WHY READ "PROBABILITY, STATISTICS AND TRUTH" BY RICHARD VON MISES NOW

Yap Von Bing
Department of Statistics and Data Science, National University of Singapore
vonbing@nus.edu.sg

Richard von Mises wrote Probability, Statistics and Truth on his axiomatization of probability, which is based on the collective. Although his probability theory was supplanted by Kolgomorov, this book is still relevant to the teaching of probability, statistics, and science today. This work highlights some outstanding elements of the book and describes their implications on teaching practice.

INTRODUCTION

In 1919, Richard von Mises axiomatised probability using the notion of a collective, which "denotes a sequence of uniform events or processes which differ by certain observable attributes, say, colours, numbers, or anything else" (von Mises, 1981, p. 12). The quote is from Lecture 1, "The Collective" section, from his book *Probability, Statistics and Truth*. The book consists of six superb lectures on the foundations and applications of his system, published in three editions from 1928 to 1951. The probability of an event is its limiting frequency of its occurrence as more and more elements of the collective are observed. Thus, a collective corresponds to the current idea of a "random experiment": a sequence of independent trials performed under identical conditions. This frequency definition (or interpretation) is the best approach for introducing probability (Freedman, 1997; Moore, 1997), and it is strongly supported by the increasing presence of computer simulation in the curriculum (Batanero et al., 2016).

In 1933, Kolmogorov published his measure-theoretic axioms for probability. It suffices here to present the simplest case, where the sample space S is a finite set. The probability of a subset $A \subseteq S$, P(A), satisfies the following properties:

- [K1] $P(A) \geq 0$, for any A.
- [K2] $P(S) = 1$.
- [K3] If A and B are disjoint subsets, then $P(A \cup B) = P(A) + P(B)$.

By the 1940's, the mathematical community had decided to accept Kolmogorov's axioms; for the historical background, see Van Lambalgen (1996). Consequently, *Probability, Statistics and Truth* is not a well-known book. However, in my view, it offers valuable lessons for the teaching of probability to the beginner today, perhaps even more so given the current development in data science and artificial intelligence. The aim of this article is to highlight some important lessons from the book and to describe their relevance to the teaching of introductory probability.

HIGHLIGHTS FROM THE BOOK

In the ensuing subsections, references to *Probability, Statistics and Truth* indicate the lecture and section.

Personal Probability

In the realms of insurance or medicine, the relative frequency of death in a population is commonly used to estimate a probability of death. In Lecture I, The Probability of Death section, von Mises shows that this probability is a property of the population and not that of any particular individual within. A forty-year-old man ("Mr. X") belongs to the group G, "all men insured before reaching the age of forty after complete medical examination and with the normal premium" (p. 17). In G, a fraction, 0.011, died in their forty-first year, so that 0.011 is an estimate of the probability of death of G. But the probability does not refer to Mr. X, for he can be included in other populations, either larger (inclusive of similar women) or smaller (H: men in G who were married, assuming Mr. X was married), which may give very different estimates from 0.011. Von Mises is blunt: "It is utter nonsense to say, for instance, that Mr. X, now aged forty, has the probability 0.011 of dying in the course of the next year" (pp. 17–18). Although the relative frequency of H seems more insightful than 0.011 from G, the former may not be closer than 0.011 to the probability of death of Mr. X, largely because it is hard to define. It is even more difficult to conceive the probability of a unique event,

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

which cannot be included in any sensible collective. Von Mises' maxim: first the collective, then the probability (*Erst das Kollektiv, dann die Wahrscheinlichkeit*), is still forcefully relevant today.

Assigning Probability

In Lecture III, the How to Recognize Equally Likely Cases and The Subjective Conception of Probability sections provide much insight into the topic of assigning a priori probabilities equally to outcomes, based on symmetry or ignorance. On symmetry, von Mises concludes that "… no concrete case can be handled merely by means of an a priori knowledge of equally likely cases" (p. 73). On the principle of indifference, he says:

> I quite agree that most people, asked about the position of the centre of gravity of an unknown cube, will answer 'It probably lies at the centre'. This answer is due, not to their lack of knowledge concerning this particular cube, but to their actual knowledge of a great number of other cubes, which were all more or less 'true'. (p. 76)

Despite the authority implied by the word 'assign,' any assignment of probability is tentative, pending further verification: "… we do not actually know this [for example, each of the six sides of a die is equally likely to appear] unless the dice … have been the subject of sufficiently long series of experiments to demonstrate this fact" (p. 71). In speaking of a fair coin or a fair die, there is no commitment to believe that an actual coin or die is indeed fair. Von Mises indicates in Lecture III in the Are Equally Likely Cases of Exceptional Significance section, "The form of a distribution in a collective can be deduced only from a sufficiently long series of repeated observations, and this holds true for uniform as well as for all other distributions" (p. 74).

On Other Definitions of Probability

The Arithmetical Explanation section in Lecture IV presents a serious criticism of the classical interpretation of the Law of Large Numbers, although it may seem strangely pedantic to us who are so used to making a subconscious or unconscious leap from the classical theory to observations:

> The proposition does not lead to any conclusions concerning empirical sequences of observations as long as we adopt a definition of probability which is concerned only with the relative number of favourable and unfavourable cases, and states nothing about the relation between probability and relative frequency. (p. 108)

The same criticism applies to theorems proved in Kolmogorov's system, with which Kolmogorov seemed to agree, writing in 1933 that: "In laying out the assumption needed to make probability theory applicable to the world of real events, the author has followed in large measure the model provided by Mr. von Mises" (as cited in Shafer & Vovk, 2006, p. 91). Among the objections against von Mises' axioms of probability theory as a science, some are surely concerned with the difficulty of rigorously capturing the empirical character of the subject. In favouring Kolmogorov's axioms, the mathematicians leave out from their discourse such thorny issues, which nevertheless need to be confronted in a course on applying probability theory to real problems. The task falls on the practitioners, who will benefit very much from the knowledge of a sympathetic statistician or probabilist.

On Science

The Theory of Probability is a Science Similar to Others section in Lecture II explains the deep and clear-headed commitment of von Mises to formulate probability theory as a science:

> Like all the other natural sciences, the theory of probability starts from observations, orders them, classifies them, derives from them certain basic concepts and laws and, finally, by means of the usual and universally applicable logic, draws conclusions which can be tested by comparison with experimental results. In other words, in our view the theory of probability is a normal science, distinguished by a special subject and not by a special method of reasoning. (p. 31)

Moreover, the manner in which von Mises motivates his probability theory by analogy with well-known scientific theories is remarkable and inspirational. An aspiring scientist or mathematician will do well to study wise words such as these in the preface to the third German edition:

> As in other branches of science, such as geometry, mechanics, and parts of theoretical physics, so in the theory of probability, the epistemological position remained in the dark for a long

> time. Near the end of the nineteenth century, Ernst Mach and Henri Poincaré made decisive contributions towards the clarification of the meaning and purpose of scientific concepts. (p. *v*)

and

> The essentially new idea which appeared about 1919 … was to consider the theory of probability as a science of the same order as geometry or theoretical mechanics. In other words, to maintain that just as the subject matter of geometry is the study of space phenomena, so probability theory deals with mass phenomena and repetitive events. (p. *vii*)

The entire Sixth Lecture is devoted to statistical physics, including thermodynamics and quantum physics, showing that the frequency definition of probability applies satisfactorily to all such phenomena. If mathematics seems to drift further away from the sciences, *Probability, Statistics and Truth* illuminates a way to foster better mutual understanding.

SOME IMPLICATIONS FOR TEACHING PROBABILITY

Personal Probability

It is not hard to find questions, say in a public health course, asking the student to calculate the probability of an illness for a specific individual. We should avoid this trap, by emphasising that personal advice such as "You have a 60% probability of developing diabetes in 5 years" is not specific to the recipient but refers to a group of individuals, namely the group that gave the relative frequency of 60%. It is best that the characteristics of the group be clearly communicated. If the probability of a unique event must be taught, let it be done after building the intuition of frequency-based probability.

Assigning Probability

Many introductory probability or statistics textbooks cover the idea of assigning probability. There seems to be a habit of teaching three approaches to the beginners: equal likelihood, frequency, and the subjective way. It is better to use the frequency definition right from the beginning, to call a relative frequency an estimate of the probability, not an assignment, and to introduce the equally likely assignment as the definition for fairness. If it is necessary to teach assignment, say for other definitions than the frequency one, the instructor ought to discuss the ensuing conceptual difficulties. For instance, the word "assign" implies a certain power over the state of nature. We do not assign the length, diameter or weight of a given pipe, or the temperature of a mug of coffee; we just measure them as well as we can. I think the same holds for probability in most important applications.

On Other Definitions of Probability

Many introductory probability or statistics textbooks include the "axiomatic" definition of probability, namely [K1–3] in our Introduction. This is analogous to teaching Euclidean geometry using table, chair, and beer mug instead of point, line, and plane, Hilbert's monumental achievement in rigorising geometry notwithstanding. It seems more fruitful to develop probabilistic intuition by treating the subject as a science rather than mathematics. To say the frequency definition satisfies [K1–3] is an understatement, for these statements can be proved from the definition and the intuitive notion of independent experiments. It is precisely in this sense that the frequency definition is an interpretation of Kolmogorov's axioms: it gives rise to a logical structure that satisfies [K1–3] and also

- [K4] $P(A \mid B) = P(AB) / P(B)$

which is another axiom of Kolmogorov. The classical definition also satisfies [K1–4], but the situation is rather different for other definitions. For example, take the subjectivist definition, where $P(A)$ is a personal belief in A. Putting aside the extremely difficult task of measuring belief, it is unclear that [K1–3] should hold (Freedman 1995), or that the belief in A should be updated upon observing B according to [K4]. In fact, intuitive judgment of probability often violates the axioms (Tversky & Kahneman, 1983). Instead, [K1–4] must be imposed on the subjectivist definition: it is an artificial interpretation of Kolmogorov's axioms (Batanero et al., 2016, p. 7). The fact that the frequency interpretation stands strongly apart from the rest presents a serious challenge to the common perception that it is merely one of several more-or-less equally sound interpretations of Kolgomorov. In any case, as argued brilliantly by von Mises, the frequency definition seems to be the correct approach to fulfil the purpose of probability theory as a science, and in today's world, as an indispensable tool for data analysis and making decisions or predictions, whether in statistics or data science.

On Science

Although this subject might appear somewhat remote from the usual narrative, educators of probability and statistics will benefit from a better understanding of the worldview of some leading mathematical scientists of a century ago. If mathematics seems to drift away from the sciences since then, *Probability, Statistics and Truth* illuminates a way for us to foster a more unified outlook among our students.

CONCLUSION

A tremendous advantage of Kolmogorov's axioms over von Mises' is that they build on the theory of real analysis, which has powerful techniques for proving theorems. In teaching introductory probability, however, the frequency definition of probability seems the best one to motivate the classical inference procedures. Hence it should be introduced as early as possible, after a brief discussion of the intuitive view of probability. For random experiments with finitely many outcomes, which are sufficient for almost all applications, [K1–4] can be proved with elementary arguments, for instance via the box model with numbered tickets (Freedman et al., 2007); for a more rigorous approach, see Kerrich (1946). This article highlights excerpts from *Probability, Statistics and Truth* that carry high pedagogical value on personal probability, assignment of probability, comparison of several definitions, and philosophy of science. The implications for teaching have more to do with what to teach rather than how. In particular, there are no separate descriptions for different types of students, although the highly accessible textbook (Freedman et al., 2007) offers a pedagogy that seems quite adaptable to a wide range of student populations.

If von Mises lost the battle of the mathematical axiomatization of probability (Van Lambalgen, 1996; von Mises, 2014), he won it in the scientific sense. His lasting legacy might be less about the mysterious objects known as collectives than establishing the frequency definition of probability as the most appropriate for scientific use, and also for learning probability theory.

REFERENCES

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

Freedman, D. A. (1995). Some issues in the foundations of statistics. *Foundations of Science*, *1*, 19–39. https://doi.org/10.1007/BF00208723

Freedman, D. A. (1997). Discussion. *The American Statistician*, *51*(3), 266–268. https://doi.org/10.1080/00031305.1997.10473976

Freedman, D. A., Pisani, R., & Purves, R. (2007). *Statistics*. Norton.

Kerrich, J. E. (1946). *An experimental introduction to the theory of probability*. Einar Munksgaard.

Kolmogorov, A. N. (1933). *Grundbegriffe der Warcheinlichkeitsrechnung* [Basic concepts of probability theory]. Springer. https://doi.org/10.1007/978-3-642-49888-6

Moore, D. S. (1997). Bayes for beginners? Some reasons to hesitate. *The American Statistician*, *51*(3), 254–261.

Shafer, G., & Vovk, V. (2006). The sources of Kolmogorov's Grundbegriffe. *Statistical Science, 21*, 70–98.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293–315. https://doi.org/10.1037/0033-295X.90.4.293

Van Lambalgen, M. (1996). Randomness and foundations of probability: Von Mises' axiomatisation of random sequences. *Statistics, Probability and Game Theory*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series*, *30*, 347–366. https://doi.org/10.1214/LNMS%2F1215453582

Von Mises, R. (1981). *Probability, statistics and truth*. Dover Publications.

Von Mises, R. (2014). *Mathematical theory of probability and statistics*. Academic Press. https://doi.org/10.1016/C2013-0-12460-9