FUNDAMENTAL CONCEPTS AND THEIR KEY PROPERTIES IN PROBABILITY – HOW TO IDENTIFY THEM AND PROVIDE SUSTAINING INTUITIONS

Manfred Borovcnik
Department of Statistics
Alpen-Adria-Universität Klagenfurt, Klagenfurt, Austria

Probability is highlighted by theoretical concepts, which are far from being intuitive. The first step is to identify key concepts; the second is to clarify these concepts not only using mathematical tools by illustrating either their value in context or their specific properties. We focus on facets of probability that allow for linking its various interpretations, or that link probability to statistical inference. 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 follows the stages of emergence of a relation between the investigated concepts.

BACKGROUND

We re-state fundamental problems with probability that arise from the specifity of this concept and the purpose for which it has been "designed". This specifity requires meta strategies that go far beyond the instruction of the mathematical details. In particular, we show suitable tasks and interactive animations, which are designed to overcome learning obstacles.

Paradigms of Science

To identify the character of the concepts used in a theory, it is helpful to have a panoramic look around how these concepts are used and for which purpose. Which are the philosophical paradigms of science that regulated the emergence of these concepts (see Batanero, Henry, & Parzysz 2005)? For probability, it is the situations of the unknown either in the future or in the past. Especially when it comes to take precautions to safeguard against severe impact in the future, probability models are one of several approaches to analyse the situation and provide decisions.

Pure randomness may be defined by the lack of control or the lack of any method of predicting the future outcome with certainty. We may see randomness as an idea embedded in several paradigms that compete with the way how to predict the future and explain the process of development. To make it simple, Borovcnik (2011a) sees three main paradigms:

- Randomness and divination: randomness is the expression of God's will.
- Randomness and causality: randomness is without specific causes or reasons as opposed to causality, which embodies specific conditions in an experiment.
- Random evolution vs creationism: evolution without specific reasons vs created by God.

There are multiple interrelations between these paradigms. "If an all-knowing God exists, the question arises as to whether this may be reconciled with randomness and free will. As randomness means unpredictability it conflicts with omniscience. […] The debate between Darwin's evolutionary theory and creationism is similar to the causality–randomness debate in physics. Is nature created by a supernatural being or has it evolved solely due to random effects?" (p. 74). We have to restrict ourselves here. As the conceptions seem to touch archetypal ways of thinking, they should not be neglected in teaching as otherwise they would continue to overlap with thinking that arises from the learned concepts.

A Laplacean Demon who would know everything and use causal laws to predict the future, would let us use probability as a replacement or substitute for this wider knowledge as we do not know the background. We would need probability only as long as we have not yet detected those causal laws behind the investigated phenomenon.

Divination would make probability superfluous as if God decides it is not to the human beings to investigate these decisions (that view might also explain the relatively late theoretical conceptualisation of probability and related concepts; see Borovcnik & Kapadia 2014). On the other hand, randomness as simple development without any deeper reason would prevent that a person influences – in any perceivable way – the course of an action so that one might interpret the

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org [© 2018 ISI/IASE]

final outcome as God's decision; in fact this has been used in the ancient oracles such as Delphi. Historically, we trace also the habit of outsourcing a crucial decision, such as the election of a person, to randomness in order to "hear" God's decision (and to avoid responsibility for the decision). This aspect of probability coincides with the idea of fairness and equal chances that have played an eminent role in the emergence of the concept.

Archetypal Ways to Approach Probabilistic Situations

For any educational effort in probability (and statistics) it is essential to consider archetypal ways of thinking that have proven successful in other domains but are much less promising in stochastic situations simply because of their nature, which is genuinely characterised by an element of unpredictability. The following strategies have to be carefully taken up in teaching. Of course, here and there the background assumptions of a random situation may be simulated and by repeatedly being confronted with the low success rate of such strategies in artificial random problems, the students may re-consider their approach. However, more importantly, an open classroom discussion about the relative merits and disadvantages of such strategies might be much more helpful to convince the learners that their strategies are inappropriate in stochastic situations. For the following, we summarise ideas in Batanero and Borovcnik (2016) that go back to the research of Kahneman, Slovic, and Tversky (1982) starting from the early 1970's.

Search for causal connections. If one finds a causal connection, either the probabilities are much clearer or even superfluous.

Search for patterns and similarities. This comprises strategies such as representativeness, by which a probability judgement is influenced by a perceived similarity to the generating process in the background. All kinds of evolving patterns of random sequences may be subsumed here.

Ease of access to information, which may summarise the heuristics availability and anchoring. According to the availability strategy, probability is intuitively approximated to the ease of recalling relevant cases from memory. However, there is nothing more unreliable than personal memory. According to the anchoring phenomenon, probability judgements are highly influenced by given or indirectly provoked information. Here, the information could be arbitrarily set into a timely connection with an incident to judge.

A less cognitively oriented commingling of strategies is the following one, which is not least important for decisions in face of the unknown.

Replacing information by imitating other people (the herding effect). A personal judgement (and a decision based thereupon) may be influenced by what others do. With hindsight, information may be re-interpreted to fit and justify the actions chosen, which is called confirmation bias.

The search-for-patterns strategy is grossly misled in case of stochastic problems. If the basic assumption of randomness applies, then any (awkward) pattern may occur without any gain in information. Thus, the search for patterns does not help; at least it does not help as long the character of randomness of the situation is not put under serious doubts. Therefore, early experiments and analyses in probability should focus on *summary statistics* and not on patterns how the results were actually generated. See also the experiments suggested below.

Archetypal ways to approach probability problems do not change by education; they can only be eased out slowly by teaching the formal concepts in a way that the learners have a strong intuitive access to the formal level of the concepts. There are plenty of studies that well-educated people regress to these "primitive" strategies when they are cognitively overstrained (say, at least by time pressure that prevents a full analysis, see, e.g., Gigerenzer 2002).

THE PROBLEM AND THE PRESENT APPROACH

The Role of Intuitions, Fundamental Ideas, and Stochastic Thinking

Fischbein (1975, 1987) views the individual concept acquisition determined by the interplay between primary raw intuitions and theoretical input from learning tasks, rules, and concepts that bear a more formal aspect and induce – besides learning mathematical rules – secondary intuitions, often impregnated by pictures or paradigmatic situations. If the secondary intuitions are strong enough, they overwrite the primary intuitions effectively, if not, if the theoretical concepts remain isolated, learning will not have a sustaining effect and individuals will return to their primary intuitions in case they need a judgement or help.

Arguments for a lack of an operational concept acquisition for the term probability may be seen in Borovcnik (2011a). One is the lack of feedback from using an adequate strategy in probability situations; that means that one may use a nonsensical strategy to find, e.g., the numbers in the state lottery and win, or one may have a suitable model for a random phenomenon and yet the decision proves wrong. Another indication is the ongoing discussion of fundamental ideas of stochastics, stochastic *thinking*, statistical reasoning, etc., which points into the direction of a form of thinking rather than merely applying a mathematical theory (routinely).

So far, no one has arrived at a reasonable list of fundamental ideas of probability, the initial attempt of Heitele (1975) looks more like the contents of a textbook rather than fundamental ideas. Borovcnik (1997) has centred fundamental ideas around the idea of information, the specific type of probabilistic information and the operations on information such as revise (conditional probabilities) or increase the level of information (increase sample size), or transfer information from samples to populations if the information is randomly obtained (statistical inference).

A further indication of the relevance of intuitions versus operations may be seen in the circumstance that there are numerous research studies on statistical literacy and a quite convincing attempt to operationalise the construct of statistical literacy (Gal 2002) but – so far – rarely a paper on probability literacy (with the exception of Borovcnik 2005, 2017, or Batanero and Borovcnik 2016). To describe the construct of probability literacy is an open issue (see Batanero, Chernoff, Engel, Lee, & Sánchez 2016) though there are rudimentary attempts by Gal (2005). Finally, the abundance of paradoxes (Székely 1986) and the theoretical character of probability as advocated by Steinbring (1991) are further arguments that might be brought forward to follow the intuitions approach. Spiegelhalter (2014) highlights the sophisticated character of probability.

Design Principles for Dynamic Applets to Support Concept Acquisition

A more direct approach beyond the mathematical exposition of the theorems is a basic requirement of statistics education not only for students of studies different from mathematics. Also, the focus within mathematics lies heavily on the derivation of the mathematical connections and their logical proof relative to axioms and optimising criteria. For example, the Central Limit Theorem is hardly open to a full proof even to mathematics students. And in the proof, the used concepts – the characteristic function, e.g. – precludes understanding of the most relevant parts. However, it is not only the convergence of the distribution of the standardised statistics under scrutiny to the standard normal distribution. This theorem incorporates also the speed of convergence to the limiting distribution, which is highly influenced by the shape of the distribution of a single random variable (Borovcnik 2015). To clarify such issues enhances the Central Limit Theorem and the resulting importance of the normal distribution (even for non-parametric statistics). Our principles in designing the applets are signified by the following ideas:

- An abstract concept of probability may be illustrated by its traces in relative frequencies in repeated experiments regardless of the connotation of probability.
- The meaning of parameters of probability models may be enhanced by showing the effect of systematic changes of these parameters.
- Central theorems (limiting theorems) are enhanced by simulation; however, rather than “showing” a limiting behaviour, which cannot be represented materially, we focus on a stable pattern, which inspires a thought experiment that finally enhances the theorem.

All the applets share the exploration of a dynamic change of the initial situation. Like in a video, one looks at the different stages of emergence of a relation between the investigated concepts: Identify the key concepts, Clarify the concepts; Link to interpretations; Link to inference; Restate problems with them; Overcome the problem by suitable applets.

TO IDENTIFY KEY CONCEPTS

Variation and the Quality of Probability Information

A basic difficulty is to understand the kind of information that lies 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 with no preference is all-too easily linked to the probability of ½; this equiprobability bias of Lecoutre (1992) seems most prompting if there are two possibilities. Such an issue is best clarified by an open discussion in the form of an empirical interview with varying situations (see Borovcnik & Peard 1996).

How does this expression link to repeated experiments? It is vital to avoid a too strong focus on the pattern in which the series develops as – if the randomness hypothesis applies – the pattern is arbitrary and each pattern is equally likely except one considers pattern as a *class* of outcomes. There is much research on random sequences and the individual judgement of probability of sequences that have to be compared (see, e.g., Chernoff 2013). For probability as a concept, the speculation about patterns is irrelevant as long as one does not consider the randomness hypothesis violated and then it is a matter of a quite complicated statistical test to apply in a replication of the data (not on the same data!).

For a conceptual understanding of a probability statement it is vital to understand that the range of variation shrinks with larger series and that at the same time the single trials are fully subjected to their random character. That means also that there is no rule of compensation, which is often believed after teaching efforts to demonstrate the stochastic convergence of the relative frequencies towards the underlying probability. Rather than illustrating such a law of convergence, which cannot be done in any finite series, it is essential to show the clear pattern of shrinking in the variability of the random sequences. A suitable experiment for this purpose is to select random digits from 0 to 9 repeatedly and determine the relative frequency of each of the digits.

Not with ever-increasing length of the random sequence. Only with two stages, let us fix 50 and 1000 trials. And then repeat the whole simulation scenario often. What can be seen is that the digits ahead vary in an erratic way, which shows that the randomness in the background is fully at work. What also can be seen is that the range of variation is considerably smaller with 1000 random digits than with only 50. A stable pattern of relative frequencies emerges with full-fledged random fluctuation for each trial, which is not a contradiction seen from this angle. It also shows that – assumed that the sample is in fact drawn randomly – the larger sample size provides the smaller variability. As a thought experiment, already here in an initial experiment, one may speculate about a shrinking towards the benchmark of 1/10 as the probabilities for the random digits are the same from the outset. We show one simulation in Fig. 1 but the reader should bear in mind that the static picture gets it life only if the repetition of the whole simulation scenario is played like an animated film to demonstrate the effect just described (see Borovcnik 2011b).

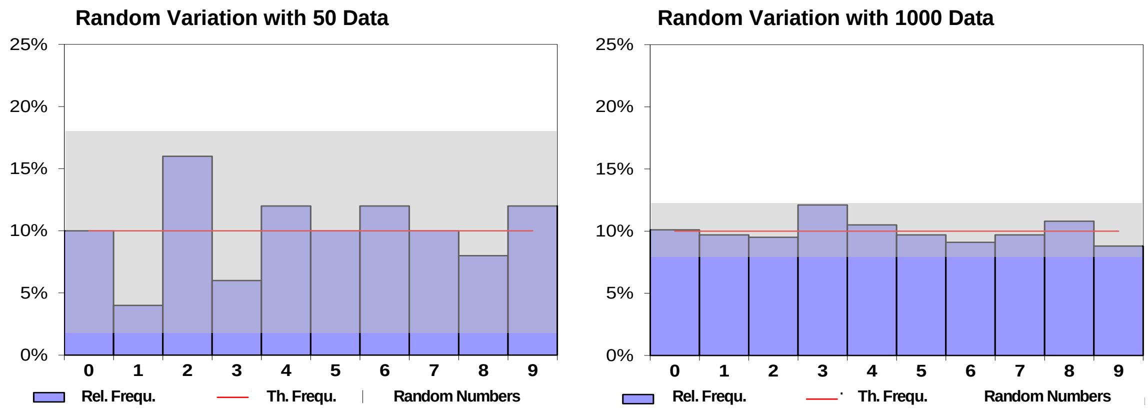

Figure 1. Range of random variation of random digits with a few and with many data

Other Key Properties of Probability

The following sources contain more about key properties of probability: Borovcnik (2011b, 2012, and 2015), Batanero and Borovcnik (2016). They cover the aspect of dynamic applets that may show – beyond and parallel to mathematical considerations – key properties of the concept of probability. A library with applets is available in Borovcnik (n.d.). The applets have been developed and tested in university courses for several years now. The feedback from the students has helped to improve them.

Pricing the unknown. Probability serves to exchange uncertainty and risk involved against money. This is part of the insurance contract where the client 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 get into a position of certainty (no risk) but pays in advance for it. The insurance company – on the other side – leaves the position of certainty and takes the risk over from the client and gets a payment for this. Probabilities are the key for determining the price of the contract. The example serves also to discuss the various interpretations of probability (subjectivist for the client, frequentist for the company) and the utility of the impact (utility for the client, simple money for the company). For details, see Batanero and Borovcnik (2016).

Estimating an unknown proportion (probability). The usual experiment is to repeat a binary random trial very often and show the limiting behaviour of the curve of the development of the relative frequencies of the investigated event. Also here, to avoid the empirically not reachable phenomenon of convergence, we prefer the Freudenthal (1972) variant of the experiment, which compares the distribution (!) of the result of the experiment at one specific sample size when the scenario is repeated quite often (with this sample size). To make it more specific, we perform the binary experiment 5 times and use it to estimate the unknown probability so that we have one estimate. Now, we repeat the whole experiment 1000 times and get 1000 estimates. We then determine the frequency distribution of the estimates of the unknown probability based on a sample of 5 and draw a bar graph of the distribution. We renew the whole setting now generating samples of 20 (and do this again 1000 times). Finally, we compare the pattern of distribution of the estimation of the unknown probability based on 5 and on 20 data. Again, we see the shrinking effect from above.

This basic experiment avoids an obscure limiting behaviour of the relative frequencies and focuses instead on two snapshots illustrating the phenomenon: a larger sample size allows estimating the unknown probability with much more precision. The experiment also shows already in the initial phase what probabilities are for and how we can link probability and statistics and how we will operationalise the concept of statistical information and how we may transfer estimates from samples to populations (finite, or the processes that generate the samples).

Central Laws. It is helpful to re-think the various laws of probability in comparison to each other. One example is to compare the distribution of several statistics related to a simple coin tossing experiment, which is performed 20 times. This sample is repeated often (more than 1000 times) and the following statistics are analysed: Statistic 1: Number of Heads minus Tails, Statistic 2: Average number of Heads minus Tails; Statistic 3: Standardised number of Heads minus Tails. The whole scenario is repeated with the basic tossing experiment performed 40 and finally 100 times. The details of the experiment may be found in Borovcnik (2015); the analysis illustrates the various kinds of limiting behaviour (again via a thought experiment by extrapolating the stable pattern found). Statistic 1 diverges (!), Statistic 2 converges to one (!) point (Law of Large Numbers), Statistic 3 converges in probability to the standard normal curve (Central Limit Law).

There are many more experiments of this kind in Batanero and Borovcnik (2016). They refer to conditional probabilities, to the addititivity of expected values, to the additivity of the variance in case of independent samples, informal comparisons of a probability distribution that represents the population and single sample results (which leads to informal inference) and further aspects of introductory probability with a seamless transition from probability to inferential statistics. For conditional probabilities there is also a multiplicative version of the Bayesian formula based on the so-called odds, which provide a much more suitable representation of probabilistic information for conditional probabilities.

CONCLUSION

It is essential to back up the intuitions by suitable experiments. Rather than investigating limiting experiments, which is not possible, we investigate a stable pattern at a fixed sample size. Then we compare this with a larger sample size (preferably four times as much). To compare the pattern allows an insight about the kind of probabilistic law, which stays behind. To integrate animations, analogies and simulations is more than only motivating; it may enhance the concepts much more and change intuitions sustainingly than a one-side mathematical approach would do.

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. 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.

Borovcnik, M. (1997). Fundamentale Ideen als Organisationsprinzip in der Mathematik-Didaktik. *Schriftenreihe zur Didaktik der Mathematik der Österreichischen Mathematischen Gesellschaft (ÖMG)*, *27*, 17-32.

Borovcnik, M. (2006). Probabilistic and statistical thinking. In M. Bosch (Ed.), *Proceedings of the Fourth Congress of the European Society for Research in Mathematics Education* (pp. 484-506). Barcelona: European Society for Research in Mathematics Education.

Borovcnik, M. (2011a). Strengthening the role of probability within statistics curricula. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics–challenges for teaching and teacher education.* (pp. 71-83). New York: Springer.

Borovcnik, M. (2011b). Key properties and central theorems in probability and statistics – Corroborated by simulations and animations. *Selçuk Journal of Applied Mathematics*, *12*, Special Issue of Statistics, 3-19.

Borovcnik, M. (2012). Multiple perspectives on the concept of conditional probability. *Avances de Investigación en Didactica de la Matemática*, *1*(2), 5-27.

Borovcnik, M. (2015). 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. (n.d.). *Excel files for teaching sophisticated concepts. Library of virtual stochastic experiments*. Online: wwwg.uni-klu.ac.at/stochastik.schule/Boro/index_inhalt.

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.

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

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.

Gal, I. (2002). Adults’ statistical literacy: Meanings, components, responsibilities (with discussion). *International Statistical Review*, *70*(1), 1-51.

Gal, I. (2005). Towards “probability literacy” for all citizens: Building blocks and instructional dilemmas. In G.A. Jones (Ed.), Exploring probability in school (pp. 39-63). Dordrecht: Kluwer.

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

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

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

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

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

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.