

Chapter 10

Selected Material on the Work of Richard von Mises



Boris Hessen

The Statistical Method in Physics and the New Foundation of Probability Theory

1. A Collective and its Distributions

Formulation of the Problem

Exact natural science has been more and more conquered by the statistical method. However, in spite of its wide and fruitful application there is no sufficient clarity regarding its place among the number of other methodological tools and its methodological essence. And this is primarily reflected in the views on probability theory.

Probability theory is a mathematical tool for the study and expression of statistical laws. Therefore, it is only natural that our view of probability theory is determined by the evaluation of the character and significance of statistical laws in the general system of laws established by science. And conversely, the evaluation of probability theory leads to a certain view of statistical laws which it studies and expresses.

A very definite view of probability theory as an area of study where, due to the limitation of our knowledge, the full investigation of phenomena is impossible has prevailed since the time of its founders, Bernoulli and Laplace. We know something of the phenomena, but our knowledge is extremely incomplete. Laplace said, "Probability is relative, in part to this ignorance [of natural laws], in part to our knowledge".¹

¹BH: Laplace, *A Philosophical Essay on Probabilities*, 1908, p. 11. TN: Laplace (1902, p. 6).

TN: Translated from *Estestvoznanie i Marksizm (Natural Science and Marxism)*, 1, 1929. (Sections 1 and 2 and the Summary).

B. Hessen (✉)
Abingdon, UK
e-mail: chris.talbot@protonmail.com

Probability theory is a result of our partial lack of knowledge. It does not *overcome* it but only partly makes up for it. This is why probability theory provides surrogate and not complete knowledge, and this makes it different from other mathematical disciplines.

According to Laplace probability theory is applied in the domain of chance. There is no fundamental or objective difference between the phenomena studied by mechanical methods and by those of probability theory. Each “curve described by a simple molecule of air or vapour is regulated in a manner just as certain as the planetary orbits; the only difference between them is that which comes from our ignorance.”² We call chance a necessary phenomenon whose reasons we do not know. Chance is an unknown necessity and therefore, it has a purely *subjective* character.

This is why Laplace, by defining the area of application of probability theory as an area where chance is king, i.e. lack of knowledge, gives probability theory a subjective and transient character. The more knowledge we gain, the less is the area of application of probability theory.

Therefore, it is but a temporary crutch for our lack of knowledge and a surrogate for factual knowledge.

This subjective interpretation of probability theory traditionally has passed into modern science and thus made the definition of both probability theory’s application and its tasks, and of the methodological value of its results, very difficult and obscure.

There is no doubt about the significance of R. von Mises’s [contribution to the] precise limitation of the area of application of probability theory and his clear definition of its tasks and, as it were, his giving it completely “equal rights” with the other mathematical disciplines, and mainly his clear definition of the objects studied by probability theory. Von Mises’s approach is completely opposite to that of Laplace. The subject of probability theory and the area of its application are determined by the objective properties of the object studied (collective) and not by our partial lack of knowledge.

The object of study by probability theory is defined as objectively, precisely and definitely as is the object of study by geometry. Its laws and conclusions regarding the object of study are as strict and precise as those of geometry.

The task of science is to study the objects and processes of the external world.

The study of phenomena makes it necessary to solve physical and mathematical problems and tasks. The character of these problems is determined by the specific structure of the studied objects. The methods of solution are most effective when they are best adapted to the characteristic structure of the studied object. Probability theory emerged and exists because the objects of a specific structure (collective) exist and their study results in the setting of quite definite tasks that require their study and solutions. Probability theory studies and finds solutions to the mathematical problems that are posed by the study of the area of phenomena where the collective plays the key role. Therefore, probability theory does not differ from other mathematical

²TN: Laplace (1902, p. 6).

disciplines in any way in the *origin* of problems, in their *processing* or in the *character of the results* that are to be used and realised in practice.

Von Mises aims at constructing probability theory in such a way that it would be a well founded and precise tool of the mathematical processing of problems that is as adjusted to the structure of the studied object as analysis or geometry.

A mathematical definition of probability is at the basis of the edifice of probability theory. It allows us to construct probability theory as a mathematical discipline. The definition of probability is the first principle³ in Laplace's construction. This is a false definition because it contains an obvious *petitio principii*. The concept of probability is reduced to the concept of "equally possible cases". Equally possible cases are nothing other than equally probable cases. Notwithstanding the fact that this definition of probability is completely *inapplicable in practice* (we shall say more about this later) it is logically untenable.

According to Laplace, "cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence."⁴ This definition follows from a subjective interpretation of the concept of probability and well illustrates the helplessness of the subjective concept in the rationalisation of probability theory.

If the objective interpretation of probability is used as a corner stone, as is done by von Mises, then the rationalisation and the construction of probability theory should not start with the definition of probability but with the *characteristics of those objects whose peculiar structure leads us to establishing the concept of probability*.

The definition of *probability* may and should be given only after specific features of the structure of the objects to which it refers are clearly established.

This approach to the problem allows us to precisely and objectively define probability and at the same time to unequivocally limit the area of application of this concept, and therefore, *the area of application* of theory of probability.

We shall look at the definition and examination of these specific objects (called *collectives* by von Mises) whose existence is at the basis of the construction of probability theory.

2. Statistical and Dynamical Laws. Additive and Non-additive Aggregates

As soon as the kinetic theory of gases began to develop, Maxwell, with his typical foresight, posed a question of the new type of laws introduced by this theory. In one of his lesser known philosophical essays he says, "But I think the most important effect of molecular science on our way of thinking will be that it forces on our attention the distinction between two kinds of knowledge, which we may call for convenience the Dynamical and Statistical."⁵

³BH: Laplace, *A Philosophical Essay on Probabilities*, 1908, p. 15. TN: Laplace (1902, p. 11).

⁴TN: Laplace (1902, p. 6).

⁵BH: Maxwell looks into the problem of necessity and chance applied to physics in a small article published by Campbell and Garnett – Maxwell's paper was given to a philosophical group at Cambridge (Club of Seniors). The article is titled "Does the progress of Physical Science tend to give any advantage to the opinion of Necessity (or Determinism) over that of the Contingency of Events and the Freedom of the Will?" TN: See Campbell and Garnett (1882, pp. 209–213). The original English is taken here from p. 210. See Chap. 4, p. 57, n. 45.

One of the specific and key features of the difference between dynamical and statistical laws is the fact that dynamical laws describe a *single or individual phenomenon* while statistical ones describe an *aggregate* of individuals or phenomena.

Although statistical laws cannot be used in the study of individual phenomenon because they simply make no sense with regard to an individual, the application of dynamical laws in the study of aggregates or collectives is not so simple. Every aggregate consists of individuals. Therefore, on first sight it would seem that dynamical laws apply well to the study of aggregates.

However, it is not so simple.

The same dual approach could be applied to statistical laws and to the concept of probability: an aggregate consists of a vast number of individuals. The behaviour of each individual is unequivocally determined by a dynamical law. We can study an aggregate as a sum of an immense number of dynamical laws. Such an approach is possible but extremely difficult. This is why we turn to statistical laws although they are second class knowledge in comparison to dynamical laws. However, the former do successfully overcome both the lack of our knowledge and partly, the extraordinary difficulties in studying aggregates arising from the immense number of individuals.

Typical arguments for the *subjective* approach to probability are easily recognised in this reasoning. In this sense statistical laws result from the limitation of our cognitive powers. The *objective* approach to statistical laws essentially means that the *raison d'être* of these laws is not in our limited knowledge but in the characteristic structure of aggregates or objects studied by them.

By using the statistical method in the study of an aggregate we examine it as a *whole*.

An aggregate, although consisting of separate elements, is not divided into separate elements in the process of our study. We study it as a whole or synthetically. Therefore, our statistical laws are applicable to the aggregate only as a *whole* and have no sense if applied to separate elements.

But what does it mean that statistical laws are valid for an aggregate as a whole and not applicable to its separate elements? In other words, what is the relationship between the dynamical laws that govern separate elements of the aggregate and statistical laws applicable only to the aggregate as a whole? Two fundamentally different approaches are possible here.

First: the laws that characterise an aggregate as a whole can be fully deduced from dynamical laws for the individuals. In this case, *in nuce* these laws are already expressed in each individual law. *Laws for the whole are a quantitative sum of laws for single phenomena.*

According to this approach an aggregate assumes a clearly additive character. It is divided without any remainder into the sum of its elements. There is a quantitative and not a qualitative difference between the whole and its parts. Laws for the whole are a sum of individual laws just like a kilogram is a sum of grams.

It is this concept of an aggregate as a purely additive formation that is the basis of the subjective interpretation of the statistical method. Indeed, if laws for the whole are simply a sum of laws for the parts then the sum of these laws is the only way to

the precise knowledge of the laws for the whole. Since the law for the whole is a sum of laws for the parts then the law for the whole will be fully and completely known only through its cognition as a total sum of individual laws.

A dynamical law is a true and objective law. A statistical law is a roundabout way towards [obtaining] the knowledge of the law for the whole as a sum of dynamical laws, a temporary crutch for our ignorance, and although it allows us to move further and further forward, we move with the gait of a limping cripple and not of a healthy man.

A second approach is first of all based on the difference between the additive and non-additive properties of aggregates. The above approach is valid only with regard to specific properties, i.e. additive ones, of aggregates and not with regard to all their properties. Additive properties mean that an aggregate as a whole is a simple sum of parts, and the laws for the whole are a sum of the laws for the parts, just as the weight of a whole is the sum of weights of parts. There is nothing qualitatively new in the weight of the whole compared to the weights of its parts. The weight is equally easy to divide into components and put them together again. The only difference between the whole and the parts is the quantity.

However, apart from the additive properties (e.g. weight) an aggregate may have significantly different properties which are not included in each of its elements and belong only to the aggregate itself. When the aggregate disintegrates into its individual components these properties disappear in the same way as chemical properties of a compound disappear during its decomposition into components.

Non-additive properties are typical for an aggregate as a whole, and only for it. They are not virtually contained in its individual components. They are manifested only in the whole and *qualitatively* differ from the properties of individuals.

Statistical laws reflect and study precisely these non-additive properties and therefore, statistical laws by their essence clearly cannot refer to the single individuals which compose an aggregate. This is *not their flaw* but their characteristic *peculiarity* because they deal with the study of precisely those properties that are manifested only in the whole and *do not exist* in single elements.

By accepting that the presence of non-additive properties in aggregates is a peculiarity of their objective structure we thereby give statistical laws an objective character.

In this respect the relation between statistical and dynamical laws is the relation between the laws of the whole and of the part. Dynamical laws are individual laws. However, they are not sufficient for the study of the law of the whole because the whole has both additive and non-additive properties.

Statistical laws do not negate or contradict dynamical laws. They are necessary and valid in their area and dynamical laws in theirs.

M. Planck rightly notes that a dynamical law is a condition for the occurrence of a statistical law.⁶ However, this does not mean that a statistical law is reduced without a remainder to a dynamical law. This is correct only if the whole is identical to the sum of its parts.

⁶TN: Planck (1960, pp. 56–68). See also Chap. 8, p. 101, n.5.

And if this is not so (and the existence of non-additive properties in an aggregate confirms the absence of this identical relationship), then a statistical law genetically arises from dynamical laws in the same way as a whole originates from a part; however, it is not composed of or decomposed into dynamical laws but is a qualitatively new formation only inherent in the whole and not in its parts. Therefore, a statistical law is not second-class knowledge compared to a dynamical law. It is a fully and equally valid method of cognition that results from the objective structure of the objects of study.

We have mentioned earlier that probability theory is the mathematical apparatus of the statistical method. The reasoning applied above to statistical laws can be applied to the concept of probability. Most importantly, probability characterises an *aggregate as a whole* and not its every single element.

Major mistakes in the interpretation of probability arise when one tries to apply the concept of probability to a single individual which is a part of a collective.

As von Mises correctly notes, this is the essence of Marbe's mistake⁷; the latter assumed that the longer only boys continue to be born, the higher is the probability of a boy being born again.

Probability characterises an aggregate as a whole. It does not refer to a single individual in the aggregate and cannot be deduced from it. Probability is a non-additive property of an aggregate.

However, the concept of probability cannot be applied to every aggregate that possesses non-additive properties.

Therefore, we now turn to the definition and description of properties of aggregates which are objects for the application of probability theory. According to von Mises' terminology we shall call these aggregates a *collective*, as opposed to other aggregates.⁸

Preface to On Causal and Statistical Laws in Physics⁹

Philosophical and methodological problems in physics featured rather prominently at the Fifth Congress of German physicists and mathematicians in Prague.

P. Frank's report "How modern physical theories contribute to the theory of cognition"¹⁰ focused on the issues of epistemology. *Von Mises's* report whose translation is published in this issue of our journal focused on the problem of a physical law.

The development of quantum mechanics resulted in a new approach to the interrelation between dynamical and statistical laws. The problem of a statistical method as a separate way of expressing physical laws became most urgent. The extraordinary

⁷TN: von Mises (1964, p. 184).

⁸TN: This concludes the first two sections of Hessen's article. The remaining sections are an exposition of von Mises' statistical work, closely following von Mises (1957).

⁹BH: A report given at the Fifth Congress of German physicists and mathematicians in Prague on 16 September, 1929. *Naturwissenschaften* 1930, transl. by E. L. Starokodamskaya. TN: Translated from *Uspekhi Fizicheskikh Nauk (Advances in Physical Sciences)*, X, No 4, 1930, pp. 437–439. The Preface is to a Russian translation of von Mises (1930).

¹⁰TN: Frank (1930).

development of statistical physics drew attention to the problems of the theory of probability. The classical explanation of probability theory based on the subjective concept of chance has clearly become unsatisfactory.

In his article below *von Mises* attempts to approach the relationship between statistical and dynamical laws using his concept of the probability theory based not on the subjective understanding of randomness but on the notion of a collective.

In his epistemological ideas *von Mises* generally joins *P. Frank* who adopts a Machist stance.

Von Mises's explanation of regularities in physics could not help but reflect his unsatisfactory epistemological views.

As to the cardinal problem of causality in physics, although *von Mises* does not adopt the extreme views defended by *Heisenberg* and *Dirac* who reject the elementary causality of physical phenomena, he still does not give a sufficiently clear explanation of this problem. However, he tries to demonstrate that the concept of statistical laws does not exclude the concept of causality.

Equally unacceptable are *von Mises's* arguments on the relationship of physical and philosophical concepts which essentially echo *Frank's* idea of "natural philosophy".

A healthy grain of protest against philosophical dogma and fantasies of the philosophy of science is completely devalued by the rough pragmatic approach to solving this problem. If we adopt *von Mises's* view and agree that the sole task of philosophy is to adapt to the solutions of this problem given by physics, then it becomes unclear why we need a philosophical examination of the problem and why it is being discussed at a congress of physicists.

All this is because *von Mises* does not draw a distinction between philosophical dogmas and true philosophy.

The unsatisfactory and inconsistent character of *von Mises's* philosophical conceptions result in his metaphysical explanation of the concept of limit and limiting case that play a large role in the uncertainty relation.

Von Mises does not pose the question of the relationship between an indefinite approximation to a given limit and a limit as initially given in every single act of observation. Therefore, no correct solution is found to the problem of divisibility and indivisibility, a very significant problem for the uncertainty relation.

Despite all these drawbacks the article by *von Mises* below is of great interest because it attempts to consider the problem of the relationship between statistical and dynamical laws using a new approach contrary to the theories that find the sole solution to the problem in the expulsion of causality from physics.

Editor's Note—CT

Von Mises's theory, like Smolukowski's, stresses the objective nature of probabilities. As Hessen explains in the first article above, he developed the concept of a "collective", a special kind of aggregate with "non-additive properties." Thus a probability relates to the collective and not to individual elements in it.

This is not the place to consider the history of statistics and the central role played by von Mises's approach, especially in the period that Hessen was writing.

We merely note that Andrey Kolmogorov, the Russian founder of the current, widely used, approach to probability and statistics, developed in the 1930s, based the application of probability on von Mises's ideas.¹¹ As von Plato explains, an unnecessarily dismissive attitude to von Mises seems to have developed in recent decades.

Hessen has to distinguish between von Mises's statistical methodology, as it could be applied in science, and his positivist philosophy. Just the fact that he wrote about von Mises was enough to foster attacks from Stalinist critics.

According to von Mises the theory of quantum mechanics showed that there was "indeterminism" from the macro-scale all the way down to the micro.

Until recently, we thought that there existed two different kinds of observations of natural phenomena, observations of a statistical character [macro], whose exactness could not be improved beyond a certain limit, and observations on the molecular scale [micro] whose results were of a mathematically exact and deterministic character. We now recognize that no such distinction exists in nature.¹²

Even allowing for the Heisenberg principle, according to von Mises all measurements could now be seen to have a purely statistical character. Von Mises does not distinguish between the macro and the micro level.

Von Mises's philosophy had led, as Hessen writes above,¹³ to a "metaphysical explanation of the concept of limit" which in turn supported his purely statistical approach to quantum theory. For Hessen understanding the concept of limit required the realisation that there is a relationship between the "indeterminate approximation to a limit" and the actual limit that was "initially given." This is not very clear, let us attempt to explain.

It seems that Hessen is pointing to the fact that at the quantum level, a quantitative process could lead to a limit where there is a qualitative change in outcome. This outcome could include discrete or indivisible "jumps", for example in the energy levels of an atom which emits distinct characteristic spectral lines. Such jumps do not happen at the macro level.

Engels explains this "problem of divisibility and indivisibility" with the following example:

If we imagine any non-living body cut up into smaller and smaller portions, at first no qualitative change occurs. But this has a limit: if we succeed, as by evaporation, in obtaining the separate molecules in the free state, then it is true that we can usually divide these still further, yet only with a complete change of quality. The molecule is decomposed into its separate atoms, which have quite different properties from those of the molecule.¹⁴

¹¹ von Plato (1994), Chap. 6, especially p. 179.

¹² von Mises (1957, p. 217).

¹³ See p. 137.

¹⁴ *Dialectics of Nature*, Engels (1988, p. 358).

References

- Campbell, L., & Garnett, W. (1882). *The life of James Clerk Maxwell with the selection of his correspondence and occasional writings*. London: Macmillan.
- Engels, F. (1988). *Marx Engels collected works* (Vol. 25). Moscow: Progress Publishers.
- Frank, P. (1930). Was bedeuten die gegenwärtigen physikalischen theorien für die allgemeine erkenntnislehre? *Erkenntnis*, 1, 126–157.
- Laplace, P. S. (1902). *A philosophical essay on probabilities* (F. W.Truscott & F. L.Emory, Trans.). London: Chapman and Hall.
- Planck, M. (1960). *A survey of physical theory*. New York: Dover.
- von Mises, R. (1930). Über Kausale und Statistische Gezetzmäßigkeit in der Physik (On Causal and Statistical Laws in Physics). *Die Naturwissenschaften (The Science of Nature)*, 18(7), 145–153.
- von Mises, R. (1957). *Probability, statistics and truth* (p. 1928). Dover, New York: Second revised English edition. Original German edition published by Springer.
- von Mises, R. (1964). *Mathematical theory of probability and statistics*. New York: Academic Press.
- von Plato, J. (1994). *Creating modern probability, its mathematics, physics and philosophy in historical perspective*. Cambridge University Press.