Von Mises' Frequentist Approach to Probability

Milo Schield[1] and Thomas V. V. Burnham[2]
[1] Augsburg College, Minneapolis, MN
[2] Cognitive Systems, San Antonio, TX

Abstract

Richard Von Mises (1883-1953), the originator of the birthday problem, viewed statistics and probability theory as a science that deals with and is based on real objects rather than as a branch of mathematics that simply postulates relationships from which conclusions are derived. In this context, he formulated a strict Frequentist approach to probability. This approach is limited to observations where there are sufficient reasons to project future stability – to believe that the relative frequency of the observed attributes would tend to a fixed limit if the observations were continued indefinitely by sampling from a collective. This approach is reviewed. Some suggestions are made for statistical education.

1. Richard von Mises

Richard von Mises (1883-1953) may not be well-known by statistical educators even though in 1939 he first proposed a problem that is today widely-known as the birthday problem.[1] There are several reasons for this lack of recognition. First, he was educated in engineering and worked in that area for his first 30 years. Second, he published primarily in German. Third, his works in statistics focused more on theoretical foundations. Finally, his inductive approach to deriving the basic operations of probability is very different from the postulate-and-derive approach normally taught in introductory statistics. See Frank (1954) and Studies in Mathematics and Mechanics (1954) for a review of von Mises' works.

This paper is based on his two English-language books on probability: Probability, Statistics and Truth (1936) and Mathematical Theory of Probability and Statistics (1964).

2. von Mises' Empirical-Inductive Approach

The approach used by von Mises is very different because it is empirically inductive: generalizations are formed from observations. For example, the probability operations of union, intersection and conditionality do not appear at the beginning as postulates; they are derived as conclusions which are only applicable in certain circumstances. As a scientist, he always started with, focused on, maintained connections with, and tested his results on the reality being represented. According to von Mises (1964),

> The probability concept used in probability theory has exactly the same structure as have the fundamental concepts in any field in which mathematical analysis is applied to describe and represent reality. Consider for example a concept such as velocity in mechanics. While velocity can be *measured* only as the quotient of a displacement s by a time t, where both s and t are finite, non-vanishing quantities, velocity in mechanics is *defined* as the limit of that ratio as $t \to 0$, or as the differential quotient ds/dt. It makes no sense to ask whether that differential quotient exists "in reality." The assumption of its mathematical existence is one of the fundamentals of the theory of motion; its justification must be found in the fact that it enables us to describe and predict essential features of observable motions.

For von Mises,

> the subject of probability theory is long sequences of experiments or observations repeated very often and under a set of invariable conditions. We observe, for example, the outcome of the repeated tossing of a coin or of a pair of dice; we record the sex of newborn children in a population; we determine the successive coordinates of the points at which bullets strike a target in a series of shots aimed at a bull's-eye; or, to give a more general example, we note the varying outcomes which result from measuring "the same quantity" when "the same measuring procedure" is repeated many times. In every case we are concerned with a sequence of observations; we have determined the possible outcomes and recorded the actual outcome each time.
>
> We call each single observation in such a sequence an *element* of the sequence, and the numerical mark that goes with it is called its *label value,* its *attribute,* or simply, the *result.* We assume that each time there are at least two possible results. The set of all conceivable outcomes of the observations is called the *label set.* Each

[1] www.teacherlink.org/content/math/interactive/probability/lessonplans/birthday/home.html

> outcome may be thought of as a "point" in a space known as the *label space* or *attribute space*; many authors also use the term *sample space.*

The word 'chance' is used by von Mises to describe any process having a 'stable limiting frequency.'

> *First basic assumption.* Consider a finite label space ***S*** containing *k* distinct labels a_1 , a_2, ..., a_k, and a sequence $\{x_j\}$, where each x_j is one of the a_i. Let a_i be a fixed label; among the first *n* elements of the sequence $\{x_j\}$ there will be a number n_i of results carrying the label a_i. The number n_i depends on *n*, on a_i, and on the sequence $\{x_j\}$. In our notation the subscript refers to the subscript of a_i, and *n* to the total number of trials. The ratio n_i /n is the *frequency* or relative frequency *of* a_i among the first *n* elements of $\{x_j\}$. For our purpose a frequency concept is needed that is independent of the number *n.* We therefore introduce our *first basic assumption. The sequence can be indefinitely extended and the frequency* n_i/n *approaches a limit as n approaches infinity.* We write Lim (n_j /n) as $n \to \infty = p_j$, $i= l, 2,...,k$. This limit, for which, necessarily, $0 \leq p_j \leq 1$, is then called the *limiting frequency of* the label a_i or also the *chance* of the label *a* within the sequence under consideration.

For von Mises, the property we call randomness can be explicated – or even defined – by the impossibility of devising a successful system of gambling.

> *General idea.* The notion of "chance" established so far leads to a value that is independent of the number of experiments. But the commonly accepted ideas of a game of chance imply another sort of independence also.
>
> A boy repeatedly tossing a dime supplied by the U. S. mint is quite sure that his chance of throwing "heads" is ½. But he knows more than that. He also knows that if, in tossing the coin under normal circumstances, he disregards the second, fourth, sixth, ..., turns, his chance of "heads" among the outcomes of the remaining turns is still ½. He knows—or else he will soon learn—that in playing with his friends, he cannot improve his chance by selecting the turns in which he participates. His chance of "heads" remains unaltered even if he bets on "heads" only after "tails" has shown up three times in a row, etc. This particular feature of the sequence of experiments appearing here and in similar examples is called *randomness.*

He identifies 'insensitivity to place selection' as the essential feature of randomness (i.e., of random sequences):

> In order to arrive at a more precise formulation of randomness we introduce the concept of *place selection.* This notion applies to a specific manner in which an infinite subsequence is derived from an infinite sequence of elements. We say that this *subsequence has been derived by a place selection if the decision to retain or reject the* nth *element of the original sequence depends on the number n and on the label values* x_1, x_2, … x_{n-1} *of the (n - 1) preceding elements, and not on the label value of the nth element or any following element.*
>
> From experience with sequences representing games of chance we gather that in this type of sequences, the chance of a label is *insensitive to place selections.* For example, if we sit down at the roulette table in Monte Carlo and bet on red only if the ordinal number of the game is, say, the square of a prime number, the chance of winning (that is, the chance of the label red) is the same as in the complete sequence of all games. And if we bet on zero only if numbers different from zero have shown up fifteen times in succession, the chance of the label zero will remain unchanged in this subsequence. Insensitivity to place selection, it is seen, is equivalent to what may be called the *principle of impossibility of a successful gambling system.*
>
> The banker at the roulette acts on this assumption of *randomness* and he is successful. The gambler who thinks he can devise a system to improve his chances meets with disappointment. Insurance premiums are successfully computed from formulas based on the assumption of randomness, and in a great number of statistical investigations (see, for example, the observations of "runs" in birth records), the theory based on this assumption proves to hold. We also add the following consideration. The "probability of casting 3" with a given die under given circumstances may be considered as a "physical constant" *p* of the die just as its density, etc**.** To find experimentally an approximate value of this p we cast the die repeatedly and note the results. Our randomness principle states that the chosen mode of notation of the results does not influence the value of *p.*

His second assumption is that the impossibility of a successful gambling system necessitates sequences that are insensitive to place selection. He reserves the term probability for those sequences that have a limiting frequency and that are insensitive to place selection:

> Thus the idea of our *second basic assumption* leads to the following statement:
>
> *In probability theory we shall deal (not exclusively) with such particular sequences satisfying the first basic assumption for which it may be assumed that the limiting frequency, or chance of any label* a_i *is insensitive to place selections. If this holds true, the chance will also be called probability, or more completely, the probability* $p(a_i)$ *of encountering the label* a_i *in the sequence under consideration.*

He notes two ways to create a chance sequence that is non-random. One way is to construct a sequence from scratch by using a rule. For example to construct a coin outcome sequence where head always follows tail and tail always follows head. The limiting frequency of heads is 50%, but this sequence is certainly not insensitive to place selection. A second way is to transform an existing random sequence by applying a rule. He gives this example: suppose we take a random sequence and derive a new sequence by adding the first and second results, the second and the third, the third and the fourth, and so on. The possible labels in the new sequence are 0 (from 00), 1 (from 10 or 01) and 2 (from 11). "It can be shown that they appear with chances ¼, ½, ¼, respectively; but in a 0 can never be followed by a 2 and a 2 never by a 0." In both cases, the created sequence has a limiting value but lacks randomness.

Having noted two properties of sequences (a limiting frequency and insensitivity to place selection), he continues:

> It is useful to denote a well-defined concept of the theory by a concise term. We shall apply the term *collective* to a long sequence of identical observations or experiments, each experiment leading to a definite numerical result, provided that the sequence of results fulfills the two conditions: existence of limiting frequencies and randomness. (Often the word *population* is used in a similar sense, although sometimes one refers to a population when it is left undecided whether randomness prevails.) The word collective will also be applied to the infinite sequence of numbers (labels) which forms the theoretical counterpart of the long sequence of experiments. As in other branches of science, we use the terms (collective, label, frequency, probability, etc.) on the one hand in their abstract mathematical sense, and on the other hand with respect to a concrete field of application, for example a game of chance, an insurance problem, a problem of genetics, or of theoretical physics.

Making a probability statement about an isolated individual is not meaningful or permissible for von Mises.

> The concept of a collective is thus at the basis of our probability theory. The term probability is meaningful for us only with regard to a clearly defined collective (or population). This is the meaning of our original assertion … that any probability statement relates to a collective, an aggregate phenomenon, rather than to an individual isolated event.
>
> The probability of dying within the coming year may be p_1 for a certain individual if he is considered as a man of age 42, and it may be $p_2 \neq p_1$ if he is considered as a man between 40 and 45, and p_3 if he is considered as belonging to the class of all men and women in the United States. In each case, the probability value is attached to the appropriate group of people, rather than to the individual.

Making probability statements about a particular outcome depends critically on the choice of labels in the collective.

> Another example: if one draws thirteen times from a box containing the 26 letters of the Roman alphabet and the letters obtained in succession form the word "Massachusetts," one would be inclined to say that something very surprising has happened, that the probability for this event to happen is extremely small. On the other hand, each of the $26^{13} \sim 2.4 \times 10^{18}$ combinations of 13 letters out of 26 has, under normal conditions of drawing, the same probability. The obvious answer to this apparent discrepancy is that in our first reaction we think of a collective with a two-point label set only, the labels being: meaningful word or meaningless combination of letters; the probability of the first label is then indeed very much smaller than that of the latter. If, however, the label set consists of all the 26^{13} possible combinations, all these labels have equal probability.
>
> A similar example: it is not true that in a lottery with 1,000,000 lots the result "900,000" is less probable than any other number, if the collective has as labels the number of the lot drawn; it *is,* however, an "improbable" result if the attribute, in a different collective, consists of the number v of zeros, $v = 0, 1, 2, ..., 6$, at the end of the number of the lot drawn (here $v = 5$)."

Next von Mises introduces four types of operations on collectives (or transformations of collectives). The first operation, place selection, may be used to generate specific collectives that have the same limiting frequency and that may be useful in solving some problems. The other three operations are used to generate basic principles of probability that are typically assumed or postulated. The second operation, mixing, is used to derive conditions under which probabilities can be added. The third operation, partition, is used to derive the formula for a conditional probability. The fourth operation, combining, is used to derive the conditions under which probabilities can be multiplied. A mathematician may find these derivations to be somewhat ad hoc. But operations of this kind are necessary to build a science of probability based on firm connections with reality. To simply postulate and derive is not science for von Mises.

> We take it as understood that probability theory, like theoretical mechanics or geometry, is a scientific theory of a certain domain of observed phenomena. If we try to describe the known modes of scientific research we may say: all exact science starts with observations, which, at the outset, are formulated in ordinary language; these

> inexact formulations are made more precise and are finally replaced by axiomatic assumptions, which, at the same time, define the basic concepts. Tautological (= mathematical) transformations are then used in order to derive from these assumptions conclusions, which, after retranslation into common language, may be tested by observations, according to operational prescriptions.
>
> Thus, there is in any sufficiently developed mathematical science a "middle part," a tautological or mathematical part, consisting of mathematical deductions. Nowadays, in the study of probability there is frequently a tendency to deal with this mathematical part in a careful and mathematically rigorous way, while little interest is given to the relation to the subject matter, to probability as a science.
>
> This is reflected in the fact that today the "measure-theoretical approach" is more generally favored than the "frequency approach" presented in this book. Cramér (1937) very clearly expresses this point of view. "Following Kolmogorov (1933), we take as our starting point the observation that the probability p(A) may be regarded as an additive set function of the set A. We shall, in fact, content ourselves by postulating mainly the existence of a function of this type defined for a certain family of sets A in the space to which our variable point X is restricted and such that P(A) denotes the probability $X \in A$." And Halmos (1944): "Probability is a branch of mathematics. Numerical probability is a measure function, that is, a finite, non-negative, and countably additive function P of elements in a Boolean σ-algebra B such that …"
>
> Now, such a description of the mathematical tools used in probability calculus seems to us only part of the story. Mass distributions, density distributions, and electric charge are likewise additive set functions. If there is nothing specific in probability, why do we define "independence" for probability distributions and not for mass distributions? Why do we consider random variables, convolutions, chains, and other specific concepts and problems of probability calculus?

For von Mises, a science is first built inductively from reality – not by assuming some arbitrary postulates.

> The way we see it, probability is a highly mathematicized science, but it is not mathematics, just as hydrodynamics is not a branch of the theory of partial differential equations—although it suggests interesting and difficult problems in this field. Our aim in presenting probability theory as a mathematical science is to incorporate into the basic assumptions the idealized basic relations to reality, as this is done in mechanics and in other sciences. This approach has been criticized in the following comments: "the von Mises definition involves a mixture of empirical and theoretical elements which is usually avoided in modern axiomatic theories. It would, for example, be comparable to defining a geometrical point as the limit of a chalk spot of infinitely decreasing dimensions." The "mixture of empirical and theoretical elements" is, in our opinion, unavoidable in a mathematical science. When in the theory of elasticity we introduce the concepts of strain and stress, we cannot content ourselves by stating that these are symmetric tensors of second order. We have to bring in the basic assumptions of continuum mechanics, Hooke's law, etc., each of them a mixture of empirical and theoretical elements. Elasticity theory "is" not tensor analysis. Our definition of probability is comparable less to the "limit of a chalk spot" than to the definition of velocity as the limit of length per time or to the definition of specific density as the limit of mass per volume. Yet, all these definitions also have something of the chalk spot: the transition from observation to theoretical concepts cannot be completely mathematicized. It is not a logical conclusion but rather a choice, which, one believes, will stand up in the face of new observations.

For von Mises, a science is first built empirically and inductively from reality so it is verifiable – not by assuming or positing some axioms or postulates with no necessary basis in reality. Verifying conclusions derived from such axioms is not sufficient to validate these axioms since they may generate other conclusions that cannot be verified.

> Let us now state more specifically what we consider to be the essential features of the theory presented so far. First of all: it is a frequency theory of probability, in contrast to a "measure theory" of probability. Of course, in a "frequency theory" as well as in a "measure theory" the concepts of sets and their measures play an important role. Likewise, the concept of frequency appears in every measure theory of probability. However, by "frequency theory" or "frequency approach" we mean the following: Whenever the term probability is used it relates to a (limit of a) frequency; and this frequency must be approximately verifiable, at least conceptually. If we say that the probability of "6" for a die equals 1/5, this implies that if we toss the die 10,000 times "6" will appear about 2000 times. The verification need not be so immediate but, directly, or indirectly by its consequences, some verification should be conceivable. This is the view of v. Mises, Tornier, and Wald. On the other hand, if probability is defined as the measure of a set and a relation to frequency is not incorporated in the theory but follows somewhere as an "obvious" interpretation and with no necessary relation to verifiability, we speak of a measure theory of probability. As representatives we name Kolmogorov, Cramer, Fréchet, and, to some extent, Laplace.

One consequence of this difference in approach involves the interpretation of the law of large numbers.

> Forming with the symbols 0 and 1 all combinations of n symbols we obtain 2^n different combinations (binary numbers). It can be shown that, for large enough n, the great majority of these 2^n numbers contains n/2 zeros and n/2 ones. More precisely: Jacob Bernoulli derived (1713) the following theorem: the larger the n, the larger the proportion of those binary numbers in which the relative numbers of zeros (or of ones) deviates from ½ by less than a given ε. Obviously this is a purely arithmetic property of numbers (of binomial coefficients). But Bernoulli himself and most authors state the result in the following way: if one throws a "true" coin long enough, it is almost certain that the relative number of heads will deviate by less than ε from ½. Certainly, this does not follow from the combinatorial theorem. The transition from that arithmetic theorem to a statement about "occurrence" can be justified only by defining a true coin (or any coin of probability p for "heads") in a way which establishes a connection between p and the frequency of occurrence of the event.
>
> The logical situation remains the same in the case of the sharper result expressed in the "strong" law of large numbers. We maintain the simple fact that if the deductive mathematical part of a theory is developed from axioms merely specifying the nature of the basic variables and functions involved, then relations to experience cannot be "derived mathematically." One cannot get out what one has not put in.

A second consequence is the issue of whether 'probability' is meaningful in situations where it is unverifiable.

> Great care must be given to incorporate the relation between theory and experience in an explicit and responsible way. This leads to our second comment. In measure theories of probability where all deductions are based on the measure definition, the relation to frequency is then often introduced as a more or less vague afterthought.
>
> From the works of Tornier, Wald, and Copeland, it appears that probability in a verifiable sense cannot be assigned to all measurable sets and not even to all Borel sets. The postulate worded by Cramér, that "any probability assigned to a specific event, must, in principle be liable to verification" is in contrast to axioms which assign probabilities to all Borel sets. In fact, if we are serious about the principle, instrumental for probability as a science, that an approximate verification should be conceivable for any probability statement then we arrive at the result that the sets of a well-defined field – quite different from that of Borel sets – and only these, should be assigned probabilities. We can use such sets and their measures freely in our deductions and conclusions, but, in the end, "probability" in contrast to measure should be restricted to situations which may be conceptually verified. In a sense, this is comparable to the auxiliary use of complex numbers in mechanics and electrotechnics.

For von Mises, the focus on the process of forming inductive generalizations that are empirically verifiable was much more important than the form of the result.

> While we think that these features of frequency theories are essential, it is, perhaps, less important whether a theory is presented in an axiomatic form or not. One can certainly present our theory in an axiomatic form. We preferred the approach presented here where, to quote one example, the fact that probability is a σ-additive set function *followed* from its frequency-explanation.
>
> The reader might be reminded that our collective is already of considerable generality. [Next] we shall extend this mathematical theory (with verifiability as the guiding principle) and shall obtain a mathematical theory of probability which is at the same time a scientific theory of probability.
>
> We add, finally, that our setup seems of some pedagogical value. For the student of our theory it is natural to analyze each problem with regard to the collectives involved and the operations employed. Thus, the probabilistic side of a problem is stressed in contrast to the mere analytic side. Many apparent contradictions are resolved and difficulties removed by an analysis of the collectives and operations. It needs no effort to find in the literature on probability instances of conceptual mistakes juxtaposed with faultless mathematics.
>
> Today, we see at one extreme the "consumer of statistics" who wishes to apply ready-made statistics to his problem, be it in medicine, education, or linguistics. His desire is all to often to use statistical methods like prescriptions, recipes, whose application should require very limited statistical and even less probabilistic knowledge and understanding and a minimum of mathematics. At the other extreme, stand those mathematicians who are exclusively interested in the mathematical aspect of some problem, some part of the theory, and who therefore consider probability to be a branch of mathematics; who teach us that probability "is" measure theory, Boolean algebra, etc., and dismiss frequency theory as "awkward mathematically." Between the extremes, the conception of probability theory as a mathematical science leads to a frequency theory of probability, much in need of the mathematicians' ideas and ingenuity, but free of a confusion of task and tool.

3. Comments by Others

Two groups found von Mises' approach irrelevant. First, those who see the theory of probability as a part of mathematics and thus not needing an empirical-inductive basis; second, those who see scientific theories as merely an arbitrary choice of axioms that can be used to generate interesting conclusions which may not be empirically verified. This difference in approach may explain some of the comments by others on von Mises approach.

Dodd (1936) was positive; he encouraged others to study von Mises' approach.

> Mises distinguishes his theory from: (1) the theory of Laplace and his followers who use "equally likely cases"; (2) the 'theory of Venn, Fechner, Bruns, and Helm, who use the first axiom without getting to the second; the theory of Copeland, Reinhenbach, Popper, Tornier, Kamke, and Dőrge, who in some way restrict the arbitrariness of the choice of the subset under Axiom 2. To those especially interested in foundations and ultimate conceptions, the book will be of extremely great value.

However, von Mises' ideas were not unanimously well received. Aleksander Ostrowski said of him, "Only with the appointment of Richard von Mises to the University of Berlin did the first serious German school of applied mathematics with a broad sphere of influence come into existence. Von Mises was an incredibly dynamic person and at the same time amazingly versatile like Runge. He was especially well versed in the realm of technology." But he also wrote: "Because of his dynamic personality his occasional major blunders were somehow tolerated. One has even forgiven him his theory of probability."

Feller (1957) indicated that a proper theory of probability should have a Frequentist meaning and verifiability but didn't require that such a theory be developed empirically and inductively as did von Mises. Feller upheld the rule-oriented nature of mathematics which gave a strong rebuttal to von Mises' approach. (Emphasis in original.)

> strong law of large numbers ... theorem on the impossibility of gambling systems ... together describe the fundamental properties of randomness which are inherent in the intuitive notion of probability and whose importance was stressed with special emphasis by von Mises.
>
> Axiomatically, mathematics is concerned solely with relations among undefined things. This property is well illustrated by the game of chess. It is impossible to "define" chess otherwise than by stating a set of rules. ... Similarly geometry does not care what a point and a straight line "really are." They remain undefined notions, and the axioms of geometry specify the relations among them: two points determine a line, etc. These are the rules, and there is nothing sacred about them. We change the axioms to study different forms of geometry, and the logical structure of the several non-Euclidean geometries is independent of their relation to reality. Physicists have studied the motion of bodies under laws of attraction different from Newton's, and such students are meaningful even if Newton's law of attraction is accepted as true in nature.
>
> In contrast to chess, the axioms of geometry and of mechanics have an intuitive background. ... Certainly intuition can be trained and developed.
>
> Even the collective intuition of mankind appears to progress. Newton's notion of a field of force and of action at a distance and Maxwell's concept of electromagnetic "waves" were at first decried as "unthinkable" and "contrary to intuition."
>
> In applications, the abstract mathematical models serve as tools, and different models can describe the same empirical situation. *The manner in which mathematical theories are applied does not depend on preconceived ideas; it is a purposeful technique depending on, and changing with, experience.* A philosophical analysis of such techniques is a legitimate study, but it is not within the realm of mathematics, physics or statistics. The philosophy of the foundations of probability must be divorced from mathematics and statistics, exactly as the discussion of our intuitive space concept is now divorced from geometry.

Kolmogorov (1963)[2], whose rival axiomatization was better received, was less severe:

> The basis for the applicability of the results of the mathematical theory of probability to real 'random phenomena' must depend on some form of the frequency concept of probability, the unavoidable nature of which has been established by von Mises in a spirited manner.

Initially Kolmogorov relied in part on von Mises approach. According to Vovk and Shafer (2003),

> This formalization [Kolmogorov (1933)], *measure-theoretic probability*, has served and still serves as the standard foundation of probability theory; virtually all current mathematical work on probability uses the measure-theoretic approach. To connect measure-theoretic probability with empirical reality, Kolmogorov used

[2] www-history.mcs.st-andrews.ac.uk/Biographies/Mises.html

> two principles, which he labeled A and B. Principle A is a version of von Mises's requirement that probabilities should be observed frequencies.

But ultimately Kolmogorov turned away from the von Mises approach.

> As we have just seen, Kolmogorov was thinking in 1963 about using algorithmic complexity as the starting point for a frequentist definition of probability, in the style of von Mises. By 1965, however, he was thinking about defining randomness directly in terms of algorithmic complexity, in a way that also allows us to connect randomness more directly to applications, without any detour through the idea of frequency. We call this approach Kolmogorov's *finitary theory of algorithmic randomness*.

A book by von Mises (1956) countered with his views on science and life based on the doctrine of positivism.

> Positivism does not claim that all questions can be answered rationally, just as medicine is not based on the premise that all diseases are curable, or physics does not start out with the postulate that all phenomena are explicable. But the mere possibility that there may be no answers to some questions is no sufficient reason for not looking for answers, or for not using those which are attainable.

Kline (1980) upheld mathematics as an empirical science and supported the need for verifiability – at least in the long term – in the physical world.

> sound mathematics must be determined not by any one foundation which may some day prove to be right. The "correctness" of mathematics must be judged by its application to the physical world. Mathematics is an empirical science much as Newtonian mechanics [is]. It is correct only to the extent that it works and when it does not, it must be modified. It is not a priori knowledge even though it was so regarded for two thousand years. It is not absolute or unchangeable.

Ultimately, the issue is one of philosophy. Is the theory of probability to be a branch of formal mathematics or is to be an independent science – a science of data? By the end of the 20th century, there were still references to von Mises. See Dalen (1990), van Lambalgen (1996) and Vovk and Shafer (2003). But there seemed to be no prominent statistician who supported von Mises' view that the theory of probability should be an empirical science rather than an applied branch of mathematics.

So why was von Mises unsuccessful in his quest for an empirically and inductively based theory of probability? His need to relocate from Germany to Turkey and to Harvard may not have allowed him to educate intellectual successors. More fundamentally, he may have been unsuccessful because he was opposing the dominant trends in the philosophy of mathematics – trends such as formalism, intuitionism and logicism. Whatever their differences, these trends all agreed on the power of the mind to postulate mathematical entities and relationships independent of any empirical content or limits. Von Mises' battle was not just a battle with mathematicians – his battle was ultimately with those philosophers of science who could not embrace an empirical inductive approach for mathematics.

4. Summary

One might summarize von Mises as objecting to any arbitrarily-postulated theory of probability whether a purely axiomatic development that gives no guidance about valid applications or a measure theory approach in particular that is known to give unrealistic answers.

5. Recommendations

One of von Mises' recommendations was to reserve 'probability' for just those processes that generate sequences that are random – that have a limiting value and that are insensitive to place selection. As one commentator said at our JSM presentation, "That train has already left the station."

Statistical educators should expose students to an empirical-inductive approach to the foundations of statistics. Given the ongoing disarray among philosophers on the nature of science, students should be broadly educated. This need not require an exposure to the totality of such an induction as presented by von Mises, but it should indicate the connections and the movement from observable facts to unobservable principles.

Statistical educators should uphold von Mises' observation that probability always requires a context for it to have unambiguous meaning. Saying that the 13 letter combination of 'Mississippi' is as likely as any other combination of letters is improper. It assumes, but fails to state, the context is 'all 13 letter combinations.' In a different context, meaningful combinations versus nonsense combinations, meaningful combinations are far less likely than nonsense combinations. By reiterating the crucial importance of context, statistical educators can help to locate statistics as a science involving an empirical application of mathematics rather than as a very narrow branch of mathematics. We often talk about "the probability of heads" since the context is readily determined from the outcome, but we should noted that "the probability of death" is completely ambiguous.

Statistical educators should disavow and disallow each and every statement of probability that applies to a single isolated individual when made without any reference to their membership in a group. For example, journalists may say, "your chance of dying in a plane crash is less than your chance of being killed by lightning." This is only true if "you" is a code for "you as a randomly selected member of the population." Neither of these probabilities is constant for all members of the population. Fliers are much more likely to die in a plane crash than non-fliers. Golfers and farmers are much more likely to be struck by lightning than office workers. A more extreme and obviously misleading example is to say, "your chance of giving birth in the next year is about 1.5%" when there are 15 births per 1000 population. Obviously this chance does not apply to most of the population.

Statistical educators should disavow assuming that outcomes have a 'natural' context. They should no longer criticize students who say that getting a mixture of heads and tails from random flips of a fair coin is more likely than getting a sequence where the first half are all heads and the half are all tails. It all depends on the choice of labels.

Statistical educators should emphasize that extreme care must be taken in making probability statements about sequences whose relative frequencies change with time. While the relative frequencies of social phenomena such as the unemployment rate, the birth rate and the marriage rate can change considerably over time, so can the relative frequency of biological phenomena. For example, the excess percentage of males among US births has decreased from 6% in 1920 to less than 5% after 1990.

Statistical educators should emphasize that using 'probability' to describe sequences that are not potentially infinite (eg. results of elections or of sports matches) is ill-advised. Using such sequences as a source for probabilities may satisfy a psychological need but these 'probabilities' are not scientific – they are not reliable evidence.

Statistical educators might consider deriving the theory of probability for continuous distributions by refounding them on intervals. Some of the axioms may not be derivable but this approach could be shorter and more teachable than von Mises' approach to continuous distributions.

Finally, statistical educators should support those engaged in the philosophy of science in their quest to formulate statistics as a science of data. Perhaps these philosophers may be able to apply or modify von Mises' empirical-inductive approach to today's world. This seems to us to be a very worthy task.

Acknowledgments

This work was performed under a grant from the W. M. Keck Foundation "to support the development of statistical literacy as an interdisciplinary curriculum in the liberal arts." This paper may be found by searching for *2008SchieldBurnhamASA.pdf* at www.StatLit.org.

References

Cramér, H. (1937). *Random Variables and Probability Distributions*. Cambridge

Cramér, H. (1953). "Richard von Mises' work in probability and statistics", *Ann. Math. Statistics* 24, pp. 657-662.

Dodd, Edward L. (1936). Reviewed Work: *Wahrscheinlichkeit, Statistik und Wahrheit* by Richard Von Mises *Journal of the American Statistical Association*, Vol. 31, No. 196. (Dec., 1936), pp. 758-759.

D.v. Dalen, "The War of the Frogs and the Mice or the Crisis of the 'Mathematische Annalen'", *The Mathematical Intelligencer* 12 (1990), No.4, pp. 17-31.

Feller, William (1957). *An Introduction to Probability Theory and Its Applications*. Volume 1, 2nd edition, Wiley.

Frank, P. (1954). "The work of Richard von Mises: 1883-1953", *Science* 119, pp. 823-824.

Halmos, P. R (1944). The Foundations of Probability, *Amer. Math. Monthly* 51 pp. 493-510.

Kline, Morris (1980). *Mathematics: The Loss of Certainty*. Oxford University Press (1982 paperback)

M. van Lambalgen, "Randomness and foundations of probability: von Mises' axiomatisation of random sequences", in *Statistics, probability and game theory*, Hayward, CA, 1996, pp. 347-367.

Presented to Richard von Mises by Friends, Colleagues and Pupils, *Studies in Mathematics and Mechanics*, New York, 1954.

von Mises, Richard (1936). *Probability, Statistics and Truth*, 2nd rev. English ed., New York, Dover, 1981

von Mises, Richard (1956) *Positivism: A Study in Human Understanding*, G. Braziller, 1956. (Pb, Dover, 1968).

von Mises, Richard (1964). *Mathematical Theory of Probability and Statistics*. Edited and complemented by Hilda Geiringer. New York, Academic Press.

Vovk, Vladimir and Glen Shafer (2003). *Kolmogorov's contributions to the foundations of probability*. The Game-Theoretic Probability and Finance Project, Working Paper #6. www.probabilityandfinance.com