CONFUSION ABOUT THE CONCEPT OF PROBABILITY

MATHIEU THIBAULT, VINCENT MARTIN

Almost 20 years ago in this journal, Pratt (1998) argued that a difference between the meanings attributed to the concept of probability in everyday and in mathematical language is an obstacle to the learning of probability. For our part, we have noted that there is confusion around the term 'experimental probability' (*probabilité fréquentielle*) [1]. In Québec, the term 'experimental probability' seems to have been introduced in the school curricula in the 2000s. It is commonly used in the school system. It can be found in several Québec educational resources such as ministerial documents (such as the 'Québec Education Program' and the 'Progression of Learning', at the primary and secondary levels), textbooks and digital resources (*e.g.,* 'Buzzmath' [2]). It is quite regularly found in problem statements such as: when a six-sided dice is rolled 100 times and the event 'getting a 3' occurs 19 times, the 'experimental probability' of obtaining a 3 is 19%.

Over the past decade, the term 'experimental probability' has also been used by Québec researchers in theses and dissertations, scientific journals, proceedings of scientific conferences and professional journals. However, some authors in other countries have questioned the expression 'experimental probability'. In the US, Konold *et al.* (2011) clearly state that there is confusion between the concept of 'probability' and the observation of frequency-based estimates:

> At some point, these frequency-based estimates of probability began being referred to as 'experimental' or 'empirical' probabilities and defined in a way that obscured the fact that observed ratios were estimates of the probabilities, rather than probabilities themselves. (p. 70)

In the European context, Batanero and Borovcnik (2016) criticise the use of the term 'experimental probability', as it is the frequencies that are empirical, whereas the probability is theoretical.

> Even though the relationship between probability and relative frequencies is fundamental for the comprehension of probability and statistical methods, this relationship is not always understood as some students confuse frequency with probability. The expression 'empirical probability' is unfortunate in this regard as a probability is always theoretical: only the frequencies are empirical. (p. 10)

Synonymous expressions are found in both English and French: experimental probability (*probabilité expérimentale*), empirical probability (*probabilité empirique*) and estimated probability (*probabilité estimée*). Of these expressions, the last is the one that seems the least problematic, as it emphasises the imprecise nature of what is observed by the frequentist approach. In order to continue the reflection, we propose to distinguish the notion of true probability and the three probabilistic approaches.

Definitions of probability

A probability is a magnitude used to qualify or quantify the randomness of the occurrence of an event. We will see later that this definition leads to particular interpretations depending on the probabilistic approach adopted. For example, here is how 'probabilité fréquentielle' is defined in a textbook widely used in Québec in Secondary IV (students aged 15-16):

> La probabilité fréquentielle est une estimation faite à partir de résultats observés suite à plusieurs réalisations d'une expérience aléatoire. On doit avoir recours à une expérience aléatoire lorsqu'on ne dispose pas d'un modèle permettant de calculer une probabilité théorique. Lorsque l'expérience aléatoire est effectuée un grand nombre de fois, la probabilité fréquentielle constitue une bonne estimation de la probabilité théorique d'un événement. [3] (Boucher, Coupal, Jacques & Marotte, 2009, p. 183)

This definition states that 'experimental probability' should be used when the 'theoretical probability' of an event cannot be computed. The use of 'experimental probability' then allows the 'theoretical probability' to be estimated by experimenting, as in the example in Figure 1 (from the same textbook) which shows the results when a paper clip is thrown 300 times.

However, are the 'experimental probabilities' presented in this example (75%; 13%; 12%; 0%) really probabilities? Accepting these measures as probabilities would imply that probabilities can vary from one experiment to another, which seems to distort the concept of probability in the sense of a magnitude used to qualify or quantify the randomness of an event. Are these 'experimental probabilities' not relative frequencies, that illustrate an observed trend? It seems to us that the observed frequency of an event makes it possible to infer (with varying degrees of confidence) the probability associated with that event, but that it is not the probability itself.

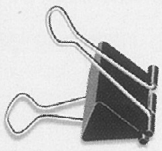
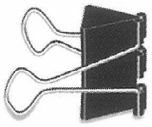
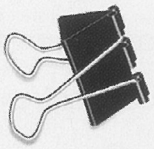
Looking at textbooks, we note that the definitions attributed to probability in cégep [4] and university differ greatly from those used at primary and secondary levels. For

Authorised translation of Thibault, M. & Martin, V. (2018) Confusion autour du concept de probabilité.
For the Learning of Mathematics **38**, 1

Exemple :

Soit l'expérience aléatoire « Lancer un pince-feuilles et noter sa position finale ».

Voici la compilation des résultats de 300 lancers.

Résultat (position finale)				
Nombre de réalisations	225	39	36	0
Probabilité fréquentielle	$\frac{225}{300} = 75\ \%$	$\frac{39}{300} = 13\ \%$	$\frac{36}{300} = 12\ \%$	0 %

Remarque : Même si on n'a pas observé un des résultats en effectuant l'expérience aléatoire, on ne peut pas conclure que ce résultat est impossible.

Figure 1. Example of 'experimental probability' in a textbook (Boucher et al., 2009, p. 183).

example, here is the 'empirical definition of a probability' proposed by a textbook used in several Québec institutions at the college level (students aged 17-20):

> Soit A, un événement associé à une expérience aléatoire. Si l'expérience aléatoire est répétée N fois dans les mêmes conditions et si l'événement A se réalise n_A fois, on définit la probabilité de l'événement A, que l'on note $P(A)$, comme étant la limite de la fréquence relative n_A / N lorsque N devient infiniment grand, ce qu'on écrit :
>
> $$P(A) = \lim_{N \to \infty} \frac{n_A}{N}$$
>
> [5] (Ouellet, 1998, p. 177)

Obviously, there is a gap between this definition of 'experimental probability' and the previous one with respect to the level of formalism. But we can also notice that this last definition presents the probability as a limit. This highlights the evolutionary character of the observed relative frequency, an element that was not put forward in the definition of Boucher *et al.* (2009).

In a university-level probability textbook, Ross (2007) cautions against the notion of 'experimental probability'. Indeed, since a random experiment cannot be repeated infinitely, the frequency obtained is never precise enough to reach the exact probability. The relative frequency can therefore only tend towards a limit value, which is the probability. It is the law of large numbers that tells us that it is likely that this frequency is the limit value of the probability. However, Ross (2007) points out that it is problematic to define the probability as a limit frequency:

> [La probabilité de l'évènement] est définie comme la limite du pourcentage du nombre de fois où [l'évènement] survient par rapport au nombre total de répétitions. C'est donc la fréquence limite de [l'évènement]. Bien que [cette définition] soit intuitivement commode et qu'elle doive toujours rester à l'esprit du lecteur, elle possède un sérieux inconvénient. Nous ne savons en fait pas si [le nombre de fois où l'évènement survient dans n répétitions] va converger vers une limite constante qui sera la même pour chaque séquence de répétitions de l'expérience. Dans le cas d'un jet d'une pièce par exemple, peut-on être sûr que nous obtiendrons de nouveau la même proportion limite de piles si l'expérience est entièrement répétée une deuxième fois ? [6] (p. 33)

In this excerpt, Ross mentions that it cannot be taken for granted that the relative frequency converges to the probability of an event in a random experiment. He questions whether it is possible to know with certainty that there will be convergence to a limit value at infinity, given that infinity is a theoretical concept that cannot be 'verified' experimentally.

Probabilistic approaches and the notion of true probability

This reflection on the expression 'experimental probability' can be situated conceptually by addressing different probabilistic approaches. Generally speaking, the work carried out in mathematics education distinguishes three main probabilistic approaches.

First, the theoretical approach is treated most regularly and is therefore considered as the classical approach (Batanero, 2014). Historically, the origin of this approach goes back to the work of Laplace. At that time, it was the only (and therefore the first) existing approach, so the concept of probability was attached to it. In this approach, probability is computed from the ratio between the number of favourable and possible outcomes of a given event, when all outcomes are considered to be equiprobable. Thus, one cannot (theoretically) compute the event 'getting a 4' with

an unbalanced dice. To approach the probability of such an event, one can measure its relative frequency following a series of observed data. This is the frequentist approach, which von Mises (1928/1952) and Renyi (1966) were among the first to tackle in their work. However,

> la fréquence relative d'un évènement a tendance à devenir constante lorsque le nombre d'expériences ou d'observations augmente. Cette stabilisation de la fréquence relative d'un évènement dans une longue suite d'expériences ou d'observations, répétées dans des conditions uniformes, a été observée depuis fort longtemps. Cette observation amène à formuler le principe suivant: si dans les mêmes conditions, on répète une expérience aléatoire indéfiniment, la fréquence relative d'un évènement associé à cette expérience tend vers une valeur limite. [7] (Ouellet, 1998, p. 177)

The value at the limit is used to approximate the probability of the event under consideration. However, in the same way that 'theoretical probability' has been associated with the theoretical approach, the idea of 'experimental probability' has become associated with the frequentist approach.

Finally, there is another, almost marginal, way of dealing with probability in the scientific literature and in educational resources: the subjective approach. In this approach, which is the topic of Lindley's (1980) seminal work, an individual evaluates quantitatively or qualitatively the strength or degree of a belief through a more or less intuitive analysis of the information at hand. This assessment, which is subjective in that it varies according to the person making it, is usually referred to as the 'subjective probability'. Since the subjective approach does not lead to the probability itself, but rather offers a way to approach this probability, we propose to use the term 'probabilistic opinion' instead of 'subjective probability'.

Thus, we see three probabilistic approaches to approximate a probability. More precisely, 1) a probability is computed in the theoretical approach; 2) a relative frequency is observed in the frequentist approach and 3) a probabilistic opinion is expressed in the subjective approach. Beyond these probabilistic approaches, Konold *et al.* (2011) proposed to consider the notion of 'true probability'. This corresponds to a unique and specific value that represents the probability of an event occurring, but which remains abstract since it cannot be known exactly. This true probability is not equivalent to the probability associated with real-life phenomena. Rather, the notion of probability is derived from a model that emerges from the phenomenon under consideration. This modelling is, at best, a good approximation of the phenomenon. For example, if we are interested in the probability of getting 'tails' with a coin, we will say that it is 1/2, but there is no good reason to believe that a coin has exactly the same probability of falling on heads or tails since it can never be perfectly balanced. Indeed, the theoretical approach assumes that one is working with 'perfect' situations (*e.g.,* perfectly balanced coins with no thickness), assuming that the situation does not change and is not affected by the conditions of the experiment. While the probability of getting tails is the same for all coins, the true probability of this event for a given coin could be 0.4999993697. One way to approach this true probability is to collect data from the real situation and compare the probability with the relative frequency resulting from the frequentist approach. However, the true probability may change over time if the coin is slightly modified as a result of many throws (*e.g.,* with wear). This leads us to say that the true probability is not a fixed value since it depends on a multitude of parameters. It is therefore contextually located and can change over time.

Konold e*t al.*'s reflection on the notion of true probability implies that the relative frequency being stabilised in an experiment can be used to approximate the true probability (and not the probability as a result of the theoretical approach), which cannot be known. This notion of true probability is important especially in the throwing of an irregular object, for which it is not possible to compute the probabilities of obtaining one or other of the sides. For example, Konold *et al.* use a deer bone with irregular sides. They argue that the frequentist approach then allows the relative frequency to tend towards the true probability in the long run. Since this situation does not allow for the calculation of probability, these authors suggest physically performing such a random experiment. In this situation, as it is not possible to approach the true probability by the theoretical approach, it becomes necessary to carry out an approach that is in line with the frequentist approach, by accumulating trials.

Model of the relationship between probabilistic approaches and true probability

In Figure 2, we now propose a model of the relationships between the three probabilistic approaches (boxes), their output (in italics) which is obtained in a particular way for each approach (arrow with verb) and the notion of true probability (in bold).

The true probability is at the centre of the model in a closed set, as it represents the unattainable and abstract value towards which the results of the different approaches, taken individually or in combination, tend in their own way (dotted arrows in Figure 2). In the frequentist approach, it is through the increase in sample size in an experiment that a relative frequency is observed to approach the true probability. In the subjective approach, it is through the accumulation of information relevant to an individual about the event and its context that a probabilistic opinion is expressed to approximate the true probability. In the theoretical approach, the ratio between the number of favourable outcomes and the number of possible outcomes is used to compute the probability to approximate the true probability. With regard to our definition of probability, we can say that these three approaches produce outputs to approach the magnitude used to qualify or quantify the randomness of an event associated with a phenomenon. As far as the true probability is concerned, it represents the exact value (in a given context and at a given time) and abstract value of this magnitude.

In this model, a double distinction is therefore made 1) between each approach and its respective output, and 2) between the three probabilistic approaches that allow us to move towards the true probability of an event. The set dedicated to each of the probabilistic approaches is delimited by

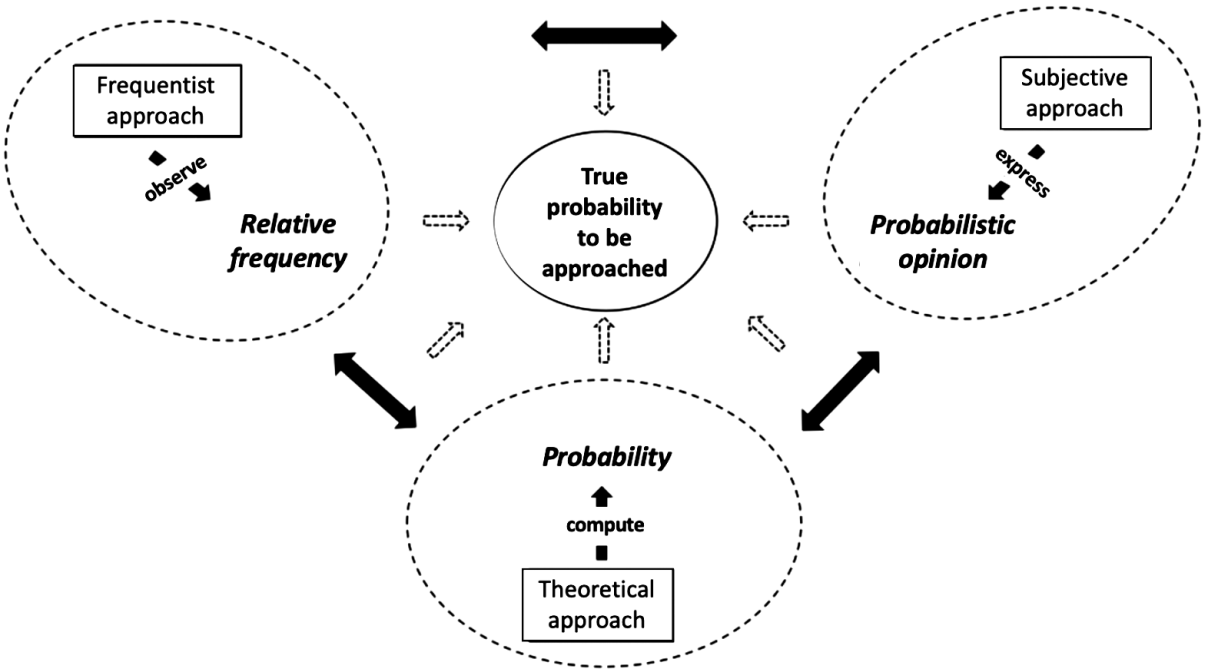


Figure 2. Model of the relationship between probabilistic approaches and true probability.

dotted lines, because these sets are not closed on themselves. Indeed, there is a possibility to go back and forth and combine probabilistic approaches. Figure 2 therefore contains three double arrows representing these possible combinations. These combinations make it possible to adopt a complementary view to approach the true probability, when the context allows it. To illustrate this idea of combinations of approaches, we suggest an example for each of the double arrows in the model proposed. That said, the detailed way to obtain results to approach the true probability by combining different approaches is something that remains to be explored.

The combination of the frequentist and theoretical approaches could be done when performing trials with an *n*-sided dice to confirm the equiprobability associated with obtaining each of the faces (*i.e.,* 1/*n* if the faces are assumed to be equiprobable). Taking into consideration both the stabilisation of relative frequencies and the calculation of probabilities leads to a probabilistic estimate that combines both approaches, thus approaching the true probability. The combination of the frequentist and subjective approaches could be done when the observation of the relative frequency is related to the expression of a probabilistic opinion. Consider, for example, a citizen who is looking for the true probability of seeing a fire truck at a given intersection in the neighbourhood he lives in. This true probability could be approximated by the citizen from his observation of the frequency of passage, combined with his probabilistic opinion based, for example, on an analysis of the proximity of different fire stations and the dynamics of road traffic. Finally, the combination of theoretical and subjective approaches could be done in a situation where the expression of a probabilistic opinion is added to the calculation of a probability. For example, a poker player could approach the true probability of winning a hand in a game. To do so, he would consider both the calculation of the probability of holding the winning hand by considering the cards he knows (i.*e.,* his own and those exposed on the table) and his probabilistic opinion based on a more qualitative analysis of elements drawn from the observation of the other players (*e.g.,* psychological factors such as their reactions and strategies) or from the dynamics of the game, etc. Since the probabilistic approaches can be combined in pairs to approximate the true probability, one could also think that a combination of the three probabilistic approaches is possible.

Thus, the theoretical approach is the only one that can answer the question "What is the probability of...?", but each of the three approaches offers the possibility of answering the question "What could be the true probability of...?" Our model thus illustrates the ways in which the true probability can be approached from one (or more) of the three approaches.

Stabilisation of a relative frequency to approximate the true probability of an event

We will now explore the idea of stabilising a relative frequency associated with an event to approximate its true probability. In order to illustrate the stabilisation of the relative frequency, we will take the example of the paper clips presented in Figure 1. According to the frequentist approach, the experimentation process must allow a random experiment to be repeated under the same conditions a certain number of times. A relative frequency is then obtained which, as it stabilises, makes it possible to approach the true probability with a level of confidence that increases as the number of repetitions increases. From this point of view, we

consider that the law of large numbers (after an infinite number of trials) indicates a stabilisation towards the true probability associated with the event, and not towards its 'theoretical probability' (as is usually said). Although it is not possible to carry out an infinite number of trials with a paper clip, a large number of throws can provide a trend. We model the situation by approximating the true probability for each of the positions. In this sense, with reference to the work of Henry (1999), the action of approximating the true probability could be seen as the development of an increasingly accurate and detailed modelling using the outputs supported by probabilistic approaches. This modelling is a mathematisation that gradually adjusts to the observed reality as the estimate becomes more precise.

When looking for the true probability in a frequentist approach, one is led to make conjectures that become more refined. For example, for each position of the paper clip, the conjecture about the true probability will not be 'confirmed' after a given number of trials, however large this number is. Indeed, the true probability can be approached, but it would only be reached in principle (assuming that it always remains fixed) with an infinite number of trials. It is quite possible that various conjectures emerge and are refined as the number of trials increases. The conjectures would therefore be rather qualitative at the beginning (one can say that one position is more probable than another). Later, as the confidence level increases (along with the number of trials), quantitative guesses would be made (*e.g.,* that one position seems twice as likely as another). The use of a large sample size to infer the true probabilities of a paper clip position therefore involves the stabilisation of a relative frequency, which refers to the law of large numbers.

The stabilisation of the relative frequency as the number of repetitions increases can be observed through the concept of the margin of error, which reflects the variability in a sample. Using the paper clip example, it is possible to compute a margin of error of approximately 5.66% for a sample of 300 trials at a 95% confidence level (or 19 times out of 20), which is a confidence level commonly used in inferential statistical analysis. This means that, 19 times out of 20, the relative frequencies obtained have a margin of error of 5.66% for each of the positions of the paper clip thrown 300 times as in Figure 1. It should therefore be remembered that, 1 time out of 20, one could obtain a relative frequency based on 300 throws of the paper clip that would be more than 5.66% away from the true probability. For the same 95% confidence level, if the object is thrown 1000 times, the margin of error is about 3.1%. The margin of error decreases to 0.98% for 10,000 throws of the paper clip and to 0.31% for 100,000 throws. It can therefore be seen that the relative frequency is stabilising as the number of throws increases, as the margin of error decreases. It is preferable to have the smallest possible margin of error to increase the confidence in inferring the true probability from the stabilising relative frequency. In this context, it is worth considering whether a margin of error of 5.66% with the results of 300 throws of the paper clip is acceptable. The effort required to obtain the results of 1,000, 10,000 or even 100,000 throws of the paper clip must be considered. In order to produce many trials and thus reduce the margin of error, a simulator can usually be used. However, in this case, this technological tool allows us to observe the relative frequency stabilising not around the true probability, but rather towards the probability computed in the theoretical approach according to what has been programmed.

Concluding remarks: clarifying an epistemological posture

At first glance, the model we propose may seem to support the idea that there is a probability that is true outside the individual who is trying to discover it. This might suggest a positivist epistemological posture that implies that there is "une réalité extérieure à la personne, indépendante et antérieure à son observation et à son expérience [et que] la connaissance ne peut être que le reflet strict de cette réalité" [8] (Jonnaert, 2007, p. 6). How can we conceive of the idea of true probability in a posture such as that of radical constructivism, which recognises the existence of a reality external to the individual, but not absolute? Reflecting on this question would be an interesting step in continuing the discussion presented here.

Translator's notes

[1] The two most common translations of the French term 'probabilité frequentielle' into English are 'experimental probability' and 'frequentist probability'. In this translation the term 'experimental probability' is used, which, whilst not carrying through the reference to frequency in the French term does capture the sense of its empirical nature. This is also the translation used in the Québec curricular documents.

[2] https://www.buzzmath.com/en/

[3] The experimental probability is an estimate made from the observed results of several realisations of a random experiment. A random experiment must be used when there is no model to compute a theoretical probability. When the random experiment is performed a large number of times, the experimental probability is a good estimate of the theoretical probability of an event.

[4] A cégep is a two to three year general and vocational college that most Quebec students start after finishing secondary school at age 17. It can lead either to university studies or to work in a trade.

[5] Let A be an event associated with a random experiment. If the random experiment is repeated N times under the same conditions and if the event A occurs n_A times, we define the probability of the event A, which we note $P(A)$, as being the limit of the relative frequency n_A/N when N becomes infinitely large, which is written:

$$P(A) = \lim_{N \to \infty} \frac{n_A}{N}$$

[6] [The probability of the event] is defined as the limit of the percentage of the number of times [the event] occurs out of the total number of repetitions. It is therefore the limiting frequency of [the event]. Although [this definition] is intuitively convenient and should always be kept in mind by the reader, it has a serious drawback. We do not in fact know whether [the number of times the event occurs in n repetitions] will converge to a constant limit that will be the same for each sequence of repetitions of the experiment. In the case of a coin toss, for example, can we be sure that we will again obtain the same limiting proportion of tails if the experiment is repeated a second time in its entirety?

[7] the relative frequency of an event tends to become constant as the number of experiments or observations increases. This stabilisation of the relative frequency of an event in a long sequence of experiments or observations, repeated under uniform conditions, has been observed for a long time. This observation leads us to formulate the following principle: if a random experiment is repeated indefinitely under the same conditions, the relative frequency of an event associated with this experiment tends towards a limit value.

[8] a reality outside the person, independent of and prior to his or her observation and experience [and that] knowledge can only be a strict reflection of this reality.

References

Batanero, C. (2014) Probability teaching and learning. In Lerman, S. (Ed.) *Encyclopedia of Mathematics Education,* pp. 491–496. Springer.

Batanero, C. & Borovcnik, M. (2016) *Statistics and Probability in High School.* Rotterdam: Sense Publishers.

Boucher, C., Coupal, M., Jacques, M. & Marotte, L. (2009) I*ntersection Mathématique. Culture, Société et Technique. 2e cycle du secondaire. 2e année. Manuel de l'élève B.* Chenelière Éducation.

Henry, M. (1999) L'introduction des probabilités au lycée: un processus de modélisation comparable à celui de la géométrie. *Repères-IREM* **36**, 15–34.

Jonnaert, P. (2007) *Le Constructivisme comme Fondement des Réformes Contemporaines des Systémes Educatifs.* Éditions des écoles nouvelles africaines.

Konold, C., Madden, S., Pollatsek, A., Pfannkuch, M., Wild, C., Ziedins, I., Finzer, W., Horton, N.J. & Kazak, S. (2011) Conceptual challenges in coordinating theoretical and data-centered estimates of probability. *Mathematical Thinking and Learning* **13**(1), 68–86.

Lindley, D.V. (1980) *Introduction to Probability and Statistics from a Bayesian Viewpoint.* Cambridge University Press.

Ouellet, G. (1998) *Statistique et Probabilités. Mathématiques au Collégial. Tome III.* Les éditions Le Griffon d'argile.

Pratt, D. (1998) The co-ordination of meanings for randomness. *For the Learning of Mathematics* **18(**3), 2–11.

Renyi, A. (1966) *Calcul des Probabilités.* Dunod.

Ross, S.M. (2007) *Initiation aux Probabilités.* Presses polytechniques et universitaires romandes.

von Mises, R. (1952) *Probabilities, Statistics and Truth.* William Hodge. (Original work published in 1928).