13th International Congress on Mathematical Education
Hamburg, 24-31 July 2016

PROFESSIONAL DEVELOPMENT FOR TEACHING PROBABILITY AND INFERENCE STATISTICS WITH DIGITAL TOOLS AT UPPER SECONDARY LEVEL

Rolf Biehler

Universität Paderborn, Germany

A 5-day long professional development course for teachers of grade 10 to 12 is presented. It will be illustrated how the course is rooted in scientific knowledge from probability education and in principles for designing professional development courses for teachers, as have been developed in the German Centre for Mathematics Teacher Education (http://www.dzlm.de/). Experiences from implementations and evaluations are reported. The content "probability, simulation and the laws of large number" will be used for concrete exemplification of activities for students and teachers.

CONTEXT OF THE PROJECT

The following paper describes the conceptual basis of a professional development course for teachers who are supposed to teach stochastics (the German name for probability and (inference) statistics) in grade 10 – 12. The teachers are from in the federal state of North Rhine-Westfalia, where, from 2014 onwards, a new state syllabus made teaching stochastics obligatory which is also included in the final central examination (Abitur). The syllabus are based on the new national standards (KMK, 2012), where stochastics has also a prominent obligatory place. A further new feature is the obligatory use of graphic calculators (GC), which can and have to be used in the final examinations, too. However, final examinations contain a "tool-free" part, where calculators are not allowed to be used. The digital tools are supposed to support teaching stochastics by providing a tool for simulation, calculation and interactive visualizations. As part of the German Centre for Mathematics Teachers Education (DZLM) we developed a four day course "stochastics compact" spread over about four months for supporting a sustainable professional development with impacts into actual classroom teaching. The course was offered three times since 2013 and reached more than 270 teachers (see Oesterhaus & Biehler, 2014 for a description of version 1 and 2 of the course). In each round, the course was improved on the basis of the feed-back provided by the teachers and by evaluations. Currently, in spring 2016, we are piloting the 4th version of the course, which was extended to 5 days is now further developed together with the regional school administration from one of the 5 administrative regions in North Rhine-Westfalia, namely Arnsberg. The new development team for the 4th version consists of myself, one doctoral student and one post graduate researcher, who had been a teacher after his Ph.D. for more than 10 years, from the University of Paderborn, and 5 mentor teachers from the Arnsberg region who have also a long-standing experience in designing and implementing PD courses themselves. We aim at offering the course for participants of all 130 schools in the Arnsberg region, who teach courses in grade 10 to 12. The collaboration can be considered as a kind of professional development for multipliers as well directed at the 5 mentor teachers, however, the collaboration is genuinely interactive and not one – way. The new course version is a result of integrating different perspectives: the perspective of research in stochastics education, theory and experience based conception for the professional development of teachers of the DZLM, the administrations' conception of "implementing" the new curricula and the practical experience of the 5 mentor teachers with regard to teaching students and with regard to doing professional development courses for mathematics teachers. The collaboration is great experience although not tension free.

DESIGN PRINCIPLES AND CONTENT OF THE PROFESSIONAL DEVELOPMENT COURSE

For understanding the goals of the course, we have to look at three levels: the structure of content from stochastics, the design principles for teaching and learning stochastics at school that are advocated in the course, and the design principles for the professional development course itself, as a course for adult learners, namely experienced mathematics teachers.

Structure of the stochastics content

The 5 days of the course have the following contents: 1. basics of probability, simulation and the laws of large number, 2. independence, Bayes rule and expected value of random variables, 3. binomial distribution as a probability model, 4. introduction into hypothesis testing with P-values, problems with interpreting small and large p-values, 5. hypothesis testing as decision making, errors of the first and second kind, including the notion of statistical power function of a test, choice of hypotheses in one sided tests (producer and consumer risk). The list of topics reflects the curriculum, which has a certain tradition in the state as non-obligatory content. Topics 1 and 2 should be covered in grade 10, the other topics in grade 11-12. The state could have chosen confidence intervals instead of hypothesis testing, but they didn't.

We will provide more insight into how we design day 1 below. We can only briefly mention some important aspects of the other days. Independence is poorly taught. We follow the idea of Truran and Truran (1997) of two different but to be related concepts of independence, the independence of events and the independence of several random experiments that has to taught as an assumption that has to checked. The latter is the most important in modelling contexts. Bayes rule is taught with natural frequencies and related to public misuses of conditional probabilities (Krämer & Gigerenzer, 2005). Expected value is introduced on the basis of a law of large numbers for averages, which is seldomly mentioned in German textbooks, and that provides a first basic interpretation, together with the a second basic interpretation of expected value as a balance point.

P-values hypothesis testing is not a tradition in German schools but we are convinced that it is easier for students to start with p-values, and then learn fixed level significance testing by the rejection condition "p-value equal or less than □"□Our general approach to testing emphasizes interpretations and misinterpretations of p-values in authentic contexts. We recently discovered that we took into account many of the ASA positions on this topic (Wasserstein & Lazar, 2016).

Teaching and learning principles

We apply three principles for designing the course for the students: *simulation, authentic modelling, intuitive grounding of concepts*: (1) use of real experimentation and simulation throughout the course for emphasizing the statistical aspects of probability and connecting data and chance (Konold & Kazak, 2008) and including "real data" in the teaching of probability; (2) taking modelling serious, i.e. make assumptions explicit and check them if possible, use authentic contexts not only artificial one; (3) intuitive grounding of concepts, by taking into account well-known misconceptions and emphasizing basic mental models for central concepts.

Principles for teachers' learning in the course, structure of teachers' knowledge

For the participating teachers the design for adult learning follows a three step role changing model: progressing from teacher in the learner's role, role of reflecting and selecting content, to teachers' role as classroom designers. We concentrate on (1) activities and possible "knowledge summaries" from the perspective of learners, (2) reflecting from the perspective of stochastics and stochastics education, providing "background knowledge" (content knowledge (CK), pedagogical content knowledge (PCK) and technological (pedagogical) content knowledge (T(P)CK), as well as "horizon knowledge" (HK) in the sense of Loewenberg Ball and Bass (2009), see also Wassong and Biehler (2010) for a model of teachers' knowledge), (3) exemplary design of learning sequences on the basis of what was previously encountered in the course together with the teachers' own expertise. Ideally, a professional development course should allow real classroom experimentation with the material and teachers coming back for collaborative reflection on their experience. This is not possible because a co-ordination with the current teaching obligations of all the participating teachers is difficult to realize. We intend to realize such an additional component for selected school teachers after they have taken our 5 day course.

Our role changing model "teachers in the learners' role, teachers' reflecting and selecting content role, to teachers' classroom implementation role" was not undebated in the collaboration with the school administration with regard to the emphasis that has to be put on the different aspects. The position was put forward that teachers should know the content already to a large extent or learn the content themselves, their role in the selection and shaping of content is regarded as limited, as we "have" the curriculum, central examinations and text books, that have taken many decision for the teachers. An emphasis is to be put on classroom implementation focusing on coping with heterogeneous students requiring differentiation and individual support, focusing on the new "competence orientation of the syllabus" instead of "on content", learning to use modern subject matter independent teaching methods and so forth. From our perspective teaching methods have to be also intertwined with stochastic content, stochastic content knowledge is not "there" in the teachers' mind because of limited or too abstract stochastics education in their pre-service training. Moreover, we see the view of the subject matter an understanding of fundamental ideas of a topic as a part of teacher development course. The material we use in the course contains activities for students but also suggestions for classroom "knowledge summaries", which usually are part of school text books. However, we found several of them inadequate or missing, among others summaries on the meaning of probability and expected value, schematic plans for doing simulations and hypothesis testing, pitfalls when interpreting hypothesis tests.

SOME EXAMPLES FROM THE COURSE

We will illustrate our design principles *simulation, authentic modelling, intuitive grounding of concepts*: with examples from day one of the five day course.

Assumptions concerning students' knowledge in probability when entering grade 10

Out approach is based on the following assumptions, which can be backed up by an analysis of German textbooks of grade 5 to 9. We assume that students who enter grade 10 do not remember probability very deeply as this was not a strong topic in grades 5 to 9. They may know how to calculate some probabilities by systematic counting under the assumption of equal probability assumptions, they have used tree diagrams with the multiplication and addition rule to calculate probabilities in compound random experiments. They may know that in some cases such as throwing LEGO blocks or thumbtacks, probabilities can only be estimated by relative frequencies in real experiments. They will have a vague memory of the "empirical law of large numbers". Usually they will have no idea, how close a frequency estimate comes to the probability depending on the sample size n. Probably most of them believe in the "law of small numbers" (Tversky & Kahneman, 1971). They have used the multiplication rule in tree diagrams but have not come across the notion of "stochastic independence". Text books often use independence without saying, even in ridiculously inadequate situations. Students will not have encountered a convincing argument, why probabilities in a tree diagram should be multiplied to calculate the probability of a path in the diagram. Probably simulation was not a topic in the curriculum. Modelling where assumptions are made explicit and checked with data was probably not a central topic.

Basics of probability, simulation and the laws of large number –teachers' in the learners' perspective

Convincing arguments for deepening the understanding of the role of sample size and a deeper understanding of the laws of large number come from various perspectives in the discourse on stochastics education, from the psychological misconceptions research and from requirements from statistical literacy (taking sample size into account when judging claims in media). A further requirement stems from the new role of simulation with digital tools that has now become a topic in the curriculum: how accurate is this method and how many repetitions should be done. From an advanced standpoint, using relative frequencies for estimating probabilities means to use confidence intervals, so what can be an elementarized version of this notion in restarting probability in grade10?

Similar to what Konold and Kazak (2008) are suggesting, we emphasize the notion of probability distribution from the beginning and introduce a "difference game activity" where the students have to find the probability distribution of the absolute difference of the two numbers, when throwing two fair dice. Guessing in the first place (by distributing 18 buttons on the 6 possible outcomes), real experimentation, systematic counting are used and simulation is introduced as a means to get more accurate estimates. The law of large numbers shows up in a new way as the stabilization of the whole distribution of the relative frequencies and not only individual freuquencies. As the probabilities can be calculated theoretically, a first rough idea of how close the relative frequencies come to the theoretical probabilities is conveyed, when sample sizes of 100, 1000 and 10000 are used. A second activity is the 10-20-test activity, which has the same structure as Kahneman & Tversky's famous maternity ward problem: students should decide, which test is easier to pass by just guessing, when all questions have two possible answers and one test has 10 questions and one has 20 questions, or whether both tests are equally likely to pass. Guessing the correct answer always leads to different answers in all groups of students and of teachers where have used this. Students have not the means for theoretical calculation (the binomial distribution comes much later) so simulation by throwing a coin an/or using a digital tool I snecessary. Simulation could only be used for just determining the passing probabilities in both test, but our activity focusses at putting the *distributions* of the *number* and of the *proportion* of correctly guessed answers (see Fig. 1).

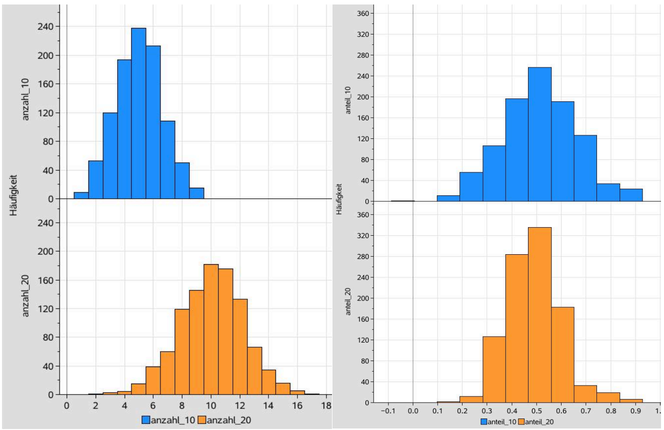

Fig.1 Simulated distributions of the *number* of correct guesses (left) and the proportion (right) of correct guesses in the 10-20-test activity

The visualization (and corresponding mental model) that should be conveyed is that the distribution of the *proportion* in the 20-test has less spread around p =0,5 than has the distribution of the proportion in the 10-test (whereas the spread increases from 10 to 20 when we focus on the number of successes). As the tests had been simulated by coins as well, the graphs in Fig. 1 can be reinterpreted as the distributions of the number or proportion of "heads" when throwing a die 10 or 20 times. This idea can be expended to sample size up to 600, where we calculate the spread of the "middle 95%" by using percentiles in the simulated distribution (see Fig. 2 left). Students can experimentally discover that (Fig. 2 right) that the 95%-middle-spread decreases with 2/squareroot –of-n. "So" we expect a relative frequency (of heads) with 95 % probability in the interval $0.5 \pm \frac{1}{\sqrt{n}}$.

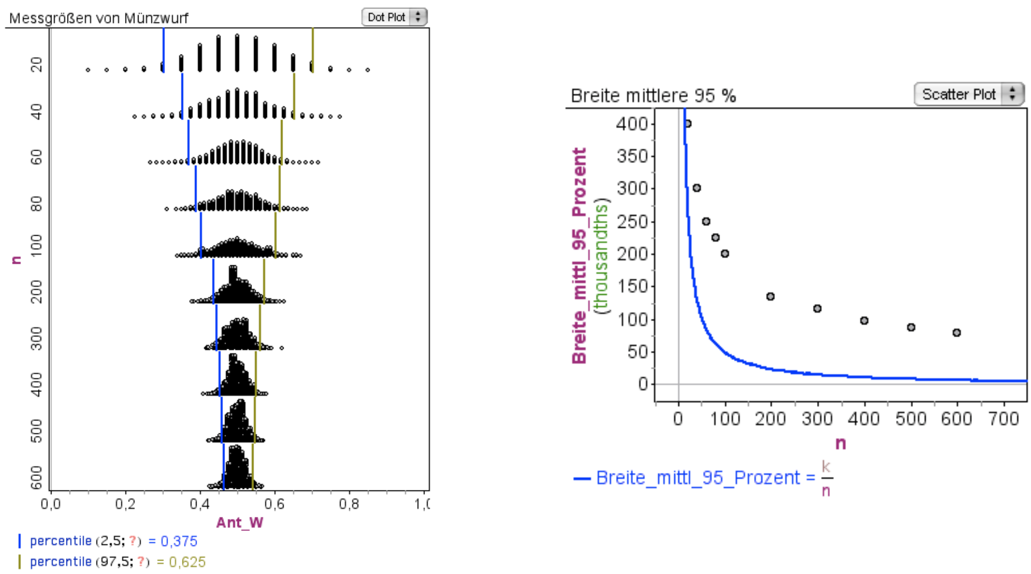

Fig.2 Decrease of the spread of the middle 95% with sample size (left) and searching for a function that predicts the dependence of the middle-95%-spread from sample size

Our students and teachers are not asked to use a digital tool to produce Fig. 2. in the classroom but work with the static picture focusing on sense making of the graph. This illustrates our "3-level-approach to simulation": simulation can be done on three levels: by students themselves, by teacher demonstrations or by analyzing results of simulations given as static pictures or as a prepared interactive visualizations. Teachers have to learn to adequately select from these options.

After some further activities, students learn that the equation $|p - f_n| \leq \frac{1}{\sqrt{n}}$ (that we call the $\frac{1}{\sqrt{n}}$ -law) can be interpreted in two ways: as an interval prediction about the relative frequency f_n when p is known and an interval estimation of p when f_n is known. Both statements hold with 95 % probability. We think that this is an adequate elementarization for students as a first step with practical value for dealing with decisions about sample size when dealing with simulations. As part of teachers' background or horizon knowledge (Loewenberg Ball & Bass, 2009) we explain how this formula is related to confidence intervals and communicate the precise frequentist interpretation of the 95% confidence probability. We also provide the mathematical background on the basis of the binomial and normal distribution from which we can derive a 95% -prediction interval as $p \pm 1.96\sqrt{\frac{p(1-p)}{n}}$, to which "our" interval is a good approximation(see Biehler and Prömmel (2013) for detailed learning trajectory for the this). According to the principle of authentic modelling, we suggest to introduce a further activity , the "North Rhine-Westfalian newborn boys activity" with real data from this state. Fig.3 shows one example from a series of visualizations we use.The scatter plot illustrates that male proportion is much more variable in small communities and the spread decreases with 1 over the square root of the "number of children born".

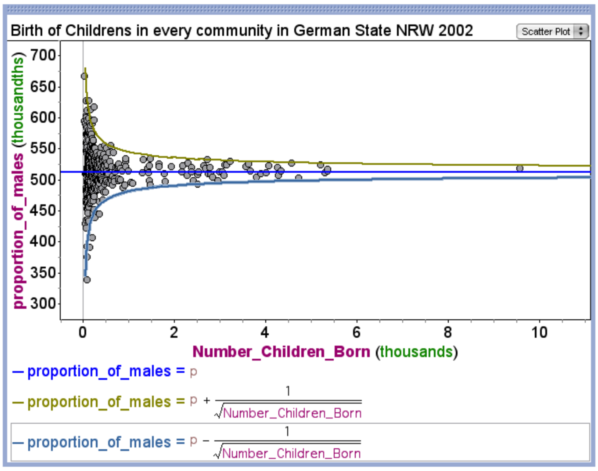

Fig. 3 Each point represents a community in the state of NRW, the proportion of male newborns in the year 2002 is shown on the vertical axis, the horizontal axis shows the number of newborn children in that community. A horizontal line is drawn at the male birth probability of 0.514, that was estimated from the data set.

Basics of probability, simulation and the laws of large number –teachers in the teachers' perspective

The *"three step role changing model"* for the participating teachers is realized as follows. With regard to reflecting and selecting content, we initiate a discussion, when and what to teach from the above content in the beginning of the grade 10 course and what could be postponed later when having taught the binomial distribution. Time constraints in the everyday classrooms and belief systems of the teachers interact in these discussions. It is not surprising that there is a high heterogeneity in every group of teachers with whom we discuss. We do perform a systematic study about our teachers' beliefs and their change during our course. Here we will provide some anecdotal evidence. Some teachers are not convinced of the elementarization and want to teach the "correct mathematics" directly or not at all, some final examination oriented teachers doubt that the square-root-n-law will have the importance that we have assigned to it, so the teaching time is not justified in their eyes. Teachers coming from a statistical or mathematical literacy point of view are more easily convinced to put emphasis on the role of sample size, even if this might not be directly relevant in final examinations. We also initiate a discussion and reflection of the role of digital tools, in particular the graphic calculator (GC), which is obligatory. In our course itself, the teachers have learned how to use GC for simulations and interactive visualizations, learning on the level of technological knowledge and technological content knowledge (TK and TCK). This reflects the state of knowledge and needs with which the teacher enter into our course. The above mentioned 3-level-approach to simulation is used to initiate a discussion on the role of simulation for learning and teaching probability as part of their emergent technological pedagogical content knowledge (TPCK). Heterogeneity of teachers plays an important role here as well. The technological knowledge with regard to the GC varies as well as the attitude towards digital tools in general and GC in particular. In general, we started our course with the belief, how important the use of digital tools and simulation in teaching and learning probability is, based on research evidence as has been elaborated elsewhere (Biehler, Ben-Zvi, Bakker, & Makar, 2013). However, this research gives most convincing results, when tools such as Fathom and Tinkerplots are used that are well-adapted and well-based on the needs of stochastics education. Graphic calculators are very different tools, even when we take into account that the TI Nspire has integrated building blocks from Fathom and Tinkerplots. For instance, it is extremely easy for students to simulate the 10-20-test problem with Fathom and Tinkerplots, without needing knowledge about the binomial distributions. Simulation in the TI Nspire or the Casio fx20 needs "binomial random numbers" or technical expertise including nested commands (see Fig. 4). Ideally teachers technological pedagogical content knowledge (TPCK) should be based on reflected experience with different digital tools, but as the GC is prescribed and time in the professional development course is limited, we focus on the GC and discuss what can be done on the three levels of simulation. For teachers, who want to use other additional tools we provide files and background knowledge for using Fathom and Geogebra.

```
=seq(countif(randsamp(ausgang,10),"richtig"),n,1,2500)
```

Fig. 4: TI Nspire: How to draw 10 "balls" from a box (called "ausgang") with 2 balls named "richtig" (correct) or "falsch with replacement and then the number of "richtig" among the 10 balls and then repeat the process 2500 times, put the simulated data into a column for a further analysis of the distribution.

As a third step, teachers are asked to (re-) design a learning sequence for re-starting probability in grade 10, based on their previous knowledge and experience, the text book they use in their schools and the new ideas they have learned in the course and the material (activities, knowledge summaries) that were presented. Initial results of our evaluation show again a vast heterogeinity of our participating teachers. The opinions are divided whether such activities are worth to do in the limited time of such a professional development day. Some teachers wish "more input" instead, feeling confident to implement what they have learned themselves, others appreciate the collaborative work with colleagues on re-designing learning sequences. Attitudes of course are also dependent on the concrete form of the "designing learning sequences activity". Moreover, we will offer other activities and tasks for those teachers who do not like the "designing learning sequences" tasks and prefer to deepen their technological knowledge concerning GCs or deepen their background and horizon knowledge. It is clear that attitudes differ with regard to where they need most support from the team of teacher educators during the meeting and where they feel confident to learn themselves at home and after the course.

CONCLUDING REMARKS

The piloting of of our course will be finished in June 2016. Afterwards, the course will be extended to reach all 130 schools of the region. We are designing instruments for studying the effect of our courses on teachers' attitudes and knowledge and intend to look into selected classrooms.

References

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for Enhancing Statistical Reasoning at the School Level. In M. A. Clements, A. J. Bishop, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Third International Handbook of Mathematics Education* (pp. 643-689): Springer New York.

Biehler, R., & Prömmel, A. (2013). Von ersten stochastischen Erfahrungen mit großen Zahlen bis zum 1/√n – Gesetz – ein didaktisch orientiertes Stufenkonzept. *Stochastik in der Schule, 33*(2), 14-25.

KMK. (2012). *Bildungsstandards im Fach Mathematik für die Allgemeine Hochschulreife (Beschluss der Kultusministerkonferenz vom 18.10.2012)*.

Konold, C., & Kazak, S. (2008). Reconnecting Data and Chance. *Technology Innovations in Statistics Education, 2*(1).

Krämer, W., & Gigerenzer, G. (2005). How to Confuse with Statistics or: The Use and Misuse of Conditional Probabilities. *Statistical Science, 20*(3), 223-230.

Loewenberg Ball, D., & Bass, H. (2009). With an Eye on the Mathematical Horizon: Knowing Mathematics for Teaching to Learners' Mathematical Futures. In M. Neubrand, et. al. (Ed.), *Beiträge zum Mathematikunterricht 2009*. Münster: WTM-Verlag.

Oesterhaus, J., & Biehler, R. (2014). Designing and implementing an alternative teaching concept within a continuous professional development course for German secondary school teachers In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA*. Voorburg, The Netherlands: International Statistical Institute, iase-web.org.

Truran, J. M., & Truran, K. (1997). Statistical Independence - One Concept or Two? Implications for Research and for Classroom Practice. In B. Phillips (Ed.), *Papers on Statistical Education presented at ICME-8 (International Congress on Mathematics Education-8) Seville, Spain, July 14-21*. Hawthorn, Australia: Swinburne Press.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*(2), 105-110.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: context, process, and purpose. *The American Statistician, online first*. doi:10.1080/00031305.2016.1154108

Wassong, T., & Biehler, R. (2010). A model for teacher knowledge as a basis for online courses for professional development of statistics teachers. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia. Voorburg, The Netherlands: International Statistical Institute*.