and a mental reflection – "abstraction réfléchissante" – on operations, should lead to an adequate conception of the concepts, independent of (existing) intuitive ideas of individual persons; important references here are Piaget and Inhelder (1951), Wittmann (1973), and Dörfler (1984).

In stochastics, on the other hand, the concept of operative term acquisition has not gotten very far. In contrast, Fischbein (1987) wrote an entire book on his interplay between primary and secondary intuitions. Especially the emotional content of intuitions seems to be decisive for the acceptance of stochastic terms, an acceptance that lies before the actual acquisition of the term and the efforts involved. A world of primary intuitions that are transformed into secondary intuitions by theoretical progress and real results: secondary intuitions that are all too often in conflict with "official" ideas, as the usage of the terms misconceptions and fundamental errors indicate. The following points highlight the great importance of intuitions:

- The discussion on stochastic thinking: Fundamental ideas are still confused with mathematical relations. Borovcnik, Bentz, and Kapadia (1991) trace *ideas* in the historic development.
- The abundance of puzzles and paradoxes: For an intuitive explanation of such paradoxes, see Winter (1992); examples are in Székelyi (1986).

The role of intuitions in teaching can be summarised as follows: empirical studies (e.g., Kahneman, Slovic, & Tversky, 1982) can show the weak points in the primary network of intuitions, i.e., where one must work on in teaching. The consequence of neglecting learners' intuitions is, that what has been learned is stored unconnectedly next to these emotionally laden ideas and in the frequent cases of conflict will always be inferior; if teaching is to be effective, these primary ideas must be actively taken up. This is what Fischbein (1987) does with his programme.

Trend towards elementarisation

Didactic efforts focussed only for a short time on the axiomatic setting of probability (Heitele, 1975), as it failed to contribute to a deeper understanding neither of probability nor of statistics. The wish to cover statistical inference made it necessary to find ways for making the concepts elementary:

- Simulation: Simulation serves to avoid probability calculations. Furthermore, this method may throw light on the relationship between probability and relative frequencies and illustrate central theorems of probability (such as Laws of Large Numbers, Central Limit Theorem) and essential characteristics of methods of inferential statistics.
- Non-parametric statistics and resampling methods: Originally, these methods should solve special cases for simplifying the pre-requisites (Noether, 1967). Until about the early 2000s, computing capacity restricted these methods to statisticians and they have not been accessible for didactical considerations. Yet, already in the 1990s, a didactical analysis clarified that resampling can serve also as a transient stage towards a fuller statistical inference.
- Exploratory data analysis (EDA): EDA has been much in use not only in the applications but also in the didactics of stochastics (Borovcnik & Ossimitz, 1987; Biehler, 2007). Yet, it follows a different paradigm in the applications. While in classical applications, after the modelling phase, there is a strict separation between reality R and model (theory) T, a tight interconnection between R and T forms the motor of the interactive analysis in EDA.

Using simulation for illustrating properties of statistical methods shifts the focus to frequentist properties of confidence intervals or of errors of type I or II, which distorts their actual meaning and neglects their *virtual* character (Spiegelhalter, 2014). Resampling solves only special cases within a restricted framework of statistical inference while EDA requires a non-scientific paradigm of modelling, which makes it less useful if one wants to transfer solutions to others.

Trend towards computerisation

A decisive question for teaching applications is the tools for calculations. Computers do not only serve for complicated calculations, but also as a conceptual support, both for direct, visual illustration of terms and for illustrating the consequences of a particular version of terms and methods:

- To visualise data appropriately and thus to make descriptive statistics meaningful.
- To use the method of simulation rather than complicated probability distributions so that more realistic problems could be set.
- To introduce methods of statistical inference, either for the calculation of the procedures or for illustrating long-run properties of the methods.
- To make the statements of central theorems of probability comprehensible without a mathematical proof. This helps to close the educational gaps in the learning of the theory.

Computers facilitated didactical innovations in probability and statistical inference that otherwise would never have been implemented. In applications, computers have led to computer-intensive methods around resampling, Big Data, and dynamic visualisations; it boosted also Bayesian methods based on a qualitative probability conception but the latter seems restricted to experts though there are a few endeavours to foster statistical inference around a parallel approach to classical inference and Bayesian methods (Vancsó, 2009; 2018). Yet, in didactics, computers materialised the concepts in a biased way, as they provoked to reduce probability and inference to mere frequentist concepts. The reduction of probability to frequencies is still ongoing. This author is sure that Varga in his intention of a complex-mathematics approach would regret such a biased view on probability.

Trend towards statistical inference

In line with the endeavour to seek for interesting applications of probability in the trend towards applications, there was a joint effort in the German didactic community to find a suitable approach towards teaching statistical inference. Probability without the potential to serve as justification for statistical inference is blind and inference without the probabilistic part is incomplete. One of the main purposes of probability (apart from decision-making, modelling risk, and reliability) is to describe the data-generating process by probability. Without random samples, it is not possible to apply the methods of statistical inference properly. In practice, vague claims that the sample is representative often substitute the random sample argument. While quota sampling requires sophisticated knowledge about the population (and then it works surprisingly well), in many areas, the random sample is just “manoeuvred” into the analysis. In the era of empirical evidence, inferential statistics got a key role for justifying conclusions from data (no matter whether it was appropriately applied or not). The random sampling argument was a wild card for getting one’s analysis “approved”.

All this in mind, researchers in the didactics of stochastics sought for viable learning paths through the acknowledged complexity of statistical inference. Early approaches such as Strick (1980) were heavily criticised that they distort statistical inference by the simplifications introduced for the sake of being able to teach it at the secondary level (see Diepgen, 1985a, for the critique, and Strick, 1985, for the answer to it). Diepgen (1985b) provides tasks where it is better to model the situation within a Bayesian framework to find a suitable decision.

Müller (1989) already introduced non-parametric considerations of statistical inference into the didactics. Internationally, as may be seen from the conferences on Teaching Statistics (ICOTS), researchers regarded the topic of inference as too difficult to teach and there is no visible development until computing capacity became easily accessible so that one could implement the non-parametric approach and resampling methods into teaching. After that, the trend broke and Cobb (2007) asked to replace traditional statistical inference by resampling methods completely (see the section on the *Trend towards Bootstrap and resampling* below). The German discussion still sought for viable ways to teach classical statistical inference (Gigerenzer & Krauss, 2001) and even investigated the Bayesian way to make inference understandable for the learners. Several meetings of the Arbeitskreis *Stochastik in der Schule* were devoted to statistical inference and Vancsó (2001) is one of the first approaches towards a parallel way of classical and Bayesian methods in teaching at high school (for the discussion, see Borovcnik, Engel, & Wickmann, 2001). Vancsó continued his work on learning paths for statistical inference by exploring the didactic value of his parallel teaching of both approaches (2009; 2018).

Yet, for statistical inference, the international discussion seems to have unanimously decided for the so-called Informal Inference with the sole orientation of statistical inference towards resampling techniques. As this approach represents a distortion of statistical inference, there remains an urgent task for the didactic community, namely to find informal ways to teach statistical inference. Attempts may be seen in Batanero and Borovcnik (2016) or Borovcnik, Fejes-Tóth, Jánvári, and Vancsó (2020). Borovcnik (2019) illustrates the potential of genuinely informal ways to teach statistical inference and discusses shortcomings of the elementarisations within Informal Inference.

Trend towards risk evaluation

Risk has been advocated by Gigerenzer (2002) or Martignon and Krauss (2009). Borovcnik (2015) provides a comprehensive analysis of risk from a didactic perspective. Risk has been historically the motor of the development of probability. Games and insurance are structurally equivalent. An exchange of the position between two stakeholders signifies both situations: The one leaves the position of security, the other leaves the position of risk: For the insurance contract, the client pays the company in advance so that the company takes his risk (with a potential negative impact) over. For the game, the client pays the casino to get into the risky position (which is now a potential win) and the casino leaves the situation of no risk. The exchange of the uneven situations has to find a price for which both partners are willing to swap roles. Risk shifts the connotation of probability towards a qualitative notion and focuses on the purpose of probability for transparent decision making.

One may regard the situation in statistical inference as decision making, which twists the focus towards Bayesian methods rather than classical inference. Didactically interesting is that statistical inference is much easier to understand under the perspective of risk and a decision-theoretic point of view (Borovcnik, 2015), especially as it shifts the focus away from a biased frequentist probability concept and thus avoids many misunderstandings in probability (Carranza & Kuzniak, 2008).

Trend towards Bayes methods

It might be of advantage to illustrate the approach by an example. In Germany, the lottery machine draws 6 out of 49 balls without replacement. The context assumes that a person comes to a distant country and sees the latest drawing: 4, 6, 15, 19, 40, and 37. As we do not know the number *N* of balls in the lottery machine, we model the initial ignorance by a uniform distribution on the interval [30, 80] for the number *N* of balls or express our subjective judgement as equiprobability. Then, this distribution changes week by week by the numbers drawn and one quickly receives very precise information, because the revised distribution for *N* contracts to one or two values after only a few weeks. Thus, the initially subjective uniform distribution for *N* gradually leads to a more objective assessment.

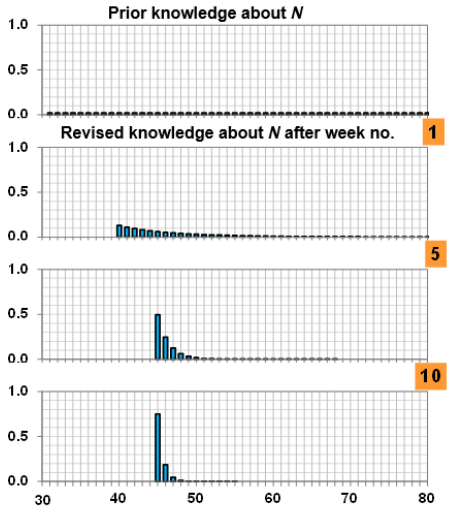


Figure 2. Prior knowledge about the number of balls revised after 1, 5, and 10 drawings

We only show the initial distribution (first line in Figure 2) and three intermediate stages, which correspond to the level of information when the lottery numbers from 1, 5 and 10 weeks are known (further lines in Figure 2). The example is from Borovcnik, Fejes-Tóth, Jánvári, and Vancsó (2020). From Figure 2, it becomes clear that the country could not be Germany.

There have been many didactic analyses about the potential of Bayes methods starting from the early times (Wickmann, 1990). In 1997, The American Statistician published a fierce dispute between Bayes and classical inference. David Moore, a statistician with high reputation in the didactics of statistics, ended the discussion by “Bayes is too difficult to teach”. Despite such comments that influenced the further development, Vancsó (2009) investigated a parallel approach of classical and Bayesian inference in order to understand both approaches better. In the statistical discipline, Bayesian inference is very popular for modelling. Yet, in the didactics of stochastics, Bayes methods have not reached wider access to the curricula.

Trend towards Bootstrap and resampling

In the section on the *Trend towards statistical inference*, we referred to the difficulty to find learning paths to this complex topic. Apart from primitive rules to mimic confidence intervals (the two-sigma rules of Strick, 1985), there was a hope that the methods of re-randomisation and Bootstrap, beginning from non-parametric methods (Noether, 1967), would provide an introductory chapter to statistical inference also for high school. While the discussion in the German community still was on the possibility to continue from this restricted form of inference towards a full methodology of inference, Cobb (2007) advocated for a radical change: replace the classical statistical inference by the methods of resampling (Efron & Tibshirani, 1993; Lunneborg, 2000). Rossman (2008) followed this advice and Stohl Lee, Angotti, and Tarr (2010) provide suggestions for teaching alongside the new paradigm. For a didactical discussion of these methods, see Borovcnik (2019).

Meanwhile, the label *Informal Inference* designates this approach. Within this paradigm, inference grounds exclusively on the data and apart from reshuffling the data, there is no reference to any probability distribution in the background. Clearly, the approach is computer-intensive and only the growing computing facility enabled the method. The approach is intuitive but leads to a restricted form of statistical inference. As DelMas (2017) argues, the randomisation methods should replace classical inference (and thus bypass Bayesian inference). For a substantial critique of resampling methods from a statistical point of view (not a didactical one), see Howell (n.d.). Informal Inference comprises two techniques of data shuffling – Bootstrap and resampling.

Estimation. Informal inference reduces it to Bootstrap intervals. While in classical inference, mathematical properties of the process of random sampling determine the sampling error, informal inference investigates it empirically by sampling from the given data with replacement: Instead of sampling from the population (the true distribution function F), one samples from the initial sample (an estimate of F). Bootstrap yields approximate confidence intervals; one has to apply very sophisticated techniques to improve their quality so that their intuitive simplicity gets lost.

Hypothesis testing. Informal inference replaces it by randomisation tests. A typical case is the comparison of two groups, a treatment and a control group where an act of randomness establishes the original attribution of "patients" to the groups. Introducing randomness should guarantee an even attribution of all non-random influential factors to the two groups so that the difference between them becomes purely random apart from treatment effects. This allows judging the differences by a statistical test. In the classical framework, one has to formulate a null hypothesis and an alternative hypothesis. This is not necessary within the re-randomisation framework.

We give a short example for the Bootstrap interval and refer the reader to Borovcnik (2019) or DelMas (2017) for the details of the randomisation test. Note that Bootstrap does not use any hypothesis; the "no difference" hypothesis in re-randomisation setting is "natural". Given: a sample of size *n* with mean and SD for a specific variable. How precise is the mean of the sample as a measurement for the population? *Time* relates to workload (in hours) for a seminar. Rather than sampling again from the population, which is impossible, we sample from the first sample (with replacement). The first Bootstrap yields a new measurement of the mean of the population (see Figure 3). This is analogue to how physicists would measure an unknown quantity. We then investigate the precision of our measuring device by repeating the measurement, which provides the Bootstrap distribution of (the "repeated" measurements of) the mean.

Raw data

Nr	Times
1	12
2	2
3	6
4	2
5	19
6	5
7	34
8	4
9	1
10	4

n	Mean
10	8,90

1. Bootstrap

Nr	Times
10	4
1	12
8	4
9	1
2	2
9	1
7	34
5	19
1	12
8	4

Mean
9,30

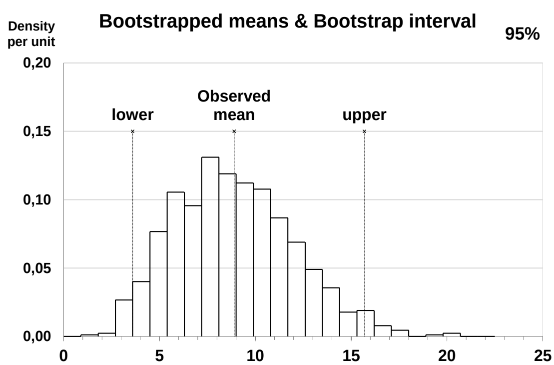


Figure 3. Left: Original sample on *Time* and first resampling of this sample
Right: Bootstrap distribution for the mean with quintile markers for the 95% Bootstrap interval

Trend to Big Data and machine learning

Big Data are massive data that originate as a by-product of other processes. Instagram has information about who likes what types of contributions or whom. Big Data are required for adapting algorithms for autonomous cars. Big Data are at the core of self-learning algorithms for diagnosing diseases from images such as CTs. We may investigate Big Data by machine-learning algorithms, which are mainly an exploratory use of regression trees or factor analysis. The problem is that Big Data often lack a clear structure and are so big that computational problems are huge. There are projects that make it accessible, which problems occur with Big Data (e.g., Hassad, 2020).

We illustrate potential and problems of Big Data by the famous flop of Google's data-aggregating tool, Google Flu Trends. Researchers designed this tool to estimate the spread of influenza-like illnesses from online search activity. Yet in 2012, the algorithm consistently and significantly resulted in overestimation. It predicted more than double the actual prevalence of flu as reported by the Centers for Disease Control and Prevention. Experts attributed this failure to the purely mechanical nature of Google's algorithm, which was free of theory and neglected temporally relevant factors (Butler, 2013). One neglected confounding factor was the fear-mongering news about the Flu, which resulted in an unduly high level of related search activity by those with and without flu symptoms. This has implications for the way we think about and reason with Big Data. We can be curious what the upcoming research will bring with it and how it will develop that kind of new thinking that leads off the traditional lines of statistical inference. The reader might learn about the new epistemologies and paradigm shifts with Big Data in Kitchin (2014) or in Prodromou (2017).

Trend towards civic statistics

More recently, a big European project on teaching political education uses statistics on society in order to develop the skills that are required for an active participation in societal decisions (Engel, 2019). The crux with such an approach is that it makes use of Big Data, dynamical visualisation of multivariate data, and methods of machine learning in order to structure the data, without explicit building the knowledge about the used methods and their restrictions. The approach makes the impression as if it were possible to decide societal issues objectively by statistical analyses without further input of values – as if the used statistics were free of any model with a restricted validity and as if it were possible to implement the results without value judgement. For example, if one arrives at a statistical gender wage gap, the question is what to advice to young women. The answer depends on values and a decision depends on the priorities between vocation (which does imply career only for a minority of persons) and family.

Trend towards dynamic visualisation

Visualisation embodies concepts in material and visible form while the concepts are theoretical (Steinbring), or metaphorical (Spiegelhalter, 2014). Schematic diagrams may enhance concepts. Visualisation is also supporting data analysis; descriptive statistics and EDA make extensive use of it. Dynamic visualisations have an authentic power and insinuate real “developments” while one can only represent the results of models (see Prodromou, 2017). The convincing power of such visualisations binds the audience emotionally while one would seek rational arguments and interpretations. For example, Rosling (2009) explains the development between income and life expectancy over five centuries by a dynamic visualisation. No word about the concepts (income, life expectancy), no mention about the logarithmic scale used for the representation, and no discussion about the interpretation of the concept of income over such a long period. The vivid and convincing conclusion: the richer, the longer humans live. Who can question that after seeing the video with nicely coloured balls of different shape bouncing up and down like little soap bubbles?

Elementarisation

In the didactics of mathematics, there has always been a need for elementarisation. Historically famous is Klein’s (1908) *Elementary mathematics from a higher standpoint*. Also in the more recent history of stochastic didactics, one may observe a trend towards elementarisation; see also the section *Trend towards statistical inference*. We see various efforts to elementarise stochastics:

- To disclose sources of insight that go beyond mathematics or to ground insight by means other than mathematics. To build up meta-knowledge of the terms that enable to grasp mathematics intuitively or even anticipate it with non-mathematical tools. To make the terms used understandable within more complex contexts.

- Probability is also connected to hopes and expectations for the future and competes with many – archetypically anchored – strategies (Batanero & Borovcnik, 2016), which also explain or predict coincidence or possibilities for the future; these strategies stubbornly resist change through instruction. A simple control of success is missing with chance due to a lack of consistent experience, because we tend to consider the situations as individual cases.
- Probability and methods of statistical inference based on it have to do with modelling, whereby the interpretation patterns between reality and model are complicated. Spiegelhalter (2014) speaks of probability as a virtual quantity and of a metaphorical use of probability. This virtual character applies even more for risks and indices for statistical inference such as the probability of coverage of confidence intervals or errors of statistical tests.
- In addition, we do not ground the interpretation of probability as relative frequencies on limit considerations. Limiting values always correspond to a genuinely theoretical view of things; they will never be visible in reality, neither in a confirmation of convergence nor in the recording of the proximity to the limiting value, and certainly, we cannot see where the limiting value will lie. For such considerations, the sequences of relative frequencies are unsuitable due to the independence assumption, that is, due to the lack of laws as they occur in calculus.
- Rather, we need specific thought experiments (as in Batanero & Borovcnik, 2016) to identify stable patterns, by systematically increasing the sample size. From this, we may recognise a "trend" and extrapolate it by a thought exper-iment. We can use this technique also for disclosing the meaning of terms of statistical inference.

Several approaches towards inference for teaching may serve the necessary elementarisation:

- "Informal inference" (Cobb, 2007; DelMas, 2017). Resampling and Bootstrap – a simplified version of inference.

- Decision-theoretic framework for statistical inference (Borovcnik, 2019). To identify the key concepts for statistical inference.
- Bayesian decision-making (rather than inference) (Borovcnik, 2015; 2019). The role of prior probabilities and prior distributions.
- Bayesian inference (Vancsó, 2009; 2018). Including prior knowledge about a phenomenon and combining it with data to establish the new status of knowledge.
- Informal approach to classical statistical inference (Batanero & Borovcnik, 2016). Playful activities to demonstrate key concepts of statistical inference.

"Informal inference" goes beyond informally exploring probabilistic models by simulation; it aims to replace traditional statistical inference. Borovcnik (2019) has provided a detailed critique about the narrowness of the approach though it may well serve as a transient stage to statistical inference in its full complexity. We should mention the following circumstances as decisive disadvantages of Informal Inference: The fact that it replaces the concept of probability completely by relative frequencies misses the genuine diversity of probability; in particular, it conceptually misses out the degree of confidence, subjective probability, and Laplace's equiprobability. This approach distorts Bayesian problems. There is actually no connection to methods of Bayesian inference in the continuation of the curriculum; but even the more elementary probability tasks that we solve with the Bayesian formula receive a completely one-sided interpretation if one no longer has a qualitative concept of probability, because relative frequencies replace probability completely.

Besides Informal Inference as a novel approach towards a restricted version of statistical inference, there is also the didactical possibility of designing informal ways towards inference. For exploratory investigations of the implications of hypotheses in the form of probability distributions, see Batanero & Borovcnik (2016). We advocate an engineering approach: Within such a setting, we avoid to develop mathematical details minutely. The goal is to ground an overall, synthetic understanding by relating to analogies that are more familiar to the learners, or by meta-knowledge, or by contexts that allow for a direct interpretation of the concepts. To specify the purpose of concepts might also help to understand them. A list of topics in Batanero and Borovcnik (2016) may give an idea how far this approach reaches. Within the following contexts, the learners gain a natural understanding of various concepts of inference by playing with parameters.

- Single-choice exam – Playing with binomial distributions to find the probability to pass an exam or to fix a suitable threshold for passing the exam.
- Lady tasting tea – Can we perform better than a random player can? Informal inference.
- Separating good and bad quality – Impact of a rejection number. The two error types are antagonistic.
- Statistical process control as exploration of scenarios – Judge the impact of introducing rules for deciding to interrupt the production process. Informal Tests.

Conclusions

The perspectives presented here should focus further didactic research. In response to the low degree of operationalisation of stochastic concepts, the explicit elaboration of the "what purpose can the notion serve for?" is essential. This question can take the form of guiding ideas, which we can link to the concept of fundamental ideas. The strong attachment to intuitions that we successfully use in other areas of life, but which often conflict with the actual stochastic concepts, makes the guided-discovery approach an attractive form of teaching. Analogies, the design of similar but more familiar contexts, could help to embed the necessary ideas. An intuitive exploration of mathematics lies at the basis of the perspectives formulated in the following as teaching strategies. We finish by final considerations about elementarisation to balance between the approaches of "Informal Inference" and informal ways to explore statistical inference.

Guiding ideas – How the "what for?" question clarifies "how" to understand the notions

As an answer to the difficult definitions of terms from the trend towards philosophical clarification and the resulting problems of understanding, the author sees the approach of an early treatment of the question "what can the terms serve for?" An analysis of this question provides the urgently needed basis for action for the learners and thus the basis for further theoretical considerations. In order to prevent a possible misunderstanding here: the clarification of the characteristics of the terms should be done but postponed until later, in favour of an increased focus on a utilitarian core of the notions. These properties are complex, confusing and difficult; it is advantageous if, at this stage, the learner is already convinced that these notions are useful and worthwhile to study.

Such guiding ideas could contribute to the construction of task systems (according to Steinbring, 1991), which highlight the theoretical character of probability. This is in line with Varga's focus on systems of tasks. These tasks might touch issues of risk and prefer decisions to statistical inference; or they might use classical and Bayes methods in the form of twins (Vancsó, 2009). Guiding ideas as strategies show how to make use of the terms. This is something similar to fundamental ideas; the different terminology should avoid that one refers to the central mathematical terms, a way in which previous research has misunderstood the role of fundamental ideas. Guiding ideas should draw attention to the question "What is the term good for?" and the question "What is the real meaning of the term?" To locate and elaborate such guiding ideas and to develop images of a lesson based on them is an urgent task of a didactics of stochastics (for an attempt, see Borovcnik, 2018).

Teaching interview or Socratic dialogue

We have seen the role of intuitions for the understanding of stochastic terms due to the lack of an operational basis. Fischbein (1987) has written about the long-lasting effect of intuitions; empirical studies show that the world of ideas is oriented towards other (often more direct and important) questions than stochastics can answer. No wonder that the intuitive ideas for the official terms often fail, if one misses to explicitly clarify what one can actually solve with the used notions. On the one hand, one has to shed light on the application of stochastic methods for the students, for example by means of guiding ideas; on the other hand, one has to be prepared for possible misunderstandings and reinterpretations as a teacher. Furthermore, a direct confrontation of the intuitive world with the official representations is promising for a better motivation and a deeper understanding.

The well-known Falk phenomenon may show the potential of a teaching interview (see Batanero & Borovcnik, 2016). These interviews highlight that we cannot simply fade out intuitive notions from the outset by an exact representation of the theory and its images; therefore, we should aim at a direct intuitive challenge to the learners. The teaching interview, modelled on the interview of empirical research, seems to be a useful tool here. The author himself began several introductory lectures at the university with an interview phase and gained new evidence about individual ideas and found points of intervention. In empirical research, an interviewer should remain neutral, which seems impossible for principle reasons because every input, every image, every comment, even a non-verbal cue is already an intervention and obscures the original idea and changes the existing primary intuitions into secondary ones – the terms primary and secondary are here used in the sense of Fischbein' approach. In contrast to the interview in empirical research, in a teaching interview, the teacher should consciously challenge intuitions; such a procedure should resolve the attitudes of the learners and partly change them permanently. For example, if a causal thought pattern becomes apparent from the course of the interview in a stochastic situation, we can manoeuvre the learner into a situation where he can recognise this thought pattern as absurd.

There is little evidence for a good design of teaching interviews; transfer from empirical interviews is only possible to a limited extent because of the different objective (deliberate provocation) and the immediate need for intervention. There is an urgent need to give an overview on persistent discrepancies in the behaviour of individuals and the official terms, the background to these discrepancies, and a planning of conscious interventions. There seems also a lack of knowledge to interpret such interview lessons profoundly. We formulate measurements to foster interview teaching as a desideratum for teaching if long-term changes in behaviour are to be achieved (an integration of stochastic assessments in personal decisions, a sensible handling of risks in the political discussion, etc.). In a certain sense, one may compare a teaching interview to a Socratic dialogue, which forms a key element in Varga's philosophy of guided discovery (see Gosztonyi, 2015a; 2019).

Analogies

Nothing is more common in methodological proposals on key stochastic concepts than analogies. The author uses a naive view of analogy here, by establishing a clear relationship between mathematical terms on the one hand and a factual context on the other. Since the context has a more general connotation, this relation is generally not isomorphic. An analogy is useful to start from a known context and to make a mathematical concept accessible from a factual point of view. However, one can also use an analogy the other way round and start from already known mathematical terms to structure an initially vague context. Subsequently, the mapping extends to the connection between formal relations on the mathematical level and content-related relations in reality. The advantage of using analogies is that one would directly be able to understand the factual relations and transfer this understanding to the analogous formal relations on the mathematical level.

The diagnostic context (2019), for example, allows a helpful insight into the character of conditional probabilities. Analogies may provide a deeper understanding, but also anticipating the mathematical precision that is to come. This anticipation of mathematical terms by means of non-mathematical repertoire allows understanding the mathematics or even "replace" it (this is an important aspect for the teaching of non-mathematicians). The idea is to make teaching more effective through the targeted use of analogies especially when it comes to increasing the acceptance of stochastic terms on the side of the learners.

"Informal Inference" versus informal ways towards statistical inference

Coming back to the issues of statistical inference and the ways to elementarise it, we compare the informal ways of teaching statistical inference in its full complexity and the approach of "Informal Inference" that absorbs statistical inference in resampling techniques. Key issues to re-consider for an "Informal Inference" approach are:

- Conceptual understanding differs from easier access to tasks and higher solving rates. Empirical evidence that is offered to show that the approach leads to higher achievement with simple tasks for students (DelMas, 2017), this has to be balanced with the conceptual drawback as probability is reduced to frequencies and probabilistic hypotheses have disappeared from the curricula as everything can be solved by reshuffling the data.
- How to adapt the probability part of the curriculum? Should we leave the nor-mal distribution behind? Probabilistic modelling uses many other distributions (e.g., for the analysis of risks). How to deal with other approaches and inter-pretations (e.g., Bayes).
- Simulation absorbs modelling. This may result in data as facts while probability models represent a hypothetical way of thinking.
- Re-randomisation fails to permit discussing errors of type II. Bootstrap is intuitive; yet, correction for bias is complex and it fails with (small!) tail probabilities (see Borovcnik, 2019).
- How to continue the curriculum within such a setting?

The “Informal Inference” approach narrows the views on probabilistic modelling. Biehler (2014) refers to the delicate relation between informal and formal inference, as we have to make sure for which school of statistical inference we design informal procedures as a substitute or companion. There are two negative consequences of “Informal Inference”: first, one blocks the continuation of the curriculum towards Bayesian inference, and second, one reduces the concept of probability to its frequentist part, which we noted as a negative implication of current approaches towards the didactics of stochastics. Thus, the elementarisation embedded in “Informal Inference” – as attractive as it may appear on the first sight – goes beyond the aim of teaching the next generation a clear picture of the methods of statistical inference including their complexity and their potential.

We prefer informal ways to explore situations and concepts of statistical inference to the “Informal Inference” approach. The latter reduces the complexity of statistical inference permanently and obstructs continuing learning paths towards the full approach of statistical inference. We doubt that the new developments of resampling, Bootstrap, and machine learning make statistical inference obsolete. Quite to the contrary, it will become even more important in the upcoming era as there will be an urgent need to check for bias and incompleteness of data; for that purpose, we will need traditional methods of inference. Informal ways may ground suitable learning paths to disclose the nature of statistical inference, see the paradigmatic examples of Batanero and Borovcnik (2016), Borovcnik (2019), and Borovcnik, Fejes-Tóth, Jánvári, and Vancsó (2020). These provide playful activities in line with Varga’s intention to give a challenging situation over to the learners and let them explore it. We still can enrich it by gamification elements (Papp, 2017) to boost their effects on motivation and learning. To connect to the didactical ideas of Varga, it might also be advisable to learn about the ideas of the empire of random from people who were influenced by Varga directly (see Nemetz & Kusolitsch, 1999), which certainly is true for Tibor Nemetz.

Elementarisation in the sense of Felix Klein (1908) as “Higher mathematics from an elementary standpoint” may enrich teaching. Yet, elementarisations always are on the edge of a *glissement didactique*. Remarkably, Varga (1983) introduces statistical inference in his task series for 9-year olds by a non-parametric argument but he does not insist to continue on that line or reduce inference to non-parametrics. The restricted playful situation only serves to explore the situation, the criteria, and the purpose of tools that one has to develop to find a satisfactory answer to the initial problem, which deals with a typical question and a paradigmatic context for statistical inference. After such learning experiences, the learners are ready for delving deeper into the issues of statistical inference; they are ready for learning more about the complexity of the notions and the methods.

References

Barnett, V. (1982). *Comparative statistical inference*. New York: Wiley.

Batanero, C., Chernoff, E., Engel, J. Lee, H., & Sánchez, E. (2016). *Research on teaching and learning probability. ICME-13 Topical Surveys*. Cham, Switzerland: Springer online. doi.org/10.1007/978-3-319-31625-3_1

Batanero, C., & Borovcnik, M. (2016). *Statistics and probability in high school*. Rotterdam: Sense Publishers.

Bauersfeld, H. (1988). Interaction, construction and knowledge – Alternative perspectives for mathematics education. In D.E. Grows & T.J. Cooney (Eds.), *Effective mathematics teaching, research agenda for mathematics education, Vol. 1* (pp. 27-46). Reston, VA: National Council of Teachers of Mathematics.

Bender, P., & Schreiber, A. (1985). *Operative Genese der Geometrie*. Stuttgart: Teubner.

Biehler, R. (2007). Students' strategies of comparing distributions in an exploratory data analysis context. In P.L. do Nascimento (Ed.), *Bulletin of the International Statistical Institute (ISI): 56. Session of the ISI* (6 p.). Voorburg: International Statistical Institute.

Biehler, R. (2014). On the delicate relation between informal statistical inference and formal statistical inference. In K. Makar (Ed.), *Proceedings of the Ninth International Conference on Teaching Statistics*. The Hague: International Statistical Institute.

Borovcnik, M. (2015). Risk and decision making: The "Logic" of probability. *The Mathematics Enthusiast, 12* (1-3), 113-139. scholarworks.umt.edu/tme/vol12/iss1/

Borovcnik, M. (2018). Fundamental concepts and their key properties in probability – How to identify them and provide sustaining intuitions. In M.A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics* (pp.1-6). Voorburg: International Statistical Institute. iase-web.org/Conference_Proceedings.php?p=ICOTS_10_2018

Borovcnik, M. (2019). Informal and 'informal' inference – Didactic approaches to statistical inference. In J.M. Contreras, M.M. Gea, M.M. López-Martín, & E. Molina-

Portillo (Eds.), *Actas del Tercer Congreso International Virtual de Educación Estadística* (pp. 1-17). Granada: Univ. Granada. www.ugr.es/~fqm126/civees.html

Borovcnik, M., Bentz, H.-J., & Kapadia, R. (1991). A Probabilistic Perspective. In R. Kapadia & M. Borovcnik (Eds.), *Chance Encounters: Probability in Education* (pp. 27-72). Dordrecht: Kluwer.

Borovcnik, M., Engel, J., & Wickmann, D. (Eds.) (2001). *Anregungen zum Stochastikunterricht: Die NCTM Standards 2000 – Klassische und Bayessche Sichtweise im Vergleich*. Berlin: 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., & Ossimitz, G. (1987). *Materialien zur Beschreibenden Statistik und Explorativen Datenanalyse*. Vienna: Hölder-Pichler-Tempsky.

Butler, D. (2013). When Google got flu wrong. *Nature, 494*(7436), 155-156. doi.org/10.1038/494155a

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*.

Cobb, G.W. (2007). The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education 1*(1).

Davydov, V.V. (1967). Beziehungen zwischen der Theorie der Verallgemeinerung und der Lehrplangestaltung. In E. Budilova et al. (Eds.), *Untersuchungen des Denkens in der sowjetischen Psychologie* (pp. 253-269). Berlin: Volk und Wissen.

DelMas, R. (2017). A 21st century approach towards statistical inference – Evaluating the effects of teaching randomization methods on students' conceptual understanding. In *Proceedings of the 61st World Statistics Congress*. The Hague: International Statistical Institute.

Diepgen, R. (1985a). Probleme eines Statistikunterrichts nach STRICK-Muster. *Stochastik in der Schule 5*(2), 14-20.

Diepgen, R. (1985b). Eine Aufgabensequenz zum statistischen Hypothesentesten. *Stochastik in der Schule 5*(2), 22-27; *5*(3), 17-38.

Dörfler, W. (1984). Actions as a Means for Acquiring Mathematical Concepts. In: *Proc. Eighth International Conference on Psychology in Mathematics Education* (pp. 172-180). Sydney: Mathematical Association of New South Wales.

Efron, B., & Tibshirani, R.J. (1993). *An introduction to the bootstrap*. New York – London: Chapman & Hall.

Engel, J. (2010). On teaching bootstrap confidence intervals. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Seventh International Conference on Teaching Statistics* (pp. 1-6). Voorburg: International Statistical Institute.
https://iase-web.org/Conference_Proceedings.php?p=ICOTS_8_2010

Engel, J. (2019). Statistical literacy and society: What is civic statistics? In J.M. Contreras, M.M. Gea, M.M. López-Martín, & E. Molina-Portillo (Eds.), *Actas del Tercer Congreso International Virtual de Educación Estadística* (pp. 1-17). Granada: Univ. Granada. www.ugr.es/~fqm126/civeest.html

Fischbein, E. (1987). *Intuition in science and mathematics. An educational approach*. Dordrecht: Reidel.

Fischer, R. (1984). Offene Mathematik und Visualisierung. *mathematica didactica 7*, 139-160.

Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Dordrecht: Reidel.

Gigerenzer, G. (2002). *Calculated risks: how to know when numbers deceive you*. New York: Simon & Schuster.

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

Glasersfeld, E.v. (1991). *Radical constructivism in mathematics education*. Dordrecht: Kluwer.

Gosztonyi, K. (2015a). *Tradition and reform in mathematics education during the „New Math" period: A comparative study of the case of Hungary and France*. PhD Dissertation. Universities Paris Diderot and Deberecen.

Gosztonyi, K. (2015b). The 'New Math' reform and pedagogical flows in Hungarian and French mathematics education. In K. Krainer & N. Vondrová (Eds.), *Proc. Ninth Congress of the European Society for Research in Mathematics Education* (pp. 1709-1716). Dortmund: ERME.
www.mathematik.uni-dortmund.de/~erme/index.php?slab=proceedings

Gosztonyi, K. (2017). Understanding didactical conceptions through their history: A comparison of Brousseau's and Varga's experimentations. In T. Dooley & G.

Gueudet (Eds.), *Proceedings of the Tenth Congress of the European Society for Research in Mathematics Education*. Dortmund: ERME. www.mathematik.uni-dortmund.de/~erme/index.php?slab=proceedings

Gosztonyi, K. (2019, Nov.). Tamás Varga's reform movement and the Hungarian "Guided Discovery" approach. Plenary lecture, *Varga 100 – Connecting Tamás Varga's 'Legacy and Current Research in Mathematics Education*. Budapest. varga100.sciencesconf.org/resource/page/id/8

Hassad, R.A. (2020). A foundation for inductive reasoning in harnessing the potential of Big Data. *Statistics Education Research Journal, 19*(1), 238-258.

Hawkins, A.S. (1991). Projects, practical work and competitions. In D. Vere-Jones (Ed.), *Proc. Third International Conference on Teaching Statistics, Vol 1* (pp. 209-213). Voorburg: International Statistical Institute. iase-web.org/Conference_Proceedings.php?p=ICOTS_3_1990

Heitele, D. (1975). An epistemological view on fundamental stochastic ideas. *Educational Studies in Mathematics 6*, 187-205.

Howell, D. (n.d.). Resampling statistics: Randomization & Bootstrap. *Statistical page of D. Howell*. www.uvm.edu/~dhowell/StatPages/Resampling/Resampling.html.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Kapadia, R., & Borovcnik, M. (Eds.) (1991). *Chance encounters: Probability in education*. Dordrecht: Kluwer.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1). doi.org/10.1177/2053951714528481

Klein, F. (1908/2004). *Elementary mathematics from a higher standpoint. Vol. I: Arithmetic, algebra, analysis, Vol. II: Geometry, Vol. III: Precision mathematics and approximation mathematics*. Berlin: Springer.

Lunneborg, C.E. (2000). *Data analysis by resampling: concepts and applications*. Pacific Grove, CA: Duxbury Press.

MacKenzie, D.A. (1981). *Statistics in Britain 1865-1930 – The Social Construction of Scientific Knowledge*. Edinburgh: Edinburgh University Press.

Martignon, L., & Krauss, S. (2009). Hands-on activities for fourth graders: A tool box for decision-making and reckoning with risk. *International Electronic Journal of Mathematics Education 4*(3), *Special issue on "Research and Developments in Probability Education"*, 227-258. www.iejme.com/volume-4/issue-3

Müller, A. (1989). Parameterfreie Verfahren zur Beurteilung von zwei abhängigen Stichproben im Unterricht. *Stochastik in der Schule 9*(1), 41-54.

Nemetz, T., & Kusolitsch, N. (1999). *Guide to the empire of random*. Budapest: TypoTex.

NOC (1994). *Proceedings of the Fourth Conference on Teaching Statistics*. Voorburg: International Statistical Institute. iase-web.org/Conference_Proceedings.php?p=ICOTS_4_1994

Noether, G. (1967). *Elements of noparametric statistics*. New York: Wiley.

Papp, T.A. (2017). Gamification effects on motivation and learning: Application to primary and college students. *International Journal for Cross-Disciplinary Subjects in Education 8*(3), 3193-3201.

Piaget, J., & Inhelder, B. (1951). *La genèse de l'idée de hasard chez l'enfant*. Paris: Presses Universitaires de France. *The origin of the idea of chance in children*. London: Routledge & Kegan Paul.

Prodromou, T. (2017). *Data visualization and statistical literacy for open and Big Data*. Hershey, PE: IGI Global.

Rosling, H. (2009). 200 years that changed the world. *Gapminder*. www.gapminder.org/videos/200-years-that-changed-the-world/.

Rossman, A. J. (2008). Reasoning about informal statistical inference: one statistician's view. *Statistics Education Research Journal 7*(2), 5-19.

Spiegelhalter, D. (2014, April). What can education learn from real-world communication of risk and uncertainty? Invited lecture *Eight British Congress on Mathematics Education*, Nottingham.

Steinbring, H. (1991). The theoretical nature of probability in the classroom. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters* (pp. 135-167). Dordrecht: Kluwer.

Stohl Lee, H., Angotti, R.L., & Tarr, J.E. (2010). Making comparisons between observed data and expected outcomes: students' informal hypothesis testing with probability simulation tools. *Statistics Education Research Journal 9*(1), 68-96.

Strick, H.-K. (1980). *Einführung in die beurteilende Statistik*. Hannover: Schrödel.

Strick, K.-H. (1985). Anmerkungen zu R. Diepgen: Probleme eines Statistikunterrichts nach STRICK-Muster. *Stochastik in der Schule 5*(2), 21.

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

Tall, D., & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics 12*, 151-169.

Vancsó, Ö. (2001). Klassisch und Bayesianisch gedacht – Erfahrungen mit meinen Studenten. In M. Borovcnik, J. Engel, & D. Wickmann (Eds.) (2001), *Anregungen zum Stochastikunterricht: Die NCTM Standards 2000 – Klassische und Bayessche Sichtweise im Vergleich* (pp. 171-176). Berlin: Franzbecker.

Vancsó, Ö. (2006). *Classical and Bayesian statistics from didactical point of view*. Dissertation: University of Debrecen.

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

Vancsó, Ö. (2018, July). How visualisation using software helps understanding classical and Bayesian statistics. *Invited paper "Teaching Probability in School – Understanding and Linking it to Statistics."* ICOTS 10, Kyoto. www.researchgate.net/profile/Oedoen_Vancso.

Varga, T. (1970). Probability through games. A sample of three games. In ICMI (Ed.), *New trends in mathematics teaching II* (pp. 424-440). Paris: UNESCO.

Varga, T. (1972). Logic and probability in the lower grades. *Educational Studies in Mathematics* 4(3), 346-357.

Varga, T. (1982). New topics for the elementary school math curriculum. In Th. C. O'Brien (Ed.), *Toward the 21st Century in Mathematics Education* (pp. 12-34). Edvardsville, IL: Teachers' Center Project, Southern Illionis University at Edvards-ville.

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

Vygotsky, L.S. (1967/2004/2012). *Voobrazhenie i tvorchestvo v detskom vozraste*. Moscow: Prosveshchenie. English translation: Imagination and creativity in childhood. *Journal of Russian and East European Psychology, 42*(1), 7-97.

Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes* (Ed. M. Cole, V. John-Steiner, S. Scribner, & E. Souberman). Cambridge, MA: Harvard Univ. Press.

Wickmann, D. (1990). *Bayes-Statistik*. Mannheim: Bibliographisches Institut.

Winter, H. (1992). Zur intuitiven Aufklärung probabilistischer Paradoxien. *Journal für Mathematik-Didaktik* 13, 23-53.

Wittmann, E. (1973). *Grundfragen des Mathematikunterrichts*. Braunschweig: Vieweg.

MANFRED BOROVCNIK
UNIVERSITY OF KLAGENFURT, INSTITUTE OF STATISTICS, AUSTRIA
E-mail: manfred.borovcnik@aau.at