Zetetiké

DOI: 10.20396/zet.v28i0.8657607

Contexts, analogies, and tasks that expose the purpose of the key concepts of probability

Contextos, analogías y tareas, que exponen el propósito de los conceptos clave de probabilidad

Manfred Borovcnik [1]

Abstract

Probability is and will continue to be a virtual concept. This specificity requires meta strategies that go far beyond the instruction of the mathematical details and ask for a sensible use of simulation. We suggest focussing teaching efforts on explicitly exposing the students to the purpose of the concept of probability. The purpose shows the character of probability indirectly as the required steps for solving a task make the properties appear natural in the context. We elaborate suitable tasks and interactive animations, which are designed to overcome learning obstacles. We focus on three aspects of probability: The character of probabilistic statements, the transparent use of probability for decisions under risk, and informal inference considerations in the early probability education. An essential criterion of teaching is how far it allows learners a more direct access to the concepts on an intuitive level. For designing didactic animations, our principles are characterised by the following ideas: A dynamic change is explored in comparison to the initial situation. Like watching a video, one looks at the different stages of emergence of a relation between the investigated concepts.

Keywords: Interplay of intuitions; Purpose of concepts; Probabilistic thinking; Dynamic animations.

Resumo

La probabilidad es y seguirá siendo un concepto virtual. Esta especificidad requiere meta-estrategias que van mucho más allá de la instrucción de los detalles matemáticos y piden un uso sensato de la simulación. Sugerimos enfocar los esfuerzos de enseñanza en exponer explícitamente a los estudiantes al propósito del concepto de probabilidad. El propósito muestra el carácter de probabilidad indirectamente como los pasos requeridos para resolver una tarea hacen que las propiedades parezcan naturales en el contexto. Elaboramos tareas adecuadas y animaciones interactivas, que están diseñadas para superar los obstáculos de aprendizaje. Nos centramos en tres aspectos de la probabilidad: El carácter de las afirmaciones probabilísticas, el uso transparente de la probabilidad para las decisiones bajo riesgo y las consideraciones de inferencia informal en la educación temprana de probabilidad. Un criterio esencial de la enseñanza es hasta qué punto permite a los alumnos un acceso más directo a los conceptos en un nivel intuitivo. Para el diseño de animaciones didácticas, nuestros principios se caracterizan por las siguientes ideas: Se explora un cambio dinámico en comparación con la situación inicial. Como si se tratara de ver un vídeo, se observan las diferentes etapas de la emergencia de una relación entre los conceptos investigados.

Palabras clave: Interacción de intuiciones; Propósito de los conceptos; Pensamiento probabilístico; Animaciones dinámicas.

Submetido em: 22/11/2019 – **Aceito em**: 27/01/2020 – **Publicado em**: 03/03/2020

[1] Doctor in Statistics and its Didactics from the University of Klagenfurt (AAU). Professor at the Department of Statistics, Alpen-Adria Universität Klagenfurt (AAU), Klagenfurt, Austria. E-mail: manfred.borovcnik@aau.at

DOI: 10.20396/zet.v28i0.8657607

1 Introduction – Identifying the key concepts

Despite all educational efforts to materialise the concept of probability through the extensive use of simulations to illustrate its impact in repeated applications, probability is and remains a virtual concept. Not only Spiegelhalter (2014a, b) addressed probability as a metaphor; that is, we need qualitative visual images and conceptions in our mind to talk about probability and understand its meaning. An indicator of the difficulties with probability can also be seen in the abundance of misconceptions (e. g., Kahneman, Slovic, & Tversky, 1982) and paradoxes (Székely, 1986). There are several studies and monographs that reveal the complex field of probability as seen from a broader perspective of philosophy; to name but a few here: Batanero, Henry and Parzysz (2005), Batanero, Chernoff, Engel, Lee and Sanchez (2016), and Bennett (1999). On the diversity of educational responses, the reader can be guided by Chernoff and Sriraman (2014), focusing from the general theme on probabilistic thinking, and Jones, Langrall, and Mooney (2007), with research on probability compared to classroom realities, i. e., the abundant problems known in the teaching of probability.

The didactic intentions to explain or clarify the nature of probability are multiple. Steinbring (1991) develops the idea of probability as a theoretical concept that falls apart and is left without a deeper meaning if only specific aspects are handled within a teaching approach. Steinbring explicitly referred to the negative implications of focusing too narrowly on Laplace’s equiprobability and neglecting the aspect of relative frequencies and vice versa. Carranza and Kuzniak (2008) speak of the duality of probability and refer to the inseparable aspects of relative frequencies and subjectivist ideas of probability; they outline the serious disadvantages of neglecting qualitative (subjectivist) aspects of probability.

We can use the term complementarity rather than duality to highlight the inseparable connection of a concept with its various parts. The term originates from the discussion of quantum theory in the 1920s; it has been introduced by Niels Bohr who stated that complete knowledge of phenomena in atomic dimensions requires a description of the properties of both waves and particles and that it cannot be reduced to a single aspect without completely losing meaning (Bohr, 1934/1927). Complementarity has been transferred to the didactics of mathematics in the 1980s to describe the didactic dilemma resulting from the inseparable aspects of concepts (see Borovcnik, 1992).

Another complementarity of probability can be seen in its close interrelationship with statistical inference. A revolution in stochastic education occurred in the late 1980s, as it witnessed the first serious approaches to statistical inference, when researchers recognised the strong connection between probability and statistical inference (Borovcnik, 1992): Descriptive statistics were no longer considered sufficient to clarify what the role of probability should be in teaching; instead, the meaning of probabilistic concepts is expanded and clarified within an inferential framework: probability without statistical inference methods is void of meaning, while statistical inference cannot be understood without probability. The purpose of probability – in this sense – has always been to design statistical

DOI: 10.20396/zet.v28i0.8657607

inference methods that allow partial knowledge (on a “sample”) to be generalised to a larger entity (called a population, which could be described by a probability distribution).

Gigerenzer (2002) is a milestone in linking probability to risk, which has been a twin partner to probability since its early development. The relationship between probability and risk is so close that, after a more detailed inspection, its aspects could not be separated, indicating once again a complementarity. Borovcnik (2016) even goes so far as to redesign probability literacy in the light of risk. Reformulating the purpose of probability as finding or inventing methods to find reasonable decisions under risk greatly improves the perception of probability.

In order to overcome the negative impact of unilateral approaches to the teaching of probability and statistics (and other disciplines of science and mathematics), Fischbein developed his ideas on primary and secondary intuitions (Fischbein, 1975, 1987). For the didactic use of developing an interaction between the raw primary intuitions that people have before any teaching and the exposure to mathematical tasks that require the application of concepts that students have not yet developed, one can return to the work of Fischbein or Borovcnik (1992).

In Borovcnik (2019b), a theoretical framework has been established to identify the fundamental ideas with probability and to develop sustainable intuitions. The eminent role of the interaction between raw primary intuitions and emerging secondary intuitions that students acquire through their continuous work with tasks has been advocated. Secondary intuitions are the ultimate goal of teaching efforts: such intuitions not only neutralise misleading primary intuitions (which would otherwise be highly resistant to change through teaching interventions), but also support a holistic (“ganzheitlich”) understanding of solutions through mathematical methods. Such a type of understanding would also work without knowing every mathematical detail of how the task was solved. A link between mathematical concepts and secondary intuitions can be established by analogies (on the role of analogies, see Simons, 1984).

Making the purpose of the concepts to be learned would also introduce a boost in the understanding and acceptance of the new concepts and the methods that are based on them. The purpose of a concept (if known) leads to immediately focus on the required learning steps; it makes the steps understandable and makes them appear “natural”. The mathematics to be learned acquires the character of a tool and one becomes more familiar with the tool, the more one uses it. Working with the concept as a tool to serve a purpose also avoids endless discussions of how a concept can be understood (which can be really confusing, especially for early learners) by showing what the concept can be used for. If the purpose seems reasonable to the students, they will be more willing to follow the learning steps required to complete their cognitive network for their conceptual understanding of concepts that have already been widely used by them.

This also follows more closely the way in which the concepts have been used at the stage of their emergence: they have been applied implicitly for a specific purpose. It is a pity

DOI: 10.20396/zet.v28i0.8657607

that we have lost the connection with the origin of probability and its conceptualisation within physics, since probability genuinely serves the purpose of understanding physical phenomena by structuring them through emerging theories, and probability has been a cornerstone of the theorisation of physics (see, for example, von Plato, 1994).

Reviving that high level of concepts, to focus on the purpose of the concepts, can be seen as a primary objective of teaching interventions. We use a hermeneutic form of argument, and relate the fundamental ideas of probability to their mathematical and philosophical background; the aim of this work is to make the purpose of probability transparent by developing tasks that are appropriate to that "purpose". As probability is a theoretical concept, we introduce the idea of the purpose of the concept and of several tasks that can be handled with it. It may be better to make it clear for what purpose the concept is to be used rather than to explain and elaborate on the question what properties the concept might have. The ultimate goal of teaching interventions should be to provide students with *intuitive access to key concepts*:

- Explore the type and quality of probability information (Chapter 2).
- Purpose of probability – to compare risks and prepare decisions (Chapter 3)
- Explore a probabilistic model for making better decisions (Chapter 3)
- Measure or estimate an unknown quantity (Chapter 4).
- Learning from theory – central theorems.
- Update of a qualitative probability judgment by data (Bayesian approach).

The first point touches on the core character of probability statements about which there is so much confusion. The second and third points (and the last one as well) refer to decision making and a qualitative component of probability and refer to obtaining and updating qualitative information on the credibility of statements or events. The fourth point establishes an early connection of the concept of probability with statistical inference, which is one of the main purposes of probability: to justify decisions between options modelled by probability. The fifth point prepares a connection between primary intuitions and secondary intuitions (Fischbein, 1975, 1987) and provides a suitable structure for the intuitions that should arise from the teaching of parts of the mathematics behind probability and statistics. It should also strengthen the connection between probability and statistical inference, which in turn revises conceptions of probability in a "last educational step".

The last two points in the above list will not be dealt with further in this document. The project "Fundamental concepts and their key properties in probability" has other articles in preparation on these aspects, which are also contained in Batanero and Borovcnik (2016), Borovcnik (1992, 2015a), or Vancsó (2009, 2018).

In this paper, we address the students or trainees in a generic way. However, we definitely target students in the final years of the gymnasium (high school). We have

DOI: 10.20396/zet.v28i0.8657607

experimented with the ideas presented with non-mathematics students at the college level (these students are no more inclined to learn mathematics nor are they more talented or capable than the target group). The contexts and tasks discussed here are comparable to those of Batanero and Borovcnik (2016), *Statistics and Probability in High School*. The tasks originate from a broader project on Fundamental Ideas in Stochastics. A key objective of this project is to develop meta-knowledge for students that can partly replace the sometimes quite complex mathematical relations and theorems of the discipline.

Probability is defined mathematically by means of axioms and the axiomatic basis of the concept justifies its interpretation. There are two main interpretations of probability, the frequentist (FQT) and the subjectivist (SJT), mathematically put on a solid basis by Kolmogorov (1933) or by Finetti (1937/1992). The third relevant interpretation is the concept of equiprobability (Laplace, or APT, a priori probability theory), which has been a leading idea in the development of physics (especially in thermodynamics) but which has not been able to be used for a solid basis of the concept of probability from a mathematical point of view. See Borovcnik and Kapadia (2014) for a discussion of these interpretations of probability from an educational point of view and for the special notation that is used here (FQT, SJT and APT).

The purpose of probability in the context of a task makes extensive use of one or more of these interpretations of probability, which can coexist even in the same situation (making probability no easier to understand). Considerations on the quality of probabilistic information (Chapter 2) refer to Laplace's equiprobability (APT); the comparison of risks and the search for better decisions refers to utility and the subjectivist interpretation of probability (SJT). The visualisations in the tasks about the measurement of an unknown probability are more explicitly linked to probability as a frequentist concept (FQT). The work being carried out within the framework of the above-mentioned project on fundamental ideas is an elaboration on statistical inference, which is openly divided between the frequency-based (classical inference) and subjectivist (Bayesian inference) methods.

As an additional observation, it is important to note that when we use the simulation method, we generate frequencies linked to related probabilities that – in the first view – connect the illustrations and the interpretation of probability obviously to a frequentist concept. However, the simulation method is widely used within the Bayesian-inference approach to determine solutions (also in the form of probability distributions). This means that despite the view of probability that one has (and this view is qualitative or subjectivist in the Bayesian framework), one can take advantage of the simulation method. Applying the simulation carries with it the possibility of being misinterpreted by the students, who could then associate frequentist aspects with subjectivist concepts. Therefore, the simulation method has to be used with care to avoid such a change in the probability connotation to a frequency (and therefore biased) interpretation corner.

With the tasks, which are discussed in the following sections, we cover all aspects of probability including the various conceptions and touch on the aspects of probability that

DOI: 10.20396/zet.v28i0.8657607

build the foundation of the key properties of the concept of probability. As we have stated, our aim is for students to understand these key properties, that are arising naturally from the context and purpose of the tasks used.

2 Exploring the type and quality of probability information

Probability is a theoretical concept and a specific value of the probability of an event under scrutiny is simply a descriptive number, an index that should convey a *potential*. This reference information is expressed differently in real situations. It is essential to have a sense of how large the size of the variation is; in other words, how large the margin of variation or error of the relative frequencies (that are the only instrument available) is so that we can materialise the index, which we call probability.

We show a simulation scenario that improves the notion of margin of error, which provides a stable intuition on the effect of the length of a random series: the larger the database, the smaller the variation in relative frequencies. While the first experiment is situated in the initial phase of teaching probability, the second reveals the variation of a sample from a normal distribution. Contrary to general expectations, samples of a distribution resemble the population in shape only if the sample size is very large.

2.1 The margin of error for random events

A major task in the early stages of probability education is to clarify and support the relationship between probabilities and relative frequencies of events in repeated trials of a randomised experiment. This relationship is justified later in the course by the Law of Large Numbers, which describes a sophisticated connection, which is completely different from convergence in calculus, and which always attracts misleading ideas about convergence. We explore variation in randomness (not the convergence) to improve understanding of the type and quality of probability information rather than to illustrate a law of convergence, which cannot be done in any finite series. Our goal is to show a pattern of reduction in the variability of relative frequencies. A suitable experiment for this purpose is to select random digits from 0 to 9.

A basic difficulty may be identified in the issue of understanding the type of information inherent in a probability statement. What does a probability of ½ really mean? It can be linked to a fair decision between two possibilities. On the other hand, a situation without preference is too easily linked to equal probabilities; this equiprobability bias of Lecoutre (1992) seems more inciting if there are two possibilities. The best way to clarify this issue is through open discussion in the form of an empirical interview with varying situations (see Borovcnik & Peard, 1996).

How does this expression relate to repeated experiments? It is vital *to avoid a focus, which is too strong on the pattern in which the series develops*, since if the hypothesis of randomness applies, each pattern is equally likely, unless one considers the pattern as a class (a group) of outcomes. There is a lot of research on random sequences and judging the

Zetetiké

DOI: 10.20396/zet.v28i0.8657607

probability of such sequences that have to be compared (e. g, Chernoff, 2013, or Borovcnik & Bentz, 1991, 1990, 2003). For probability as a concept, speculation on patterns is irrelevant as long as the hypothesis of randomness is not considered violated, and if so, it is a rather complicated statistical test that has to be applied to a replication of the data (not the same data!) to check that assumption.

In the following experiment, an idea from Freudenthal (1972) is taken up; Freudenthal suggested investigating the margin of variation with a fixed sample size instead of illustrating an obscure limit of relative frequencies towards the unknown probability by repeating the basic random experiment over and over again. For a conceptual understanding of a probability statement it is vital to understand that the range of variation is reduced with larger series and that, at the same time, individual trials are completely subject to their random character. This also means that there is no compensation rule, which is often believed after teaching efforts to demonstrate stochastic convergence of relative frequencies towards the underlying probability.

We simulate random digits (0 to 9), each with a probability of 1/10. In the statistical laboratory, we can guarantee that this probability is maintained for each individual digit, which excludes crude conceptions about favourites (we use random digits from 0 to 9 and not the emotionally charged numbers from 1 to 45 of the state lottery). We do not investigate the development of relative frequencies with the length of the random sequence. We set two snapshots in the random process, after 50 and after 1000 digits are generated. We investigate the distribution of the two scenarios and compared them. It has to be explicitly stated that we normally do not have such data (although there are some people who read the statistics of the numbers recently drawn in the state lottery).

What can be seen (in Figure 1) is that the digits in the process vary erratically, showing that the randomness in the background is working fully. It is enlightening to repeat the entire simulation scenarios (with 50 and 1000 digits) – as in a video – to witness that the digits that are “at the top” in one simulation, are not necessarily behind in the next, or vice versa. The frequencies of the individual digits actually vary without any pattern and once one recognises a pattern, it gets lost in the following repetitions of the simulation scenario.

What we can recognise is that the variation of the frequencies of the individual digits remains somewhat within a band (highlighted in Figure 1) that reflects the range – the size – of the variation. Rarely does a simulation scenario have frequencies outside this band. We get an impression of the margin of random fluctuation. What can also be seen is that this range of variation is considerably less with 1000 random digits than with only 50. A stable pattern of relative frequencies emerges with a full random fluctuation for each trial, which is not a contradiction when viewed from this angle. It also shows that – assuming that the sample is in fact taken at random (which is the advantage of our statistical laboratory, where we guarantee that the assumptions are met) – the larger the sample size, the lower is the variability.

DOI: 10.20396/zet.v28i0.8657607

As a thought experiment, already in this initial experiment, one can speculate on a further shrinking towards the reference line at the level of 1/10 since the probabilities for the random digits are the same from the beginning. We show a simulation in Figure 1, but the reader should note that the static image is only alive if the replay of the entire simulation scenario is played as an animated movie to demonstrate the effect just described.

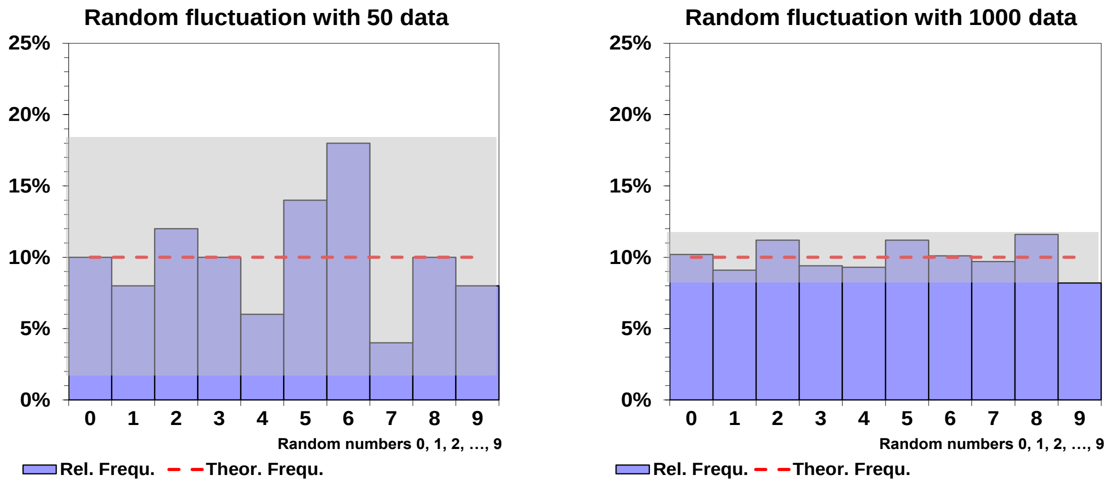

Figure 1 – Margin of the variation of random digits (1, 2, …, 9) with few and many data

Source: Prepared by the author

2.2 Sampling from a normal distribution

Random samples are used to estimate unknown parameters of the "population" (or parental distribution). Although this method already works for relatively small sample sizes, the form of the data generated does not resemble the population distribution unless we have a lot of simulated data. Next, we investigate two simulation scenarios (Figure 2):

- In Scenario 1, we generate 50 data from a normal distribution with a pre-specified mean and standard deviation.
- In Scenario 2, we investigate the same process and generate 1000 data.

DOI: 10.20396/zet.v28i0.8657607

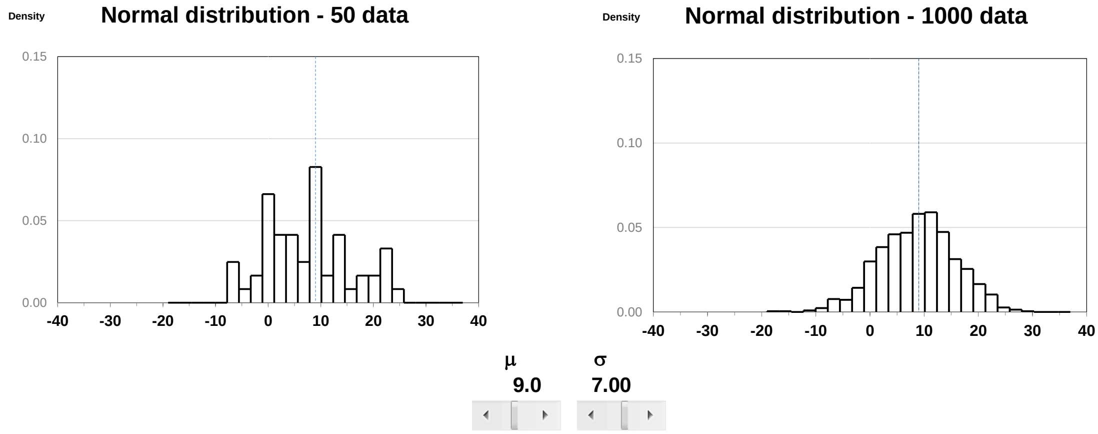

Figure 2 – Variation of samples from a normal distribution – Left with 50 data; Right with 1000 data

Source: Prepared by the author

In both scenarios, we draw a histogram to investigate the empirical distribution of the generated data. It is helpful to select other values of the mean value or the standard deviation (on the slider) to see the effect on the empirical distribution of the data: one can see that the histogram representing the generated data will shift according to the new mean value and will show less width if the standard deviation of the model is reduced. What is more important is to see the great variety in the shape of the empirical distribution of the data, which is only remotely reminiscent of a normal shape. Even for the scenario with 1000 data, the shape shows anomalies here (a strong asymmetry on the left) with respect to a normal distribution.

The scenarios reveal the “range of variation” of a probability model. Even if the main distribution is normally distributed (which is guaranteed by our statistical laboratory), the data are far from being distributed normally. The repetition of the simulation scenario allows students to witness the variation as in a video. As a result, we see the relevance of the sample size and that with small samples the conclusions are unstable. We recognise that assumptions such as a normal distribution cannot be statistically tested in an adequate manner. Variations that do not lead to a rejection of the presumption of normality are too wide and may be caused by a different parental distribution.

However, it is possible to draw conclusions on several population parameters even from smaller samples. As a consequence of the Central Limit Theorem, the distribution of various statistics extracted from a sample is approximately normal. This turns out to be an artificial concept, as we usually only have one sample from the population and therefore never experience the sampling distribution of a statistic such as the average of a sample.

3 Tasks that expose the purpose of probability

This chapter discusses tasks that reveal the purpose of probability without explicit connection to statistical inference. Probability can be used to make decisions under risk transparent and to clarify the criteria used to define what a good decision requires and how to provide a solution.

DOI: 10.20396/zet.v28i0.8657607

A basic task refers to a primary purpose of probability, which is to fix the price of uncertainty or risk: Whether we play a game of chance or take out an insurance policy, structurally the situation is the same: two stakeholders meet and exchange their roles: one partner gets rid of an uncertain situation and the other takes over this role from the first, i. e., is willing to leave the position of certainty. By introducing an index called probability, which "measures" the degree of risk, it is possible to find a price for the exchange of roles (first section below).

If one is willing to refer to the probability of a future outcome, then one may also be able to improve a decision that has to be made now. Optimisation criteria without such a probability index could lead to inferior conclusions and have a conservative impact on behaviour: probability increases the scope of strategies with a high potential for innovation but carry an inherent risk – this is a general invariant of using probability even if it can be used for an overall advantage (second section below). When using probability as a price for uncertainty, there are some basic rules to follow, which seem very natural in the context of risk (third section below).

The following sources contain more information about the key properties of probability: Borovcnik (2015a), Batanero and Borovcnik (2016). They cover the aspect of dynamic applets that can show – beyond and parallel to mathematical considerations – the key properties of the concept of probability.

3.1 Pricing the Unknown

Probability serves to exchange the uncertainty and risk involved with money. This is part of the insurance contract where the customer faces the possibility (risk) of an accident and pays the insurance premium to the insurance company. The client leaves the position of uncertainty (about the financial implications of an accident) to enter a position of certainty (without risk) but pays in advance for it. The insurance company – on the other hand – leaves the position of certainty and assumes the risk of the customer and receives a payment for it. The odds are the key to determining the contract price. The example also serves to discuss the various interpretations of probability (subjectivist for the client, frequentist for the company) and utility of impact (utility for the client, simple money considerations for the company). For more details, see Batanero and Borovcnik (2016).

Probabilities are the key to determining the price of the contract, or the basis for decision making if the outcome depends on probabilities, or rather, if outcomes are modelled by probabilities. The stakeholders who come together in a decision (as in the insurance contract) do not need to use the same model of the situation and if they use the same model, they might choose different probabilities; moreover, if they have the same values for the probability, they might add a different connotation to it: for the insurance company it is a frequentist probability averaged over many contracts in the past, for the policyholder it is a personal and qualitative probability over the event in question and the policyholder does not have an averaging point of view, since he takes the premium only once. The "risk" lies with

DOI: 10.20396/zet.v28i0.8657607

the customer with the full range of variability, while the insurance company bears almost no risk, as its result is very stable and therefore predictable due to averaging.

3.2 Optimising a decision in a situation of uncertainty

We present an example of making decisions about the number of copies to produce for a journal if the demand is modelled by probabilities and there is some additional data on the cost of copies (see Table 1, reproduced from Borovcnik, 2015b); the price per edition is 1.60 (€). The options are to produce 1000, 2000, ..., or 5000 copies. Which is the best option?

Table 1 – Production options, cost of these options, and probabilities of demand for the magazine

Option	1	2	3	4	5
Copies to produce	1000	2000	3000	4000	5000
Cost of production	2000	2200	2400	2600	2800
Demand	1000	2000	3000	4000	5000
Probabilities	0.40	0.30	0.20	0.06	0.04

Source: Prepared by the author

There are different criteria for optimising the number of copies. The maximum loss could be minimised, for example, or the expected benefits could be optimised. It is easy to see (see Borovcnik, 2015b, p. 128) that the minimax method can lead to strange decisions:

> It is obvious that no one would be willing to decide for option 1,000 as – whatever the demand will be – it will lead to a loss of –400. This reminds us to a principle of avoiding a sure loss [...]. The option 2,000 delivers a positive expected profit of 360; however, it can lead also to an even higher loss of –600 as compared with the decision for 1,000 copies. This reflects a basic property of [...] decisions under risk. Rarely can one find decisions, which are better throughout [...]. [...] the remaining [...] actions cannot be compared to each other without a further criterion and what is the better decision depends on the criterion used. To improve a situation in one respect (to have a higher expected net profit) is accompanied by the risk of higher potential losses. One may even speak of an invariant in human life as seen from a general philosophical perspective on risk. The option 3,000, which yields an expected profit of 640, is better. However, it bears the risk of a loss of –800 (if demand is only 1,000, which has a probability of 0.40). It turns out that option 3,000 yields the maximum expected profit (640) and is – in the present model – the best decision.

The concepts involved range from the frequentist to the subjectivist one, including probability and utility, and encompass a discussion of the criteria to be followed when optimising a decision. It should be noted that the task of optimising a decision is not well defined as long as the decision criterion is not set and each criterion has its own merits and drawbacks.

3.3 The expected value is an additive notion – Dispersion is not always additive

There are several ways to derive the expected value of a binomial distribution, which is the result of a binary experiment repeated *n* times independently with a probability of *p* for the event in question. It is very easy to recognise that for one judgment the expected value is *p*. If the situation is embedded in a bet – win 0 if the event does not occur and win 1 if the

DOI: 10.20396/zet.v28i0.8657607

event does occur – the fair price or expected value for this bet amounts to p. If one makes such a bet n times, then the price is n times p, which is very easy to accept without any further mathematics. However, if the bets are dependent, then this additivity – although highly counter-intuitive – is still valid, as can be seen in Batanero and Borovcnik (2016).

A comparable additivity property for the variance is valid only for independent random experiments, as is the assumption for individual trials leading to the binomial distribution. This additivity is somehow compared to Pythagoras' theorem and independence becomes an orthogonal relationship for the random variables. This is the reason behind the choice and preference for the square of the standard deviation in favour of other measures of variability for a random variable and not any consideration concerning the high quality of the visual description of the width of a distribution.

Indeed, for skewed probability distributions, neither the expected value nor the dispersion (or standard deviation) provides an adequate description of the location or the width of a probability distribution. However, these parameters are used throughout. The reason for this lies in the mathematical interrelations, mainly in the fact that the mean of the samples as a statistic is – due to the Central Limit Theorem – approximately normally distributed. This simplifies statistical inference procedures when conclusions are to be drawn from a random sample of the population.

4 Measuring or estimating an unknown quantity

In this section, we try to close the gap between probability and statistical inference. Already in the early stages of teaching probability, one should introduce informal connections for inference so that the purpose of probability is clearer. To do this, we develop an analogy between the task of estimating a probability (which is normally unknown) with the repeated measurement of a physical quantity. We have to define a measurement procedure and clarify how we can conceptualise the quality of our measuring instrument. A better instrument would not always provide a better measurement value, but in general it would be an advantage if used several times (repeatedly). This means that we have to introduce methods to conceptualise the margin of error of the measurement (first section below).

Next, we redefine the repeated coin-tossing experiment (or any other binary experiment) in the context of measurement. The relative frequency of Heads in a series of coin tosses is considered to be a measure of the unknown probability (second section below). Instead of increasing the number of tosses, we investigate three "instruments" for their measurement quality, and set their number of tosses to 5, 10, and 20 only. We just want to demonstrate that the measuring instrument with the larger series of tosses can be considered as more accurate (third section below). And from that we draw our conclusion by "extrapolation" and not by a material experiment. This thought experiment can be applied to any other task of estimating a parameter, such as the average of a population.

DOI: 10.20396/zet.v28i0.8657607

Bootstrap is a modern method for estimating an unknown parameter of a population, which can be used alternatively to a confidence interval. Bootstrap is one of the methods used in “Informal Inference”, a modern approach suggested in research on statistics education, which aims to simplify statistical inference (some would even say to replace statistical inference; see Borovcnik, 2019a). The measurement context also highlights what is actually done in a Bootstrap estimate (fourth section below).

4.1 Analogy between the measurement of physical and virtual quantities

Probability as a limit of relative frequencies is a naive perception that has also been (and still is) the focus of the definitions of probability. This forms the basis of the frequentist view of probability. It is an important task of the introductory phase of probability education to clarify this relationship between probability and relative frequencies. The usual experiment is to repeat a binary randomised trial very often and show the limiting behaviour of the development of the relative frequencies of the event under investigation. Here too, to avoid the empirically unattainable phenomenon of convergence, we prefer the Freudenthal variant (1972) of the experiment, which investigates the distribution (!) of the experimental result to a specific sample size when the scenario is repeated many times (with this sample size).

This approach relates to an analogy between measuring an unknown probability and measuring physical quantities and reveals a key idea of probability – measurement is called estimation in probability jargon.

The measurement of an unknown quantity is a process that is usually performed by a well-defined measurement procedure. However, in the context of probability, we deal with non-physical quantities that require special methods that differ from physical measurements. The average of a population is a virtual quantity that does not really exist, but it can be relevant to know it. The same is true for a proportion of people in a population who have a given property; this is a virtual number. Or, the unknown expected value of a probability model that is used to describe the distribution of a variable (in a finite population or in a data-generating process). Also, the unknown probability of an event is “measured” by the relative frequencies of the underlying random experiment. These all are virtual quantities. However, the analogy with the measurement of physical quantities can clarify the purpose and properties of such “measurements”. In fact, the measurement of a physical quantity is related to measurement errors, measurement bias (a systematic measurement error), and measurement accuracy (for the details of the analogy, see Table 2).

In the theory of physical measurements, it has been recommended to average the result of a few measurements to decrease the measurement error and increase the accuracy of the measurement procedure. Such recommendations have been established by a connection with probability theory. We use the analogy of measuring virtual quantities (such as the mean of a population or the proportion of people in a population with a specific property) with the measurement of physical quantities to improve the understanding of virtual concepts and their probabilistic measurement.

DOI: 10.20396/zet.v28i0.8657607

4.2 The task of measuring – or estimating – an unknown probability

The idea is to evaluate the accuracy of various measuring instruments by comparing them with each other. In Figure 3, we see how the curve of cumulative relative frequencies stabilises with more experiments; after 1000 binary experiments, the relative frequencies seem to be very stable. To make it more specific, we investigate three scenarios to analyse the 1000 data. We could use the coin tossing or any other binary random experiment; we could also imagine focusing on one of the nine random digits of the experiment in Figure 1.

Table 2 – Comparison of terms used in measurement theory and in probability

Category	*Theory of measurement*	*Probability theory*
Method	Measurement of physical quantities	Estimation of virtual quantities
Measurement object	The true value of a *physical* entity	Average or proportion (or other parameters) of a *real or virtual* population
Measuring process	To measure the object physically, use the instrument repeatedly	From the population, simulate data for a random sample
Measurement result	Some values of the measurement	Data of the random sample
Measurement value	The mean value of the measurements	A convenient statistic of the sample data
Naming of the outcome	A measured value of the physical entity	An estimation value of the population characteristic
Error of *one* measurement	The difference between the true value of the entity and the measurement value	The difference between the population parameter and the estimated value
Feature of *one* measurement / error	It varies in general due to uncontrollable conditions	It varies due to the randomness in the generation of the samples
Precision of the measurement	Overall error size	Variability of the estimate due to randomness
Margin of error	It is measured in terms of the standard deviation	It is measured in terms of the standard deviation
Quality of the measurement procedure	The smaller the standard error, the better the method of measurement	The smaller the standard error, the better the method of measurement

Source: Prepared by the author

- Scenario 1: We divide the 1000 data (of 0s and 1s) into blocks of length 5 and use the relative frequency in each block (sample) to “measure” (estimate) the unknown probability. We get 200 (estimated) measurements, which are signified by crosses in Figure 4 (left); the representation shows that – not unexpectedly – the measurements vary greatly and quite a few measurements have an error of greater than 0.20 (i. e., they are outside the band between 0.30 and 0.70, which is marked by thick dashed lines).

DOI: 10.20396/zet.v28i0.8657607

- Scenario 2: We divide the 1000 data into blocks of 10 in length and obtain 100 estimates of the unknown probability, which are represented in the central diagram in Figure 3. Far fewer values are beyond the band marked for measurement errors greater than 0.20.
- Scenario 3: We divide the 1000 data into blocks of 20 in length and obtain 50 estimates of the unknown probability, which are shown in the diagram to the right of Figure 3. Only two measurements have an error greater than 0.20.

The crosses representing the individual measurements of the unknown quantity show a clear pattern of reduction from left to right in Figure 3 with an increasing number of data (length of the measurement series, or sample size). We can also identify the stabilising effect of the length of the experiment on the chart of the relative frequencies (in Figure 3) with the total influence of randomness to which the individual measurements of the resulting blocks are exposed. This emphasizes that there can be no compensatory effect according to which the ensuing results would correct an existing deviation in the relative frequencies in the direction towards the limit (the probability). This compensation thinking is very popular, as can be seen in the ubiquity of the lists of numbers recently drawn in the state lottery, which “provide the basis for calculating the current odds” of the numbers for the next drawing.

4.3 Analysis of the accuracy of the three measuring instruments

The measurement analogy shows that the measuring instruments differ greatly in the quality of the measurements; the errors of Instrument 1 (Scenario 1) tend to be much greater than those of Instrument 3 (Scenario 3). We compare these measuring instruments by means of a statistical analysis of the measurement errors in our scenario. This means that we compare the pattern of the distribution of the unknown probability estimate based on repeated data from 5, 10 and 20 with a bar chart (Figure 4). Again, we see the shrinkage effect. The larger sample gives a more accurate estimate of the unknown probability. Figures 3 and 4 are slightly modified from Borovcnik (2019a).

Figure 4 especially prepares the idea of the sampling distribution of the proportion for a fixed size sample. It is essential to explicitly clarify the artificial character of our simulation study of the measuring instruments: we have repeated the measurements to obtain an empirical database for the quality of the instruments. In practice, however, we have only *one* sample and *one* measurement (estimate) of the unknown parameter. We apply the knowledge of such artificial studies to transfer the knowledge of quality to this single value and thus we can judge its reliability in the sense of the dispersion (measured by the standard deviation) of the sampling distribution of the proportion.

DOI: 10.20396/zet.v28i0.8657607

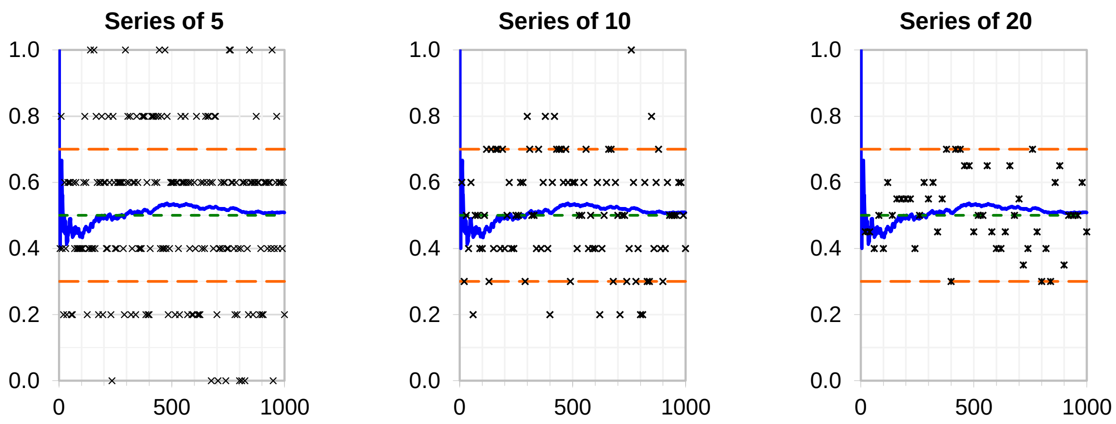

Figure 3 – "Convergence" of relative frequencies towards the unknown probability for the duration of the series of experiments compared with repeated measurements of this probability by different procedures:

Left: Probability measurement by the relative frequency of the last block with 5 trials. Middle: Measurement by the result of the last block with 10 trials. Right: Measured by the last block with 20 trials

The band within the dashed lines marks those measurements with an error of less than 0.20

Source: Prepared by the author

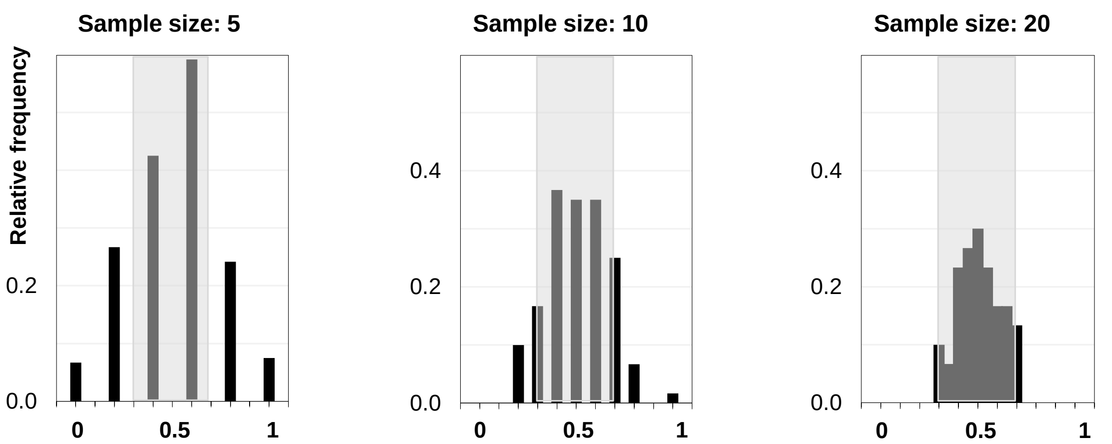

Figure 4 – Comparison of the accuracy of the three procedures for measuring an unknown probability

Left: Based on the last block with 5 data. Middle: Based on 10-data blocks. Right: Based on 20-data blocks

The shaded area signifies a measurement error of less than 0.20

Source: Prepared by the author

We explore the accuracy of the measurement procedures rather than investigate a law of convergence. This basic experiment avoids obscure limiting behaviour of relative frequencies and focuses instead on three snapshots that illustrate the phenomenon: a larger sample size allows estimating the unknown probability much more precisely. The experiment also shows already in the initial phase what the probabilities are for and how we can link probability and statistics and how we can operationalise the concept of statistical information

DOI: 10.20396/zet.v28i0.8657607

and how we can transfer the sample estimates to the population (finite, or the processes that generate the samples).

Studying in the statistical laboratory ensures that other factors could not blur the results, as we have full control over the assumptions. Repetition of the simulation scenario (in Figures 3 and 4) is very instructive, as it shows a stable pattern in the reduction effect and allows students to experience its variation (when it changes, how it changes, what is invariant, what is the range of variation or error if a quantity is measured) rather than investigating obscure limits that are not open to scrutiny.

The analysis of the 1000 data in blocks of different lengths is motivated by a coin-tossing protocol. In a computer-simulation version we could try to avoid that for size-20 samples there is much less data (only 50) compared to size-5 samples (where we have 200 data). The less data one has, the more additional variation the simulation can have, which might blur the underlying pattern. In a dynamic simulation, 1000 (or more) size-5 samples could be simulated, then size-10, and finally size-20 samples so that the empirical distributions are less prone to random-effect artefacts from the simulation and the pattern is even more clearly visible.

The experiment for averages is analogous to the current one. Only the mean rather than a proportion of a population should be measured by a sample. Once again, the measurement procedure is investigated. How accurate is the measurement of the average of a population by the average of a sample? In the statistical laboratory, we can simulate samples of the parent distribution (the population) to investigate the used measurement device. It is essential to remember that in practice one only has one measurement, as one only has one sample. The scenario of simulating repeated samples of the population, determining the mean of this sample, and continuing the sampling with the calculation of the mean provides an empirical basis for the sampling distribution of the mean. This sampling distribution is conceptually not much easier in the simulation scenario although it is materialised in the data on the repeated measurements that are generated in this sampling process; it describes the quality of the measuring instrument.

The sampling distribution is usually obtained by mathematical theorems; it is the key to statistical inference. We use the introductory experiment to relative frequencies to familiarise students with the connection between probability and statistical inference. It is also of interest to investigate through a simulation study how fast the convergence is, i. e., how much sample data is needed to make the distribution look quite similar to a normal distribution. Again, the Central Limit Theorem provides an easy description of the sampling distribution of the mean as it is approximately normally distributed.

4.4 Extension of the repeated-measurement analogy to the Bootstrap method

It is very interesting that the analogy of repeated measurements gives a natural explanation for the Bootstrap method. If we knew the population, we could simulate one sample after another, always calculating the parameter of interest; this sample parameter

DOI: 10.20396/zet.v28i0.8657607

provides a measure of the amount of population we are interested in. In Bootstrap, the simulation is performed from an approximation of the population, which is provided by the original sample. Normally, a confidence interval would be calculated on the basis of that sample. This confidence interval provides information about where the “true” parameter of the population should be.

Apart from the known problems in the interpretation of confidence intervals, the method provides rather complicated formulas for the confidence interval for parameters other than the population mean. This mathematical complexity might also be the reason for the frequent misunderstanding not only of statistical tests but also of confidence intervals. Bootstrap intervals are much easier to obtain, as only the data is needed and no further assumptions are made. The interval is obtained simply by the repeated simulation of samples from the first sample (sampling with replacement). However, the reason why the method works can be clarified much better than usually by the analogy with measurements.

Given: the data of a sample size *n* with mean and SD for a specific variable. How accurate is the sample average as a measure for the whole population? If we have access to the population, we could repeatedly take a sample from it and calculate the average value of this sample. This mean value represents the first measurement of the “unknown” mean. In the statistical laboratory (where the population including its mean is known), we can investigate the accuracy of our measurement procedure for the population mean by this method of measuring the population mean by the sample mean.

If we generate 1000 samples (or even more) and calculate the mean, we obtain an empirical basis of measurements, which we can analyse statistically as in the case of the task of measuring an unknown probability (proportion of a population). Using this simulation scenario, we provide an empirical basis for the sampling distribution of the mean and can judge the quality of the measurement procedure (the estimation of the “unknown” mean). The procedure is exactly the same as in the example of measuring an unknown probability.

If we do not have access to the population, we cannot sample from it more repeatedly and measure (estimate) the unknown average of the population. Yet, the following measurement procedure, called Bootstrap, may provide a reasonable way to measure (estimate) the unknown population mean. Instead of taking a sample from the population (which is no longer possible), we take a new sample from the already existing data (with repetition). Since this initial data set is a random sample of the population, there should not be too much error in this ‘rough’ sampling procedure. As in the case of the known population, we calculate the mean of this Bootstrap sample and “measure” (estimate) the unknown mean of the population.

The quality of this measurement procedure is evaluated by repeating the measurement, i. e., repeating to take a Bootstrap sample of the data and calculate its average. If the process is repeated 1000 times (or more), an empirical basis for the distribution of the mean is obtained (see Figure 6). We can recognise how much these Bootstrap measurements vary from sample to sample and find a central interval (of 95 %, for example) of these

DOI: 10.20396/zet.v28i0.8657607

measurements. This so-called Bootstrap provides an evaluation of the measurement procedure. It has properties similar to those of a confidence interval for the mean but has a different interpretation.

Raw data initial sample		1. Bootstrap: sample of the initial sample	
Nr	**Length**	**Nr drawn**	**Length**
1	12	5	19
2	2	7	34
3	6	10	4
4	2	7	34
5	19	8	4
6	5	2	2
7	34	10	4
8	4	8	4
9	1	7	34
10	4	1	12

Statistics of the first sample		First measurement of the mean of the population	
Mean	**SD**	**n**	**Mean**
8.90	10.39	10	15.10

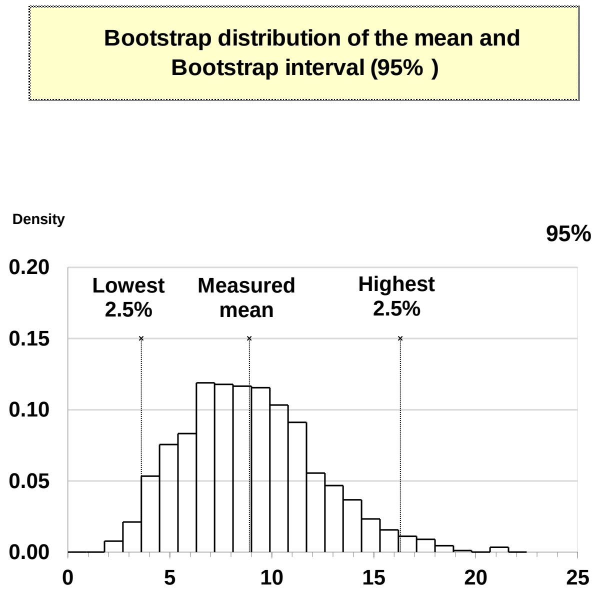

Figure 5 – Bootstrap samples for measuring the unknown mean.
Left: The first Bootstrap shows that the measurement of the unknown average is very unreliable.
Right: The Bootstrap distribution produces 95 % of repeated measurements between 3.81 and 16.12
Source: Prepared by the author

In fact, it is didactically attractive how simple the interpretation of a 95% Bootstrap interval is: it is that central interval, which contains 95% of all Bootstrap measurements of the unknown population parameter. If the measurement method has no systematic errors, then the repeated measurements should vary around the actual value. And the Bootstrap interval indicates the variability of the measurements and is the best that can be obtained with this method, which has an inherent margin of error. The analogy with the measurement of physical quantities fits perfectly.

Since the early 1990s there have been suggestions to use the Bootstrap and non-parametric statistics as a transitional stage for teaching statistical inference in order to simplify the first approach to this sophisticated topic. This Bootstrap interval is a naive but very interesting imitation of the classical confidence interval although Borovcnik (2019a) gives reasons why the Bootstrap interval method is not suitable to replace the classical confidence intervals; it could be used as an intermediate phase in teaching until the more sophisticated confidence intervals can be taught. Cobb (2007) has suggested completely replacing statistical inference with resampling methods, including bootstrapping and rerandomisation. In practice, experts would know when to use which method to estimate the average of a population.

DOI: 10.20396/zet.v28i0.8657607

5 Conclusions and perspectives

In the second chapter, we investigate the probability character of the random-digit experiment. The equiprobability of digits and the development of relative frequencies are two inseparable parts of the concept and establish a relationship of complementarity for probability.

Rather than following obscure convergence considerations, we show by our experiment that the margin of fluctuation of the relative frequencies of the digits is much smaller for 1000 random digits than it is for 50. This is an experimental discovery. The rest, that the fluctuation margin will continue to shrink in a similar mode from 1000 digits to a larger size of simulated digits, is due to a thought experiment. This is typical for our animations that we only have simulations for small sizes and we establish a model by a visual comparison of the experimental diagrams for various scenarios. Then we repeat the simulation to investigate whether the form already found will remain stable. The repetition of the simulation can be followed like watching a video.

In the third chapter, we investigate the complementarity between probability and risk. Concentrating on the risk aspect by suitable tasks makes it easier to understand the role of probability. In the insurance contract there is a price for the change of a risky position (the client at the beginning) and certainty (the insurance company at the beginning), which can only be clearly determined if one wants to estimate, presuppose, or subjectively determine the potential of the risky event to occur. What number is attributed to this probability may not be easy to solve, but a scenario regarding this number makes it clear, which decision is preferable: to take out the insurance or not to take it out.

Interestingly, the example also touches on the complementarity between frequencies and the subjectivist aspects of probability since the insurance company can use the frequencies of the insured event (e. g, the accident of a car) while the client only has qualitative (subjectivist) considerations to find a suitable probability for this event. In the journal task, there are mathematical solutions without using probabilities (to search for the minimum of the possible maximum loss, for example); yet, such strategies can lead to unacceptable solutions as is shown in the example with the copies of the magazine. The virtual role of probability in relation to risk is made much clearer by such tasks.

In the fourth chapter, we introduced informal-inference considerations in the early stages of teaching probability. These considerations make the concept of probability much more accessible and natural. Instead of “showing” a limit of relative frequencies towards an unknown probability, one would analyse the task of estimating the unknown probability by the relative frequency of a series of experiments with a fixed size. We then developed the analogy with the measurement of physical entities and investigated the accuracy of our measuring instrument.

The analysis of measuring instruments provides the knowledge that the instrument with the larger data set is more accurate. The rest is done with a thought experiment: the more

DOI: 10.20396/zet.v28i0.8657607

data, the more accurate the measuring instrument. The analogy with the measurement of a physical quantity also gives an excellent explanation for the Bootstrap method, which has been adopted by the school of "Informal Inference" in statistics education.

Probability is a virtual concept despite all efforts to illustrate its impact by simulating the random situation under investigation. The concept has many aspects that may be described as a kind of complementarity: Equiprobability and relative frequencies, frequencies and subjectivist aspects of probability, probability and risk, and probability and statistical inference. Complementarity means that separation of the aspects involved would leave the concept with a complete loss of its genuine meaning. As a consequence, simplifying the issues with a didactic objective can deform the meaning and bring the education of probability into a dilemma.

Since probability is a theoretical concept, people lack experience in it, which could correct inadequate basic intuitions. The interaction between these primary intuitions and secondary intuitions becomes relevant to establish stable conceptions in students. Borovcnik (2019b) provides a theoretical framework for identifying fundamental ideas of probability and how to develop intuitions that are stable. This framework refers to Fischbein's interaction between primary and secondary intuitions.

Developing secondary intuitions through mathematical tasks is easier if the design of the tasks and the expected learning paths are focused on the purpose of probability. This helps not only to motivate students to continue their work, but can also convince them that the concepts are useful; yet, most importantly, the purpose, once brought out to the open, allows the mathematics to be developed with a goal so that the concept begins to make sense. Above all it makes sense since it helps to solve the task. It makes sense, then, because in the context of the task and in relation to the purpose of the task, the concept can become natural with obvious properties, or at least with properties that seem reasonable (and not counterintuitive). By working repeatedly on such tasks with an explicit purpose, students can complete their cognitive network for that concept and gain a broader understanding. An understanding, as it sometimes occurs with engineers: without knowing all the mathematical details, they have a good understanding that allows them to draw their conclusions.

References

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

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. Online: doi.org/10.1007/978-3-319-31625-3_1.

Batanero, C., Henry, M., & Parzysz, B. (2005). The nature of chance and probability. In: A. G. Jones (Ed.), *Exploring probability in school* (pp. 15-37). New York: Springer.

Bennett, D. J. (1999). *Randomness*. Cambridge, MA: Harvard University Press.

DOI: 10.20396/zet.v28i0.8657607

Bohr, N. (1934/1927): *Atomic theory and the description of nature*. Cambridge, UK: Cambridge University Press (containing the *Como lecture* of 1927).

Borovcnik, M. (1992). Stochastik im Wechselspiel von Intuitionen und Mathematik (Stochastics in the interplay between intuitions and mathematics). Mannheim: Bibliographisches Institut.

Borovcnik, M. (2015a). Central theorems of probability theory and their impact on probabilistic intuitions. In J. M. Contreras, et al. (Eds.), *Didáctica de la Estadística, Probabilidad y Combinatoria, 2* (pp. 15-35). Granada, Universidad de Granada.

Borovcnik, M. (2015b). Risk and decision making: The "logic" of probability. *The Mathematics Enthusiast, 12*(1, 2 & 3), 113-139.

Borovcnik, M. (2016). Probabilistic thinking and probability literacy in the context of risk. *Educação Matemática Pesquisa*, *18*(3), 1491-1516.

Borovcnik, M. (2019a). Inferência informal e inferência "informal". *Educação Matemática Pesquisa, 21*(1), 433-460. Online: dx.doi.org/10.23925/1983-3156.2019v21i1p433-460.

Borovcnik, M. (2019b). Conceptos fundamentales y sus propiedades clave en probabilidad – Cómo identificarlos y proporcionar intuiciones que los sostengan. *Revista Eletrônica de Educação Matemática (REVEMAT), 14, Educação Estatística*, 1-25.

Borovcnik, M., & Bentz, H. J. (1990, 2003). Intuitive Vorstellungen von Wahrscheinlichkeitskonzepten: Fragebögen und Tiefeninterviews (Intuitive conceptions of probabilistic concepts: Questionnaire and in-depth interviews). Technical Reports. Klagenfurt University.

Borovcnik, M., & Bentz, H. J. (1991). Empirical research in understanding probability. In R. Kapadia, & M. Borovcnik (Eds.), *Chance encounters. Mathematics education library vol 12* (pp. 73-106). Dordrecht: Kluwer Academic Publishers.

Borovcnik, M., & Kapadia, R. (2014). A historical and philosophical perspective on probability. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 7-34). New York: Springer.

Borovcnik, M., Peard, R. (1996). *Probability*. In A. Bishop, et al. (Eds.), *International handbook of mathematics education* (pp. 239-288). Dordrecht: Kluwer.

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*. Monterrey: ICMI and IASE.

Chernoff, E. J. (2013). Probabilistic relativism: A multivalentological investigation of normatively incorrect relative likelihood comparisons. *Philosophy of Mathematics Education Journal*, *27*, 1-30.

Chernoff, E. & Sriraman, B. (Eds.) (2014). Probabilistic thinking: presenting plural perspectives. Advances in Mathematics Education (Vol. 7). New York: Springer.

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

DOI: 10.20396/zet.v28i0.8657607

de Finetti, B. (1937/1992). La prévision: ses lois logiques, ses sources subjectives. *Annales Institut Henri Poincaré, 7*, 1-68 (1937); English translation: Foresight: Its logical laws, its subjective sources. In S. Kotz, & N. L. Johnson (Eds.), *Breakthroughs in statistics* (pp. 134-174). New York, Berlin: Springer.

Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Dordrecht: Reidel.

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

Freudenthal, H. (1972). "The empirical law of large numbers" or "The stability of frequencies". *Educational Studies in Mathematics, 4*(4), 484-490.

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

Jones, G. A., Langrall, C.W., & Mooney, E. S. (2007). *Research in probability: Responding to classroom realities*. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning*, Vol. 2 (pp. 909-956). Greenwich, CT: National Council of Teachers of Mathematics, Information Age Publishing.

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

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

Kolmogorov, A. N. (1933/1956). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin; English translation: *Foundations of the theory of probability*. New York: Chelsea.

Lecoutre, M. P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics,* 23(6), 557-568.

Simons, P. R. (1984). Instructing with analogies. *Journal of Educational Psychology, 76*(3), 513–527.

Spiegelhalter, D. (2014a). Comments on probabilistic thinking. In: E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: presenting plural perspectives* (Back cover). New York: Springer.

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

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

Vancsó, Ö. (2009). A parallel discussion of classical and Bayesian ways as an introduction to statistical inference. *International Electronic Journal of Mathematics Education, 4*(3), 291–322.

Vancsó, Ö. (2018, July). How visualisation using software helps understanding classical and Bayesian statistics. Invited paper in session 6C "Teaching Probability in School –

DOI: 10.20396/zet.v28i0.8657607

Understanding and Linking it to Statistics." ICOTS 10, Kyoto, 8-13 July, 2018. Online: www.researchgate.net/profile/Oedoen_Vancso.

von Plato, J. (1994). Creating modern probability: its mathematics, physics and philosophy in historical perspective. Cambridge: Cambridge University Press.