

Mutual Influence between Different Views of Probability and Statistical Inference

Manfred Borovcnik

manfred.borovcnik@aau.at

<https://orcid.org/0000-0001-9044-5326>

University of Klagenfurt

Klagenfurt, Austria.

Recibido: 21/09/2020 Aceptado: 10/02/2021

Abstract

In this paper, we analyse the various meanings of probability and its different applications, and we focus especially on the classical, the frequentist, and the subjectivist view. We describe the different problems of how probability can be measured in each of the approaches, and how each of them can be well justified by a mathematical theory. We analyse the foundations of probability, where the scientific analysis of the theory that allows for a frequentist interpretation leads to unsolvable problems. Kolmogorov's axiomatic theory does not suffice to establish statistical inference without further definitions and principles. Finally, we show how statistical inference essentially determines the meaning of probability and a shift emerges from purely objectivist views to a complementary conception of probability with frequentist and subjectivist constituents. For didactical purpose, the result of the present analyses explains basic problems of teaching, originating from a biased focus on frequentist aspects of probability. It also indicates a high priority for the design of suitable learning paths to a complementary conception of probability. In the applications, modellers use information in a pragmatic way processing this information regardless of its connotation into formal mathematical models, which are always thought as essentially wrong but useful.

Keywords: Intuitive views. Foundations of probability. Bayesian controversy. Justification of probability. Bayes theorem. Statistical tests.

Influencia mutua entre diferentes puntos de vista de la probabilidad y la inferencia estadística

Resumen

En este trabajo, analizamos los diversos significados de la probabilidad y sus diferentes aplicaciones, y nos centramos especialmente en la visión clásica, la frecuentista y la subjetivista. Describimos los diferentes problemas de cómo se puede medir la probabilidad en cada uno de los enfoques, y cómo cada uno de ellos puede ser bien justificado por una teoría matemática. Analizamos los fundamentos de la probabilidad, donde el análisis científico del enfoque que permite una interpretación frecuentista conduce a problemas insolubles. La teoría axiomática de Kolmogorov no basta para establecer una inferencia estadística sin más definiciones y principios. Por último, mostramos cómo la inferencia estadística determina esencialmente el significado de la probabilidad y se produce un desplazamiento de las posiciones puramente objetivistas a una concepción complementaria de la probabilidad con componentes frecuentistas y subjetivistas. Con fines didácticos, el resultado de los presentes análisis explica los problemas básicos de la enseñanza, originados por un enfoque sesgado de los aspectos frecuentistas de la

probabilidad. También indica una alta prioridad para el diseño de vías de aprendizaje adecuadas a una concepción complementaria de la probabilidad. En las aplicaciones, los modelizadores utilizan la información de manera pragmática procesando esta información, independientemente de su connotación, en modelos matemáticos formales, que siempre se consideran esencialmente erróneos pero útiles.

Palabras clave: Vistas intuitivas. Fundamentos de la probabilidad. La controversia bayesiana. Justificación de la probabilidad. Teorema de Bayes. Pruebas estadísticas.

Influência mútua entre diferentes visões de probabilidade e inferência estatística

Resumo

Neste artigo, analisamos os vários significados de probabilidade e suas diferentes aplicações, e focamos especialmente na visão clássica, frequentista e subjetivista. Descrevemos os diferentes problemas de como a probabilidade pode ser medida em cada uma das abordagens, e como cada uma delas pode ser bem justificada por uma teoria matemática. Analisamos os fundamentos da probabilidade, onde a análise científica da teoria que permite uma interpretação frequentista leva a problemas insolúveis. A teoria axiomática de Kolmogorov não é suficiente para estabelecer inferência estatística sem mais definições e princípios. Finalmente, mostramos como a inferência estatística essencialmente determina o significado da probabilidade e uma mudança emerge de visões puramente objetivistas para uma concepção complementar da probabilidade com constituintes frequentistas e subjetivistas. Para fins didáticos, o resultado das presentes análises explica problemas básicos do ensino, decorrentes de um enfoque tendencioso em aspectos frequentistas da probabilidade. Também indica uma alta prioridade para a concepção de caminhos de aprendizagem adequados para uma concepção complementar de probabilidade. Nas aplicações, os modeladores usam informações de uma forma pragmática processando essas informações independentemente de sua conotação em modelos matemáticos formais, que são sempre considerados como essencialmente errados, mas úteis.

Palavras chave: Visões intuitivas. Fundamentals de probabilidade. A controvérsia Bayesian. Justificação da probabilidade. O teorema de baías. Testes estatísticos.

1. Introduction

Didactic is the melding of a discipline such as mathematics with any science that could be a promising source for illuminating this subject from a teaching perspective, such as (cognitive) psychology, sociology, philosophy of science, and so forth, all with the aim of getting theoretical and empirical insight into teaching or learning processes in order to improve actual instruction. The present paper focuses on the history of probability, on the ways, how probability can be perceived, and on how specific views of probability can be scientifically investigated. The investigation specifically directs towards the question whether a frequentist view of probability is justified and if not, which gaps in a justification process may be traced and how the view may be extended to account for a wider perception of probability, which may

guide a more transparent application of probabilistic models to real problems. From such an analysis, the field of didactics may get an orientation about the ways, how to design learning paths through a curriculum on probability.

The central idea of Section 2 is to highlight the differences in the various approaches towards probability, to show a complementarity between the different meanings of probability, to characterise the role of an axiomatic theory as a tool mediating between the different views, and to highlight the problems that arise when measuring probability within the various approaches.

Key idea of Section 3 is to investigate the justification of the various approaches towards probability, which extends and deepens the discussion about their differences. We describe distinct advantages of the approaches, which link to severe drawbacks: While more objectivistic theories of probability need the extension to statistical inference, where there are some concerns about the rationality of the suggested methods, the more subjectivist approaches would include inference within probability but they suffer from the subjective connotation of the concept of probability.

Section 4 is devoted to an analysis of statistical inference and Bayesian inference, which are the methods for inference that objectivist and subjectivist approaches towards probability provide. We use the term *statistical* inference to express a link to a *frequentist* probability view and *Bayesian* inference in connection to inferential methods based on a *subjectivist* meaning of probability. The central focus in this section is how those methods succeed to solve the inverse inference problem, which explicitly restrict probability to an objectivist meaning. The analysis shows that the various objectivist schools have their rationality gaps, which – by the end – introduces subjective elements into statistical inference through the backdoor.

In Section 5, we draw conclusions about the concept of probability, which – after the analysis – remains a complementary entity with subjectivist and objectivist constituents. The implication for teaching is to investigate learning paths that allow fostering the mutual interrelations between the various meanings of probability rather than focussing on a restricted frequentist view, which is prevalent. Concrete learning paths are still waiting to be developed and tested.

2. Probability and probability theory

The interrelations between singular concepts and a theory, in which these concepts are embedded, are mutual. On the one side, a concept gets its own meaning by the relations of the theorems within this theory. On the other hand, if someone pursues a specific meaning of a concept, then a usual way is to develop a theory around this interpretation and see how far-reaching such a theory is. Thus, the mutual relations between a theory and specific concepts clarify also the concept and that is not different with the concept of probability.

2.1 Roots of probability

Historically, the concept of probability emerged from games of chance, which were in use for a mixed purpose. The origin of such games connects to divination, to explore God's will (David, 1962). Consistently, ancient priests used games of chance, respectively the tools with which one could play such a game, for the purpose of divination. From that origin, the more general games of chance have developed primarily for entertainment. Yet, the way, how people wagered their stakes, often reminds us of the original use. Otherwise, one would not be able to understand how in the ancient Rome people wagered their own life once they had lost their fortunes and went to slavery without too much concern as – if so – it would have been God's will. The range of intuitions related to probability is wide and contradictory: to quantify speculations, hope, expectations, fear about unknown issues that may lie in the past or in the future. Mathematics was the ideal theory that should clarify feasible interpretations. By designing an axiomatic basis for probability, the expectation of the mathematics and physics community was to bring order into the chaos of contradictory interpretations.

2.2 Conceptions of probability

While it was comparably easy to find a satisfactory solution for an axiomatic basis for probability, the crux was to build up from that framework a theory of statistical inference that would allow for a connection of this conception to the real world. That is, a statistical theory of inference that would also justify the tight connection between probability and statistical inference within an unambiguously accepted theory. With hindsight, we may classify several conceptions of probability, all of which mirror huge differences in the philosophy of probability. We will present a few of the different classification schemes of the various meanings of

probability in order to illustrate that probability really is a multi-faceted concept. Barnett (1982) sees the following distinction in probability viewpoints (italics by the present author):

- Classical*: Symmetry considerations; ‘equally likely outcomes’.
- Frequency* (or frequentist): Empirical; relative frequencies in ‘repeatable’ situations.
- Logical*: Objective; intrinsic ‘degree-of-belief’ as a logical measure of implication.
- Subjective*: Personalistic; individual assessment of ‘rational’ or ‘coherent’ behaviour (p. 65).

Hacking (1975) traces two sides of probability, an epistemic and a statistical side: while the epistemic side relates probability to a degree of belief, the statistical side looks for a relation of such probabilities to empirical entities, such as the relative frequencies of an event. If the probability of A (to B) is higher than ‘1 to 1’ – in modern terms, higher than 0.50 – then it should not only be favourable to bet on A but one should also win more often the bet on A than lose it. That is, right from its documented origins (the poem de Vetula from the 13th century, see David, 1962), the logical and the frequentist view from Barnett’s (1982) list above is pre-dominant. Thus, it is advisable to put wagers according to expected gain as “you will learn full well how great a gain or loss any of them is able to be” (Bellhouse, 2000, p. 135, cited from Batanero, Henry, & Parzys, 2005, p. 20). The other source of probability links even more directly to relative frequencies: lifetables and insurance (Graunt, 1662). Borovcnik and Kapadia (2014) write on this other root of probability that combines ideas of frequencies, risk, and probability:

Graunt continues a tradition already started by Huygens to calculate empirical frequencies and deal with them as if they were probabilities without theoretical justification. For such problems like annuities, a practical solution was needed. A more theoretical problem was to estimate the mortality of diseases. These ideas are the precursors to insurance and risk (p. 15).

Batanero, Chernoff, Engel, Lee, and Sánchez (2016) distinguish the following views or meanings of probability:

- *Intuitive views*: Just stating the role of intuitive views that may play a role for later teaching of more formal views as they are pre-existing without any prior education.
- *Classical meaning*: Based on an assumption of equiprobability that may apply in games of chance but that otherwise is justified by the disputed Principle of Insufficient Reason.
- *Frequentist meaning*: Based on a “limit” of relative frequencies in a repeatable

experiment.

- *Propensity meaning*: Physical disposition, which connects between long-run frequencies and an application to single cases (as it becomes necessary in statistical inference).
- *Logical meaning*: Objective degree of believe, revised under new experience.
- *Subjective meaning*: Subjective degree of believe, revised under experience.
- *Axiomatic view*: Mathematical theory based on axioms.

New as compared to Barnett are: a) *Intuitive views*, which may assume an archetypal character according to Batanero and Borovcnik (2016). b) *Propensity meaning*, which reformulates probability statements as a physical property of an object or situation to produce events, which may be more pronouncedly called a tendency (Popper, 1959). And c) *Axiomatic view*, which leaves the exact interpretation of probability open (Hilbert, 1900) though it builds a framework for an initial intuitive conception of probability, and which was provided by Kolmogorov (1933/1977) for the ideas of relative frequencies and by de Finetti (1937) for the idea of probability as subjective degree of believe.

The classical view has served as a tool to provide *logical reasons for probabilities as numerical values*. Bernoulli (1713/1987) established a connection between the logical or the classical and the frequentist views on a theoretical basis by his *Golden Theorem*. He proved – after lifelong attempts – a restricted Law of Large Numbers (LLN) that stated a kind of convergence of an entity that can be perceived as a theoretical counterpart of empirical relative frequencies towards the underlying probability. The theorem was published only after Bernoulli's death. Bernoulli's theorem highlights a dual character of probability and relative frequencies, of the logical and the frequentist side of the concept. Later, in physics, the term complementarity has been coined to express that it is impossible to separate two aspects of a concept without severe loss of meaning. In this way, we recognise such a complementarity for probability and its empirical counterpart of frequencies (Steinbring, 1991).

Bernoulli had paved the way for a substantial interpretation of probability, namely the frequentist meaning. That implies, if one has not been able to find those cases that can be considered as equiprobable, one would still be able to apply probability, as its values would then be estimated from the relative frequencies of repeated experiments. That view paved the way for a probabilistic revolution in physics, beginning with the error theory for measurements by

Gauss (1809) and Laplace (1814/1995) and later with the development of probabilistic laws in thermodynamics (see Steinbring, 1980).

In our everyday lives, many of the occurrences of probability relate it to a personal conviction, which Good (1965) describes as psychological probability. There has not been much reference to such a subjective view of probability within the major part of the development of the concept of probability and the emergence of a probability theory. As applications of probability made it necessary to have numerical values for probability, the classical view with its symmetry considerations and other justifications such as the principle of insufficient reason came to the foreground in comparison to the logical view. Moreover, the principle of insufficient reason opened the doors to an arbitrary judgement of possibilities: Poisson is a major case for such “misuse”, especially with his examples concerning applications of probability in law. Good (1983) presents the following classification of conceptions of probability.

Since this [...] is concerned mainly with subjective and physical probability, it would be inappropriate to discuss other kinds in great detail. Perhaps a mere list of various kinds will be of interest:

- 1) *Degree of belief* (intensity of conviction), belonging to a highly self-contradictory body of beliefs. (This hardly deserves to be called a probability.)
- 2) *Subjective probability* (personal probability, intuitive probability, credence). Here some degree of consistency is required in the body of beliefs.
- 3) *Multisubjective probability* (multicredence). The name here is self-explanatory.
- 4) *Credibility* (logical probability, impersonal, objective, or legitimate intensity of conviction).
- 5) *Physical probability* (material probability, chance, propensity; this last name was suggested by K. R. Popper).
- 6) *Tautological probability*. [...] it is customary to talk about ideal propositions known as “simple statistical hypotheses.” If, for each possible result, E , of an experiment, $P(E | H)$ is equal to a number that is specified as part of the definition of H , then the probability $P(E | H)$ may be called a ‘tautological probability’, and H is a ‘simple statistical hypothesis’ (p.70).

Good introduces this classification with a remark on subjective and physical probability, which are also the principal views in the controversy of the foundations of probability, on which we base our considerations in this paper. His *degree of belief* has a great overlap with intuitive views of the Batanero et al. (2016) scheme. Good (1983) further discerns between subjective and multisubjective probability with the difference that the latter corresponds to a joint group

preference system and not to a single person. Credibility differs from subjective probability by the way that the personal judgement is based solely on logical reasons (and not on personal preferences) though probability is attached to *judgements of “persons” about reality* rather than it were perceived as *property of reality*. Good (1983, p. 71) continues with considerations about statistical hypotheses, which he introduces as tautological probabilities and concludes with a dogmatic decision that it is enough to focus on subjective probability, which does not preclude any other kind of information about the underlying probabilities, such a relative frequencies:

A full discussion of the relationships between the various kinds of probability would take us too far afield. I shall merely repeat *dogmatically* my opinion that although there are at least five different kinds of probability we can get along with just one kind, namely, subjective probability. This opinion is analogous to the one that we can know the world only through our sensations, an opinion that does not necessarily make us solipsists, nor does it prevent us from talking about the outside world. Likewise, the subjectivist can be quite happy talking about physical probability, although he can measure it only with the help of subjective probability (p.71).

Borovcnik and Kapadia (2014) use the terminology of APT, FQT, and SJT from Çınlar (2011). APT stands for *a-priori theory*, which refers to the classical meaning (in the list of Batanero et al., 2016), also denoted as Laplace’s equiprobability. A priori means that the probabilities are fixed right from the beginning (*a priori*), a terminology that might be confused with prior probabilities in Bayes’ theorem (though there is some overlap as one needs probability values a priori to use the theorem). Yet, it reflects the character of this probability much better than other names.

FQT stands for *frequentist probability* (no 5 of Good’s list, which can relate to either the long-run behaviour of frequencies or to the propensity meaning) and marks a *posterior* approach towards probability as it can only estimate values for probabilities after data is available. FQT connects the theoretical probability to something like relative frequencies and makes probabilistic statements open to an empirical check – by methods of statistical inference – against data from the real world. Finally, SJT signifies *subjectivist views* on probability; it connects logical reasons (credibility, no. 4 from the list of Good, 1983, above) and subjective probability (no 2 from Good, as probability that relates to a preference system). Borovcnik and Kapadia lay a focus on subjectivist probability as in decisive places in inverse inference (reasoning from data back to hypotheses) approaches of logical probabilities have failed.

To all these classification schemes it adds to the difficulty of clarifying the concept of probability that there is a group of researchers that want to settle the concept *within* a theory of probability, and other researchers, that advocate that the concept of probability cannot be understood if it is investigated separately from statistical inference. The argument of the latter group is supported by the tight interconnection between probability and statistical inference. These are complementary, since it is not possible to separate the concepts without severely changing their character. We will go into the details of an analytic clarification of the concept of probability as has been undertaken by Hacking (1965) and Stegmüller (1973).

The philosophical debate in the 1930s has revived the conflict between material objective interpretations of probability, mainly related to an interpretation as something connected to relative frequencies and as a degree of belief that has significant closeness to something connected to personal probabilities. Both interpretations have emerged in an axiomatic theory (Kolmogorov, 1933/1977, for the relative-frequencies conception, and de Finetti, 1937, for the degree-of-belief conception), which serves as a justification of the respective interpretation. As long as we remain within the frame of probability theory, there was a strong preference for the objectivist variant – especially faced with the need for an *empirical* view in physics where probability had taken over a lead role since the 1870s with the progress in thermodynamics by Boltzmann (see Steinbring, 1980).

However, the Bayesian controversy in the foundations looked beyond probability theory, a controversy that was fuelled by the way, either school of probability tried to develop a theory of statistical inference. If probability is something like a relative frequency in experiments, then it is the role of statistical inference to fix the conditions, under which probability can be measured by relative frequencies. Consistently, the clarification of these conditions underlying the measurement process, and the precision that can ultimately be achieved influences the significance of probability. Thus, statistical inference shapes the concept and the scope of probability. We go into the details of the controversy and clarify how the integration of statistical inference introduces a shift in rationality, objectivity, and scientific character of the methods.

Just to mention already here the resulting dilemma of the present analysis: Either stay within a purely objectivist – frequentist interpretation of probability and accept huge gaps in the theory of statistical inference, or radically change to a conception of probability that is more general and has genuine subjectivist traits but attains a sound theory of statistical inference.

Though we finally aim at clarifying the relation between frequentist and subjectivist aspects of probability with relation to statistical inference, we discuss the three main conceptions of probability, which we denoted by APT, FQT, and SJT. The role of the propensity meaning of probability and the axiomatic probability will become clear from the exposition of the ideas.

2.3 Measurement of probability and related problems

The various views attached to probability lead to completely different situations and problems when one has to obtain values for the probability in a specific situation. Below, we will speak of the measurement of probability even though measurement might provoke too tight associations to physics. According to the different conceptions of probability, the measurement of a probability in a case of application differs.

- *For APT*: We have to identify the equally “likely” cases and then to count the cases that are favourable for the event in question. The ratio of favourable to possible cases provides the probability requested; it measures probability without measuring error.
- *For FQT*: A related experiment has to be repeatable under the same conditions and the relative frequency of the event of interest provides a measurement of its probability. As the relative frequencies fluctuate from series to series, we face a measurement error of unknown degree.
- *For SJT*: If any of the fore-mentioned information becomes available, adherents of SJT would use it of course. Yet, the approach applies also to situations where such information of possibilities or on frequencies is not available or even precluded. For these situations, SJT becomes vital and could be applied in different ways. Either an engineering judgement (qualitative knowledge about similar devices, e.g.) provides numbers for the probability, or a so-called introspection will deliver a personal probability. To assist the introspection, the paradigm of a bet has been elaborated by the schools around SJT. By the odds (as measured by the stakes), which a person is willing to accept, the personal probability is indirectly elicited.

As all the measurements are unreliable, there are several approaches to avoid the measurement of probabilities. The assumption of equal probabilities for all outcomes would circumvent the necessity to measure it. A stochastic model could be used – a binomial or a

normal distribution; arguments could be used for specific values of the parameters of such a model. Alternatively, the parameters could be estimated from a “random sample” of this distribution. For SJT, but not only for this approach, robustness considerations (Berger, 1985) would provide somehow a worst and a best scenario, or a probability distribution between extreme values that are assumed to cover the range of plausible models. This would leave a pure SJT approach behind but would still have a subjectivist connotation.

Benefits of the various approaches with respect to the measurement of the probabilities.

Each of the different approaches is more appropriate in different situations and has its relative advantages. For APT, one gets the probabilities of ALL events and thus a complete mathematical model of the whole situation. Different persons might initially disagree as to what counts as the single outcomes but after some theoretical inspection and some additional measurements via relative frequencies (which shows how important it is that APT might directly connect to FQT), the measurement of the probabilities get an objective character.

For FQT, the measurement of probabilities for a specific variable leads to the empirical cumulative distribution function or an empirical density function. Although a little more indirect, the result of the FQT measurement provides approximations for all relevant events. It is worthy to note that the measurement would fluctuate from one measurement to another but – by central theorems – one could derive upper boundaries for this fluctuation by a confidence interval and thus estimate the precision of the measurement. More importantly, the measurement conditions should be undisputed, i.e., different persons would come to an agreement what these conditions should be and thus, as the measurement procedures become coordinated between various subjects working on that, it gets an objective meaning.

For SJT, the result of measurements is ideally not a number as measure of a probability for a statement (the SJT equivalent of event), but a personalist probability distribution around this probability. Of course, an empirical correction by equally likely cases or by relative frequencies would eliminate initial differences in the subjective measurement of the probability. To integrate the consideration of relative frequencies of future experiments – if they can be performed – the formalism of Bayes’ formula is applied (see Batanero & Borovcnik, 2016; for continuous problems see Barnett, 1982). From central theorems, we can expect a fast convergence to the “true” probability distribution (for a simple simulation study, see Borovcnik,

Fejes-Tóth, Jánvári, & Vancsó, 2020). However, unlike the FQT approach, a measure of the precision cannot be given directly. However, from the perspective of a scientific point of view, SJT probabilities cause serious defence movement as we may see from Fine (1973):

Subjective probability holds the best position with respect to the value of probability conclusions, [...] Unfortunately, the measurement problem in subjective probability is sizeable and conceivably insurmountable [...] The conflict between human capabilities and the norms of subjective probability often makes the measurement of subjective probability very difficult (p. 240).

Yet, the SJT approach is applicable also when no FQT information is available and even when no further direct experiments can be performed to apply Bayes' theorem and to coordinate the various subjectivist probabilities from different persons. This situation occurs especially in engineering applications such as reliability theory. One may think of no failures of a device already observed in more than the usual lifetime of it, which is the usual case for high-quality devices, so that no relative frequencies would become available even if many devices were tested. One would not be able to wait for years to get the necessary data so that one would have to decide – from the reliability engineer's perspective – on the personalist probabilities available whether to use the device for a mission (e.g., in space) or not.

Usually, the stress upon the units is enormously intensified and from such an accelerated lifetime experiment, engineers would find an estimation for the lifetime under normal conditions (Collins, Freels, Huzurbazar, Warr, & Weaver, 2013). A similar use of probabilities as a scenario figure can be observed when assessing the reliability of technical networks that are built from several devices, which are considered to be connected in parallel or in series with respect to the probability of failure. While the probability for failure of the single devices originates from a qualitative statement, the final reliability (the complement of the failure probability) gives some orientation whether one should seek more redundancy (extra units in parallel to already used ones) to improve the reliability. The final decision would be between cost of increasing redundancy and the actual increase in the resulting reliability (a probability) of the network. Borovcnik (2006) provides a simple example of a qualitative use of probability. It would not make the models better if one claimed that the probability distribution that describes the lifetime of the devices or the reliability of the network as a formal entity had an FQT meaning and that this model could be checked by a statistical test. The situation simply is void of such a test.

Problems of the measuring of probability.

Each of the different approaches also has some problems in the process of measuring probability. For APT, one has to note that this establishes only a special case; furthermore, often it is very difficult to identify the equally “likely” cases and to argue that the cases under scrutiny are equally likely. As a tool for this process, symmetry considerations are helpful. The more severe problem lies in the fact that the concept of probability remains inconsistent. There are probabilities, which do not relate to approximations of equally likely cases, not only in statistical inference (statistical tests, confidence intervals) so that the character of probability remains obscure and a proper measurement is missing.

A far more critical issue is that there can be no probabilities defined for series of events (which might be attached to relative frequencies in repeated experiments) based on equal likelihood. If in an infinite series of experiments under the same conditions under stochastic independence the probability of the event A in the i -th trial (denoted as A_i) is p , then for n experiments, we get $P(A_1 \wedge A_2 \wedge \dots \wedge A_n) = p^n$. With n increasing beyond any threshold, the resulting probability for the infinite series converges to zero! This is a serious problem insofar as such probabilities on series of outcomes are needed, for example, in physics, especially in thermodynamics where it has become substantial for theory building to formulate physical laws in the form of probabilistic laws (Steinbring, 1980).

For FQT, it is necessary to develop the concept of series of experiments that are repeatable under the same conditions plus a sort of an insensitivity-to-place-selection axiom (von Mises, 1919), which corresponds to the later formulated concept of independence (Kolmogorov, 1933/1977). For the applications, repeated experiments are an ideal concept; it is by no means clear whether and when such assumptions are met so that the resulting relative frequencies can be perceived as a feasible measurement of an unknown probability. As there is no mathematical formalism to describe the pattern of the development of the relative frequencies, no limit of relative frequencies can be read off finite series of experiments.

In addition, there are genuine one-off situations, which means situations that have no connection whatsoever with repeating it under the same conditions, not even as an idealistic experiment (Borovcnik, 2015). In such a case, a measurement of the unknown probability from the frequencies of repeated experiments is lacking. Such single-case situations occur regularly in decision making under uncertainty but also in statistical inference where the quality indices

of the methods (such as the probability of type-I and type-II errors, or the coverage probability) refer to irrelevant features of the situation: What would happen if the inference were made under the same conditions, “independently” of each other? There is no *unique* model that has been applied to the specific situation, because the next time a different model would reflect the examined situation much better.

Moreover, what should mean that the inferential situation would be repeated independently? There is no way to argue for this except just to request this assumption to be fulfilled. Even more severe, the probability of a rejection (or non-rejection) to be a false decision in the presented statistical test can neither be formulated nor calculated without extending the model to integrate quantities into the analysis, which are genuinely subjectivist, i.e., which are of SJT character and thus lack any connection to empirical frequencies in experiments.

For SJT, the core problem is the question of the scientific status of such a quantity that has been measured by introspection or by qualitative arguments reflecting implicit engineering knowledge. As it has a personal connotation, it cannot directly be acknowledged as being objective – supposed that measuring it in terms of APT and FQT is missing. It has to be noted that – even though such a probability value has a personal character and two persons can end up with different probabilities, such a measurement (and its result, a measurement value of the probability in question) is *by no means arbitrary* as it is often perceived (see no 1 and 2 in the classification of Good, 1983). There have to be quite a lot of crosschecks for consistency in the measuring process (see the foundation of SJT). The greatest difficulty in such a perception of probability as SJT and its measurements do lie in the applications of physics where we are confronted with an entity we call real world to which there are “objective” physical connotations, which would be in conflict with a subjectivist probability.

3. Foundations – justification of probability

In the analytic theory of science, a meaning of a concept is justified by developing a theory around this concept. The richer this theory is and the more phenomena of reality are reflected in this theory by derived concepts and theorems, the better is the justification for the initial meaning of the concept (Stegmüller, 1973). Such a theory provides solutions for the phenomena; the interface between this theory and its applications is characterised by this basic meaning, which also allows translating from the situation in reality to the model at theory level.

There are several criteria to judge the quality of a foundation: Does the derived theory allow dealing with as many phenomena and problems as possible? Is the development of the theory self-contained without rationality gaps? The best way to answer the second question is to formulate the basic rules for the concept in the form of mathematical axioms and then use only logic and mathematical relations to build the theory.

In Section 3.1, we see that APT only allows the description of restricted phenomena related to uncertainty (as quantified by probability) although there are obviously other situations where a probability would apply but for which the derived concept of probability cannot be used without breaking consistency. For FQT, Section 3.2, the difficulty of the indirect justification by Kolmogorov's axiomatic theory is not as easily seen as the theory works quite well with some drawbacks on defining and understanding conditional probability. The problems arise only when one tries to integrate statistical inference into the theory of probability. For SJT, as we discuss in Section 3.3, statistical inference is just a twist on conditional probability so that it is fully integrated into probability theory, yet, at the price of a subjectivist concept, which seems unacceptable from the perspective of an objective paradigm of physics.

The critique against statistical inference based on an FQT interpretation – which is full of consistency and rationality gaps – formed the core of the Bayesian controversy on the foundations (Good, 1983; Hacking, 1965; Stegmüller, 1973). In this controversy, the scientists even accepted Popper's view that statistical hypotheses can only be tested on an ongoing basis, so that they become prone to rejection, which would increase the *overall* proportion of correct hypotheses by this method of corroboration rather than seeking statements about *specific* hypotheses (Popper, 1962/1935).

In applications, the attitude is pragmatic. Regardless of the connotation of probability, it is used as a 'neutral' mathematical concept and information is filled into the models, as it becomes available (Berger, 1985). That means, even Bayesians would use FQT information if it is available (see Good, 1983); yet they would process the information by Bayesian methods. However, also non-Bayesians would use SJT information if this were the only way to tackle the problem. If there is enough data, then it would overwrite SJT information; if there is not enough data, then an analysis based on SJT is better than do nothing. For evaluating the result, more or less robustness considerations would replace an FQT or SJT interpretation in either case.

3.1 Foundation of probability using the a-priori-theory approach

The basic assumption in APT is that all (finitely many) outcomes of a random experiment (a random situation) are equally likely. This provides probabilities to any event and allows for quite a rich probability theory that culminates in the normal distribution, which occurs as an approximation to the binomial distribution. The concept of independence is circumvented by the equally-likely assumption on the (product) space of combined experiments (which also stands for the repetition of the same experiment). Consequently, it is also possible to define the concept of a random sample within the APT approach.

The binomial distribution also models a concept of theoretical absolute frequencies, so that the way is paved to connect to the frequentist conception of probability, especially as the relative frequencies (theoretical absolute frequencies divided by the number of trials) “converge” in some complex sense to the underlying probability (Bernoulli’s Law of Large Numbers). That connection opens the APT approach for an empirical test by practical relative frequencies that can be observed. It is also possible to formulate the usual methods for statistical inference in the restricted situations of repeatedly drawing (with or without replacement) out of an urn. With some artificial twists, one would also be able to extend inferential methods (via approximation of latent binomial distributions) to the normal distribution.

In this way, we get a possibility to check the equally-likelihood assumption right at the beginning for specific cases of applications by an empirical counterpart to the theoretical probability by a statistical test. There are, however, several restrictions about the APT approach:

- *We are left without probabilities* if we cannot provide equally likely cases or prove that the possible cases that are identified in a specific situation are equally likely.
- *Symmetry arguments* may support an assumption of equal likelihood for all possible cases. If they are not applicable, which is the situation when we go beyond gambling to real-life applications, then an awkward argument such as the Principle of Insufficient Reason (Laplace, 1812/1955) is applied to justify this equal-likelihood assumption.
- *For infinite series of equally-likely experiments*, there is no way to define a proper probability. This is a major disadvantage for physics, where they need to have probabilities for such series. Furthermore, since 1909, a strong Law of Large Numbers (LLN) has been proven by means of Borel-Cantelli Lemma and it states that the “probability” of all series that do not converge to the “true” probability is zero. A further

inconsistency arises in statistical tests (though this did not occur before the 1930s with the advent of statistical tests by Neyman and Pearson; see Neyman & Pearson, 1928, Neyman, 1937). The probabilities of type-I and type-II errors attain a frequentist connotation only by an artificial reproduction of the test situation but definitely are not connectable to an APT conception of probability.

- *A basic philosophic counterargument* against the APT definition of probability is, of course, that the concept of probability is defined by equal likelihood of elementary outcomes, which is nothing but a different wording for probability. Thus, the definition is circular as it uses the very concept to define the concept.

Yet, another inconsistency was linked to the “probability” that occurred in the Borel-Cantelli lemma. To solve this problem there was a need to find a better foundation for the concept of probability, which has already been recognised earlier by David Hilbert in 1900 who listed the axiomatisation of probability in physics among the most urgent mathematics problems (Hilbert, 1900).

3.2 Foundation of probability using the frequentist approach

The central idea of this approach is that probability is something like the limit of relative frequencies in repeated experiments that may be characterised as random sequences. This endeavour met the need of physics to have probability as an empirical quantity linked to relative frequencies in specific experiments. Von Mises (1919) who comes from physics took it as his life programme to find a suitable theory for probability based on the concept of relative frequencies. To achieve this goal, von Mises introduced an axiom called ‘Regellosigkeit’ that should delineate the fact that a sequence of relative frequencies cannot follow a mathematical law comparable to sequences in calculus (wherefrom it is possible to derive a limit of such a sequence if it converges). The axiom of ‘Regellosigkeit’ reflects the random behaviour of relative frequencies; since this behaviour was directly characterised in terms of empirical entities and not by the later so-called iid assumptions (independent, identically distributed), the axiom itself and the resulting theory became very complicated. Von Mises tried to repair initial inconsistencies but did not accomplish this task (Porter, 1986). Only “exotic” mathematicians devoted their energies to that project and it was not before the 1970s, that Schnorr (1971) could

find a satisfying solution. Schnorr studied algorithms for generating random numbers, so he had a vital interest as von Mises in a material definition of independence.

Meanwhile, the mathematics community already had found a viable solution to the axiomatisation by the approach of Kolmogorov (1933/1977) that indirectly justified the FQT meaning. Kolmogorov chose probability axioms that reflect idealised properties of relative frequencies and derived a mathematical theory based on these axioms that was apparently equivalent to measure theory with a finite measure normalised to 1 (the probability of the whole space). By additionally formulating the concept of independence, Kolmogorov was able to obtain the distribution of a random variable that describes the theoretical behaviour of absolute frequencies (which follows a binomial distribution). With the Bernoulli Law of Large Numbers (LLN), which continued to be valid in Kolmogorov's theory, the connection between relative frequencies (as random variable describing relative frequencies in reality) and the underlying probability (which is usually unknown) was established: the random variable of relative frequencies converges 'in probability' to the basic probability (Kolmogorov, 1933/1977, pp. 50). Because of this LLN, probability in Kolmogorov's theory was open to the frequentist interpretation of probability. This was *the* argument that Kolmogorov's approach was judged as useful for physics and regarded as the solution to Hilbert's demand of 1900.

The success of Kolmogorov was not only grounded on the justification, which his theory provided for the FQT interpretation of probability. At the same time, Kolmogorov's probability clarified directly the status of the "probability" in the lemma of Borel and Cantelli and Kolmogorov's probability was directly applicable to (infinite) series of events (random variables). Such series were not only useful for physics and for further limit theorems but also for developing the theory of stochastic processes, which deal with series of random variables. The theory of stochastic processes is that achievement why Kolmogorov is named among the three top mathematicians of the twentieth century.

The basic feature that makes a distinction between general measures and probability is the independence concept (see Steinbring, 1980, p. 251). This concept is not only an elegant solution to avoid the von Mises' Regellosigkeit axiom; it represents a hinge for the application of probability. However, the basic axioms do not cover the concept of independence, so that it has to be introduced by an additional definition as a product rule for events: Two events are independent if the joint event has a probability that equals the product of the single probabilities.

Yet, if the meaning of independence is determined by a definition solely based on probability, how can probability be explicated by the axioms plus a regress to independence? This is a vicious circle. Kolmogorov (1956) recognises this stating

[...] one of the most important problems in the philosophy of the natural sciences is [...] to make precise the premises which would make it possible to regard any given real events as independent (p. 9).

Attempts to avoid this circularity and explain independence by a lack of causal influence are a source for misconceptions as the two concepts are on distinct levels with no common ground. A further possibility to circumvent any explanation of this sophisticated concept is to separate those situations where independence is “reasonable” from those situations where it is “not reasonable” by a statistical test (a chi-squared test, or a Kolmogorov-Smirnov test).

We will see later that this solution needs more than just extending probability theory to statistical tests, because a satisfactory theory of statistical tests is not already contained within probability theory. It needs extra assumptions and concepts, which lack a consistent theoretical framework though they can be formulated with probability concepts. There are substantial counter arguments against such a test theory so that it remains not easy to check whether independence applies or not in a specific problem (see Section 4.3).

That difficulty has a serious consequence also on any statistical test of a specific probability (or of a probability distribution) as such tests are usually based on *random* samples, which amounts to an independent repetition of the same basic random experiment. If in practice the independence assumption is not guaranteed, then the concept of relative frequencies is void of any meaning so that a statistical test based on relative frequencies lacks any justification. In the practice of statistics, it has become a ritual to claim that the data on which a statistical test is based, originate from a random sample, which includes the independence of the process of how the single data has been generated. This claim is not something that can really be investigated by research methods. Furthermore, it is not something that is open for a check afterwards by independent researchers to control the statistical investigation of others. The continuous lament about the replication crisis reflects the problem just stated (Gorard & White, 2017). Replications of a study fail to reproduce the direction of the results of the original study, i.e., they fail to confirm the significance and the direction of the effects found.

These difficulties with defining the concept of independence are one fundamental reason

for adherents of an SJT conception of probability to replace it by the concept of exchangeability, which has been introduced by de Finetti (1937); see also Barnett (1982, p. 78). Exchangeability is a form of symmetry that can easily be checked by its influence on the betting behaviour: if the sequence of statements is of no influence on the elicitation of odds, then exchangeability applies. That is, to explain exchangeability needs no reference to vague causal arguments. In addition, the central theorems (such as the LLN) of probability still hold.

In summary, although Kolmogorov's axiomatisation has established a satisfactory basis for the FQT interpretation of probability, the question of when in practice we really deal with relative frequencies in the sense of Kolmogorov theory remains unsolved. For any application, independence has no more than the status of a background hypothesis, which cannot properly be examined. Kolmogorov's axiomatic theory can also be interpreted to define a probability void of any meaning and thus it is open to diverse meanings as long as there is a corresponding interpretation of the basic axioms in terms of another perception of probability. For example, the APT probability is just a special case of Kolmogorov's probability. A little bit more complicated will be the view that Kolmogorov's theory may be perceived as a common mathematical description of probability that mediates between so different perceptions as FQT and SJT. This serves only as a tool to simplify the language between different perceptions of probability, but – beware – the “words” in the common language have a distinct meaning.

3.3 Foundation of probability using the subjectivist approach

The basic entity, on which probability is applied in the SJT approach, is a statement (that includes uncertainty) rather than an event. Preferences (that may originate from a person) order these statements, which is a different way to evaluate the entity than by relative frequencies or theoretical probabilities. The actual preferences of a person can be measured by a process of elicitation, especially in the form of equivalent bets. If specific axioms are requested about a preference system (such as transitivity, consistence, or coherence, see Barnett, 1982, pp. 86), then, this preferences allow for a numerical representation (de Finetti's representation theorem, 1937). This representation is a finitely additive probability function, which is a personal variant of probability function as is depends on the preference system that can differ between various persons. It is helpful to formulate the assumptions in connection to odds in bets because in this way it becomes easier to measure the probabilities.

The Bayesian probability has slightly different properties as seen from a structural point of view. As bets can be directed only on finitely many possibilities, the representation theorem provides only a finitely additive measure for the probability function instead of the countable (complete) additivity of Kolmogorov's probability. De Finetti (1937), as an extreme advocate of the subjectivist position would have insisted on this difference. Yet, as Good (1983, p. 176) states, "I am also prepared to use an axiom of complete additivity as a mathematical convenience but do not regard it as essential to any application. It is purely metaphysical."

Consistently, from a mathematical perspective, one could smooth these differences in order to find an equivalent formulation of the Bayesian probability theory that coincides with the Kolmogorov probability theory. That means, one could perceive Kolmogorov's probability theory also in a subjectivist sense, i.e., as an overall structure arching between the two so distinct axiom systems (one about the probability of events, the other about basic properties of preferences related to statements). The goal of such an endeavour is to make it easier to compare the two probability theories (Borovcnik & Kapadia, 2014).

To this end, one would have to find a compromise on the following issues: SJTs would have to agree that numerical values for probabilities make sense (true subjectivists such as Good, 1983, prefer to have intervals for probabilities instead). The difference between independence and exchangeability should be 'ignored' for the sake of the comparison (de Finetti, 1937, would severely object that; Berger, 1985, would be less stringent on this issue). The difference between finite additivity and countable additivity should be set aside to allow for the comparison (de Finetti, 1937, with his focus on bets would strongly oppose such a levelling). Moreover, statistical inference would differ between SJTs and FQTs. Yet, with such simplifications for the *probability* part (not the inferential statistics part), we would be able to use the Kolmogorov theory as a common syntax to describe probability, yet – even though the probability language would then be the same between the so hugely distinct approaches, the interpretation of the same "words" and theorems would make the difference.

Insofar as there is APT information about equally likely cases or there is FQT information about a problem, an SJT adherent would use such information for the analysis and there would be no discussion about the objectivity of the procedure (see also Good, 1983, in his comments on the various probability conceptions). If such an "objective" information were lacking, an SJT would still be able to apply probability using qualitative information. The great

discrepancies between the approaches would occur when it comes to develop and justify the methods for evaluating the used “model” when it is not open to an FQT interpretation and consistently, FQT information is missing.

3.4 Summarising the probability positions

To summarise the discussion, if we consider only probability theory, we come to the following conclusions:

- For APT, we can establish quite a lot of probability theory and even develop methods for statistical inference for nearly all situations that are highly relevant for statistics education – at high school and in many fields of applications at the university. This applies even when we acknowledge that an APT conception is restricting the meaning of probability drastically and that there are conceptual breaks with probability as discussed in the previous sections. Instead of the usual terminology of classical or Laplacean probability, we use the term APT (the a-priori theory) to reflect that it attributes probability values to events *prior* to data.
- For FQT, we obtain an objective property of events that is free of personal judgement, which is at first sight “directly and easily linked” to relative frequencies; we can also derive statistical tests and – to some extent – interpret the quality indices of such tests from a frequentist viewpoint. FQT is a *posterior* approach as it allows to attribute probabilities only based on data by estimating the otherwise unknown probabilities.
- For SJT, the personal or Bayesian conception of probability, we also obtain a mathematical theory that is equally well justified as the Kolmogorov theory of probability. Yet, the concept is bound to personal judgement (hence also a personal probability), which is often interpreted as arbitrary and subjective, since the approach allocates probability values via preferences. Of course, besides qualitative information, also FQT information can be used if available, so that the approach becomes more widely applicable. Robustness considerations when possible (Berger, 1985) are a method to improve the validity of probability values for the investigated statements.

4. Frequentist and Bayesian inference

In the investigation of this section, we refer to frequentist methods of inference associated with an FQT interpretation of probability and an objectivistic framework for probability, whereas in Bayesian inference, we subsume methods of inference connected to an SJT interpretation and a subjectivist framework for probability.

4.1 Statistical and deterministic hypotheses

If we compare these two main ways to conceive the concept of probability, it becomes clear that there is some substantial interest to favour an FQT interpretation as it would pave the way for an empirical connection for probabilistic hypotheses. That means that theories and statements (hypotheses) would have an empirical meaning and would be open for something like a statistical test that would provide an objective check of such hypotheses against empirical facts. In this way, science should guarantee that statements would not depend on personal judgement and “taste” but would rather be open for such a test that is undisputed between different persons, i.e., free of subjective elements. If an explication of the concept of probability connected to an FQT meaning is the goal of a scientific project in the foundations of probability, then this task has to include proper methods for an empirical examination of probabilistic statements (statistical hypotheses). Such rational methods will influence the conception of probability and its objective character (i.e., independent of personal judgement).

We start with some comments on hypotheses and hypotheses testing. We have to discern deterministic hypotheses, which are generally logical all-statements, and statistical hypotheses, which include a probability statement. A typical deterministic hypothesis is “All swans are white”. Such a statement is falsified if we find one counter example, i.e., one swan that is black. A typical probabilistic hypothesis is “The probability of Head with a specific coin equals θ ”.

The term “statistical data” combines relevant knowledge from observation (its empirical component) and background knowledge (its theoretical component). This background knowledge provides the specification of the *class* of hypotheses (expressed by probability distributions), which form the basis of the empirical check (the test); for example, the family of normal distributions, if a specific statistical hypothesis (a specified normal distribution) undergoes such a test. We denote such knowledge as background *hypothesis* as there is the possibility to refer to a different choice. Independence of single data is a further background

hypothesis, which is crucial for any test as it is a requirement of the notion of a random sample, which combines single data from several experiments. The following example illustrates the function of statistical hypotheses and the relevance of background hypotheses:

We start with the assumption – the hypothesis h – that $P(\text{Heads}) = \theta$. It is not possible to examine this hypothesis in isolation, as it would seem at first sight. With a test, we usually request additionally that the number of Heads in a series of n trials follows a binomial distribution $B(n, \theta)$; this type of distribution indirectly reflects the independence of the single trials as random variables. Within the test procedure that has to be developed, this background hypothesis will not be disputed.

In a comparison between deterministic and statistical hypotheses, the following striking differences have to be noted: Deterministic hypotheses (also called deterministic laws) are relative to acknowledged data definitely falsifiable but not verifiable. Statistical hypotheses are neither verifiable nor falsifiable. The following example illustrates the situation.

The statistical hypothesis h . “Probability of Head in throwing a specific dice equals $1/6$ ”. It may seem convincing to reject this hypothesis if one gets 12 Sixes in 20 throws of this dice; yet, there is no way to exclude logically that such an event can be observed in the case that the hypothesis h really applies.

Concerning an error of the decision, for a deterministic hypothesis only a type-II error has to be considered (i.e., a hypothesis that is false is erroneously not rejected) as such a hypothesis can definitely be rejected if one finds a counter example. Different to that, for a statistical hypothesis, two types of errors have to be taken into account: the type-II error, and – additionally – the type-I error, which is the situation where a hypothesis that is true (that applies) is erroneously rejected (Borovcnik, 2015).

Even more stringent is the fact that the judgement of deterministic hypotheses can be done in isolation, i.e., a single hypothesis can be judged (as if it is judged as false, it is definitely false) whereas it is not possible to judge a statistical hypothesis in isolation as if it is rejected, it still can be “true”. If we base our decision solely on an isolated hypothesis, we cannot simply reject it. We first have to find better reasons to justify such a rejection. Better reasons are better alternatives – alternative hypotheses that have a higher support when the data is considered. Thus, the “degree” to which it still can be true reflects the strength of alternative hypotheses. Consistently, a statistical hypothesis can never be examined in isolation (Batanero, 2000).

As a due consequence of the foregoing, the procedure of checking a deterministic hypothesis provides rules for when to reject such a hypothesis, while a procedure for checking a statistical hypothesis can only provide statements about the (relative) support of hypotheses. Only reject one hypothesis in favour of another hypothesis if the other one has much higher support! Suitable support functions would be expressed by probabilities but they are excluded from an objectivistic framework. In an objectivist framework there is no probability of a hypothesis as probabilities would need an empirical connection in the form of relative frequencies in a specific experiment, which is not possible, neither for deterministic nor for probabilistic hypotheses. This restriction does not apply for a subjectivist framework as probability therein is not necessarily bound to an empirically testable element but rather is a degree of credibility, so that a probability structure for the support of hypotheses would clearly be an option for the subjectivist position.

For the objectivist position there remains the so-called likelihood (support) of statistical hypotheses relative to empirical data and further, *indirect* approaches that circumvent the concept of support of hypotheses. The Neyman-Pearson test (Neyman, 1937; Neyman & Pearson, 1967/1928) translates better support of an alternative for all *potential* data into their probabilities for type-I and type-II errors; the Fisher (1925, 1935) significance test replaces support by fiducial probabilities. All of these approaches use direct or indirect notions that lack the structure of probability so that it becomes more difficult to interpret *how* they support statistical hypotheses.

4.2 Direct inference and inverse inference

It makes sense to identify two distinct statistical tasks, the direct conclusion from a specific probability distribution (a statistical hypothesis) towards future data, and the inverse conclusion from observed data to potential statistical hypotheses.

Direct or single-case inference

What can we conclude from a statistical hypothesis (a probability distribution) for a “future” event? What does a probability statement mean for the reality? In this situation, “direct inference”, the so-called single-case rule determines the meaning of such probability statements: If the probability model (the objectivist probability) specifies $P(E)$ for an event E , then it is

possible to interpret this value as support for the statement

$e :=$ “E occurs in a concrete trial”.

The support for e is exclusively determined by the information $P(E)$. There are several attempts to justify the single-case rule (see also Borovcnik, 1984, pp. 194-199):

- *Naïve objectivists-frequentists* motivate the single-case rule by “long-run” considerations: they request “axiomatically” that “optimal” behaviour in the long-run is also optimal for the single case, although there is no logical justification for this request. Furthermore, Borovcnik (2015) illustrates the differences in decisions that are adapted to the single case and those that should optimise long-term behaviour; the optimal strategies differ completely, which highlights that it is not easy to mediate between single-case and long-run.
- *On a theoretical basis*, one would justify the rule by a reference to the likelihood support theory (Stegmüller, 1973) by which it is possible to derive the single-case rule within the logic of support. This approach, however, causes a circular justification: On the one side, Stegmüller’s likelihood rule attains its authorisation only because it provides a derivation of the single-case rule; on the other side, it is only possible to justify the single-case rule by a reference to the generalised likelihood rule.
- *The propensity meaning of probability* suggested by Popper (1959) does not provide a solution. This propensity rule resumes the applicability of a probability statement to a single case; yet again, it has no additional justification so that in a way it attains the character of an unquestionable principle rather than an elsewhere grounded property of probability.
- *A subjectivist justification* is much easier as this would be referred back to accepted odds in bets and if no further information becomes available for the single case, then the odds are completely determined by the probability, so that the probability statement in general and the statement for the single case coincide.

Yet, the direct inference really is not the core of the controversy between objectivists and subjectivists. It is essential to note that one has to separate the problem of the justification of the single-case rule from the problem of the correct application of this rule, as the probability

statement could be ambiguous. This causes difficulties regardless whether one adheres to an objectivist or a subjectivist position towards probability.

Inverse inference – Inference from data to hypotheses

How can one conclude from empirical data to probability models, which are latent behind the random process? Thomas Bayes was one of the first to formulate a problem of inverse inference; Hartley (1749) enthusiastically reports about its solution by Bayes (1763):

An ingenious Friend [Hartley refers to Bayes] has communicated to me a Solution of the inverse Problem, [...] where the Number of Trials is very great, the Deviation must be inconsiderable: Which shows that we may hope to determine the Proportions, and, by degrees, the whole Nature, of unknown Causes, by a sufficient Observation of their Effects (Hartley, 1749, cited after Hall, 2007, p. 25).

The development of the causes in the light of observations should be translated to modern language as the probability θ of “success” in the experiment after we have observed the relative frequencies in the repeated trials. Structurally, Bayes was faced with the problem of a Bernoulli chain producing the binary data with an unknown probability θ for the 1s ($1 - \theta$ for the 0s) and the need to make an inference about θ when we are faced with k 1s and $n-k$ 0s in a series of n trials. His argument about complete lack of knowledge about θ was later resumed by Laplace so that it became the famous principle of insufficient reason as an argument for using equal probabilities for the possibilities (Laplace, 1812/1955), which are here all values of θ in the interval $(0, 1)$. In fact, Bayes derived – within his experimental device – a correct argument for using a uniform distribution on this interval:

Bayes gave another *objective* reasoning for his uniform prior (see also Edwards, 1978). If you do not know anything about the combined experiment (including the parameter θ), then the stakes for any number k of successes in n trials should be the same. From this assumption it is relatively easy to justify a uniform prior for the unknown parameter θ (Borovcnik & Kapadia, 2014, p. 19).

In Bayes’ setting, θ is a parameter that has a uniform distribution; the random experiment provides a binomial distribution for the number of 1s. The inverse inference on θ relative to the data k of n is performed by a new distribution for θ (in modern terms a beta distribution) (Barnett, 1982; Good, 1983). We are only interested in the “structure” of the situation, which

can be summarised in the following way, which forms a discrete version of Bayes' formula (see a simplified case in Batanero & Borovcnik, 2016; the general case with continuous densities is in Stegmüller, 1973).

Bayes' formula: Let H_1, H_2, \dots, H_r be r exclusive and exhaustive hypotheses for a stochastic situation and E be an empirical observation. If the probabilities $P(H_i)$ were known as well as the probabilities for E under the various hypotheses, i.e., $P(E | H_i)$, then the hypotheses attain a new probability conditional to E in the following form:

$$P(H_i | E) = \frac{P(H_i) \times P(E | H_i)}{P(E)} \propto P(H_i) \times P(E | H_i).$$

Hereby, the term $P(E | H_i)$ represents the likelihood of H_i relative to the empirical data E , and $P(H_i)$ stands for the prior probability (prior information, applies before the data E becomes known). The term on the left side is also called the posterior probability of the hypothesis relative to the data E . This posterior probability attains the structure of a probability; i.e., it fulfils the Kolmogorov (1933/1977) axioms, or better, the de Finetti (1937) axioms and integrates the data into the support function.

The \propto sign in the last part of the formula line denotes a proportionality. It signifies a simplification of the formalism. As long as data E is fixed, the posterior probability $P(H_i | E)$ is proportional to the numerator and the denominator works only as a normalising constant, i.e., to normalise all probabilities for the hypotheses so that they sum up to 1.

There is obviously a great potential in this application of Bayes' formula. This formula in addition to the single-case rule and a so-called conditional conditionalisation (Borovcnik, 1984; Seidenfeld, 1979) together form the basis of a complete inductive logic. However, the price for that achievement is very high indeed, as usually there is no frequentist (objectivist) information about the hypotheses on the prior probabilities so that to accept this would mean to leave a closed objectivist position in the foundations of probability.

We mentioned that Bayes – within his experimental device – had derived a uniform distribution for the parameter θ . This argument was called Bayes' postulate and was erroneously extended to Laplace's principle of insufficient reason, which obviously represents an epistemic flaw: how can the principle of insufficient reason – overly interpreted as absolute lack of any knowledge – emerge into a uniform distribution on the possible cases, which undisputedly represents some sort of knowledge?

The Objective Bayesians (Jeffreys, 1948) tried other ways to repair Bayes' postulate to find logical justifications for it; yet, such attempts lead to irresolvable problems that may be best circumscribed by violations of the invariance principle. Of course, the prior probabilities usually are not open to an empirical control by relative frequencies in a suitable random experiment.

Overall, Bayesian inference was refuted by objectivists as a solution for the inverse inference (Venn, 1866). Objectivists request that methods for judging statistical hypotheses have to link to objectivist probabilities only (i.e., probabilities that are open to an FQT meaning). Modern subjectivists (Savage, 1962) would circumvent the justification of the uniform distribution in Bayes' postulate by grounding probability as degree of confidence. The objectivist position would discredit such attempts, however, as private and thus non-scientific hypotheses testing. Yet, it is not enough to criticise the opponent's theory based on one's own. On a neutral basis, one has to check how the various approaches of objectivists and subjectivists cope with the inverse inference. This is the target of the following section.

4.3 Objectivistic test theories

Within the objectivist school, there has emerged a strong position in favour of the likelihood tests and against the Neyman-Pearson test theory. Likelihood tests are based on the likelihood support theory (Hacking, 1965), which is thought to be constitutive for an objectivist conception of probability. To the contrary, Neyman and Pearson (1928) circumvent the support theory for hypotheses by their policy of testing hypotheses repeatedly. The Fisherian significance test (Fisher, 1925, 1935) – though intuitive – is not an acceptable solution from the perspective of foundations as it lacks an explicit consideration of alternative hypotheses as demanded by the investigations in the foundations (Hacking, 1965).

Likelihood test theory. As we have stated before, early discussions of Bayes' theorem have revealed enormous problems in finding an objective justification for the prior probabilities $P(H_i)$. Moreover, it is doubtful to attribute a probability to hypotheses in an objectivist sense, as it is not open for an empirical test in an experiment with relative frequencies (not only by Venn, 1866/1962). Thus, what was left of Bayes' theorem was the term $P(E | H_i)$, the probability of an empirical event given the hypothesis H_i , which was interpreted as support for H_i in the sense of the higher the probability for E , *the higher the support for H_i* . Such a thinking formed apparently the background of arguments such as by Arbuthnot (1712) who interpreted the fact E (80 years

with a majority of boys in birth statistics) as an argument for the hypothesis H (a divine order) as $P(E | H^c)$ is very small. The complement to the divine order was interpreted as “randomness”.

In the scientific analysis it became soon clear that the likelihood $L(H | E) = P(E | H)$ cannot be interpreted in absolute values but only compared to the likelihood of other hypotheses (Hacking, 1965, p. 59). Birnbaum (1962, p. 271) formulated the *likelihood principle* according to which nothing else apart from the likelihood function should influence the support of hypotheses. With this likelihood principle and a translation of likelihood values into the support of hypotheses, a test theory could be justified. To test the null hypothesis h_0 against the alternative h_A , the null is rejected if the likelihood ratio is larger than a pre-specified γ , i.e.,

$$\frac{L(h_A | E)}{L(h_0 | E)} \geq \gamma.$$

Neyman-Pearson (NP) test theory. This test theory still dominates statistical practice and the teaching of statistics. Neyman and Pearson (1928) pursue a mathematical strategy to find a test with optimal properties. Yet, their solution has severe rationality gaps. Neyman states that it is impossible to get knowledge (support) about the specific hypothesis under scrutiny. Rather, one has to find rules for rational decisions, which guarantee in the long run that erroneous decisions (type-I and II) do not exceed some probability levels. This is in-line with Neyman’s (1937) naïve views on a long-run interpretation of probability. Yet, it becomes obvious that one needs prior information about the frequency of h_0 and h_A in order to get a long-run frequency of erroneous decisions as the type-I and II errors are only conditional to scenarios and have no overall (unconditional) meaning.

Stegmüller (1973, p. 183) based on an example of Hacking (1965, p. 89) speaks of an intuitive plausibility of the support argument. He summarises his critique against the NP-optimal test in the following way: The likelihood test is based on intuitively plausible considerations of support of the hypotheses after the sample with the empirical data is known, while the NP test is derived by criteria, which guarantee a good behaviour of the test before the data is used. Yet, rational behaviour in the forward look can be quite irrational in the backward look (analysis after data is gathered). Stegmüller (1973) gets to the limits of the likelihood test theory via the paradox of Kerridge (p. 217); yet, he does not abandon the objectivist position:

[...] it is questionable whether the objections raised are really decisive or whether they only point to difficulties that can perhaps be overcome within the framework of subjectivist theory. [...] Perhaps the most

convincing thing that remains is the concern: [...] the necessity of a radical subjectivation of the natural sciences [...] if the [subjectivist] view of the concept of probability were to prevail. The science theorist, before he swallows this bitter pill, will look for another solution (p. 244; translation by the present author).

In conclusion, Stegmüller declares Fisher's fiducial argument as a potential way out of the crisis. This argument, however, faces the same basic difficulties as the programme of the Objective Bayesians (e.g., Jeffreys, 1948), which may be seen from Borovcnik (1984) and Good (1971). To summarise, all solutions to avoid the prior probabilities in Bayes' theorem or to find an objective justification for specific priors are doomed to irreparable inconsistencies.

5. Conclusions

To summarise the discussion: In early times, probability was related to a *comparison* of below and above 1:1 for odds; later various odds have been in use, which involved a more detailed comparison of two probabilities (for and against the outcome, e.g., in a bet, or in playing a game of chance). With Bernoulli's LLN, the focus shifted to quantifying the probability of events by a *number* between 0 and 1 as a degree of "certainty".

Anyway, absolute numbers for probability still have attached an enormous problem for the interpretation, and it has become customary, to interpret probabilities as idealised relative frequencies. Probabilities may serve as a prediction for the frequencies in a series of random experiments (or random situations that we model by probability). On the other hand, the relative frequencies in such series (which amount to a random sample) serve as an estimate for the underlying probabilities. That means that the number attached to a probability is interpreted narrowly and not even connected to a margin of error (e.g., related to the width of a confidence interval for this probability).

Yet, we use probability more often as scenario figure (Borovcnik, 2006). Such probabilities attain their strength from a comparison between scenarios and the resulting numbers for the probability in question. A good example is from reliability, where one tries to increase the reliability of a technical system by introducing redundancy. Likewise, in decision-making, the comparison of two decisions by an expected value (sum over all future states of impact \times probability of this impact) will clearly identify the better decision while it is very difficult to interpret this expected value in isolation for one decision only.

The situation with probabilities complicates in case that we deal with small probabilities. From the many empirical investigations about misconceptions of probability, we know that people tend to react and decide irrationally once small probabilities are present, so that if a huge loss relates to a small probability, they would opt for a decision to avoid that possibility. On the other hand, if there were a huge win related to a small probability, they would seek the risk of such a situation. The first occurs with the insurance of the monetary impact of accidents or similar situations, the second relates to bets and games of chance. In the insurance example, people are risk averse; in games of chance, they tend to be risk seeking (Kahneman & Tversky, 1979).

In the foundations, the controversy about the nature of probability revealed that a frequentist conception is highly prone to dispute, as a justification of an FQT meaning failed from a rational perspective when the explication of probability integrates the application of probability, namely statistical inference. However, a naïve frequentist interpretation is prevailing in statistics education, especially since the advent of computer technology, which provided generous facilities for simulating probability models to “replace” them by the frequencies of the simulated scenario. Simulation power also facilitated to implement resampling techniques as a special kind of statistical inference. Yet, it easily escapes the attention that the techniques disclose only a small part of statistical inference and even that part does not follow the rationality principles underlying statistical inference (Borovcnik, 2019). Resampling fails to integrate a discussion about alternative hypotheses and misses answers to questions about type-II errors. By the tractability of special cases – even for teaching statistical inference at lower levels (informal inference) – one tends to neglect the greater picture about statistical inference. This will remain true in the revolution of big data, as methods of statistical inference will be an essential tool to detect and remove bias from data to make reliable conclusions therefrom.

Many applications of probability, reliability and decision theory included, have to provide a wider interpretation of probability than a simple FQT meaning. The concept of probability – right since its origins – emerges from a mixture between APT, FQT, and SJT meanings, which cannot easily be separated. The shift in connotation of probability towards a biased FQT meaning decreases the scope of probability or the quality of applications. Probability is a complementary concept, which falls apart if we reduce it to one view.

References

- Arbuthnot, J. (1712). An argument for divine providence taken from the constant regularity observed in the birth of both sexes. *Philosophical Transactions of the Royal Society*, 27, 186-190.
- Barnett, V. (1982). *Comparative statistical inference*. New York: Wiley.
- Batanero, C. (2000). Controversies around the role of statistical tests in experimental research. *Mathematical Thinking and Learning*, 2(1-2), 75-97. https://doi.org/10.1207/S15327833MTL0202_4.
- Batanero, C. & Borovcnik, M. (2016). *Statistics and probability in high school*. Rotterdam: Sense Publishers. <https://doi.org/10.1007/978-94-6300-624-8>.
- Batanero, C., Chernoff, E., Engel, J., Lee, H., & Sánchez, E. (2016). *Research on teaching and learning probability. ICME-13 Topical Surveys*. Cham: Springer. https://doi.org/10.1007/978-3-319-31625-3_1.
- Batanero, C., Henry, M., & Parzysz, B. (2005). The nature of chance and probability. In A. G. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning. Mathematics Education Library, Vol. 40* (pp. 15-37). New York: Springer. https://doi.org/10.1007/0-387-24530-8_2.
- Bayes, T. (1763). An essay towards solving a problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society*, 53, 370-418.
- Bellhouse, D. R. (2000). De Vetula: A medieval manuscript containing probability calculations. *International Statistical Review*, 68(2), 123-136. <https://doi.org/10.1111/j.1751-5823.2000.tb00317.x>.
- Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. New York: Springer.
- Bernoulli, J. (1713). *Ars conjectandi*. Basel: Impensis Thurnisiorum. Originally published in 1713.
- Birnbaum, A. (1962). On the foundations of statistical inference (with discussion). *Journal of the American Statistical Association*, 57(298), 269-326.
- Borovcnik, M. (1984). *Was bedeuten statistische Aussagen*. Vienna: Hölder-Pichler-Tempsky.
- Borovcnik, M. (2006). Probabilistic and statistical thinking. In M. Bosch (Ed.), *Proceedings of the Fourth Congress of the European Society for Research in Mathematics Education* (pp. 484-506). Barcelona: European Society for Research in Mathematics Education.
- Borovcnik, (2015). Risk and decision making: The “logic” of probability. *The Mathematics Enthusiast*, 12(1,2&3), 113-139.
- Borovcnik, (2019). Informal and “informal” inference. In J. M. Contreras, M. M. Gea, M. M. López-Martín, & E. Molina-Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística*. Available from www.ugr.es/local/fqm126/civeest.html.
- Borovcnik, M., Fejes-Tóth, P., Jánvári, Z., & Vancsó, Ö. (2020). Experimente zur Einführung von Ideen und Denkweisen statistischer Inferenz im Gymnasium. *Stochastik in der Schule*, 40(1), 18-27.

- Borovcnik, M. & Kapadia, R. (2014). A historical and philosophical perspective on probability. In E. J. Chernoff, B. Sriraman (Eds. (2014). Probabilistic thinking: Presenting plural perspectives (p. 7-34). New York: Springer. https://doi.org/10.1007/978-94-007-7155-0_2.
- Çınlar, E. (2011). *Probability and stochastics*. Berlin, New York: Springer.
- Collins, D., Freels, J., Huzurbazar, A., Warr, R., & Weaver, B. (2013). Accelerated test methods for reliability prediction. *Journal of Quality Technology*, 45, 244-259. <https://doi.org/10.1080/00224065.2013.11917936>.
- David, F. N. (1962). *Games, gods and gambling*. London: Griffin.
- Edwards, A. W. F. (1978). Commentary on the arguments of Thomas Bayes. *Scandinavian Journal of Statistics*, 5, 116-118.
- Fine, T.L. (1973). *Theories of probability*. New York: Academic Press.
- Finetti, B. de (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales Institut Henri Poincaré*, 7, 1-68.
- Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.
- Fisher, R. A. (1935). *The design of experiments*. Edinburgh: Oliver and Boyd.
- Gauss, C. F. (1809). *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*. Hamburg: Perthes und Besser.
- Good, I. J. (1965). *The estimation of probabilities: An essay on modern Bayesian methods*. Cambridge, MA: MIT Press.
- Good, I. J. (1971). The probabilistic explication of information, evidence, surprise, causality, explanation, and utility (with discussion). In V. P. Godambe & D. A. Sprott (Eds.), *Foundations of statistical inference* (pp. 108-141). Toronto: Holt, Rinehart, and Winston.
- Good, I. J. (1983). *Good thinking. The foundations of probability and its applications*. Mineola, NY: Dover Publications.
- Gorard, S., & White, P. (2017). Still against inferential statistics: Rejoinder to Nicholson and Ridgway. *Statistics Education Research Journal*, 16(1), 70-75.
- Graunt, J. (1662). *Natural and political observations upon the Bills of Mortality, chiefly with reference to the government, religion, trade, growth, air, diseases etc. of the City of London*. London: Royal Society of London.
- Hacking, I. (1965). *The logic of statistical inference*. Cambridge: Cambridge University Press.
- Hacking, I. (1975). *The emergence of probability*. Cambridge: Cambridge University Press.
- Hald, A. (2007). *A history of parametric statistical inference from Bernoulli to Fisher, 1713-1935*. New York: Springer.
- Hartley, D. (1749). *Observations on man, his frame, his duty, and his expectations*. London: Richardson.
- Hilbert, D. (1900). Mathematische Probleme. *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 253-297.
- Jeffreys, H. (1948). *Theory of probability. 2nd ed.* Oxford: Clarendon.

- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-291
- Kolmogorov, A. N. (1956). *Foundations of the theory of probability*. New York: Chelsea.
- Kolmogorov, A. N. (1977). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Ergebnisse der Mathematik, 2. Band, Heft 3. Berlin: Springer (original work published in 1933).
- Laplace, P. S. de (1951). *A philosophical essay on probabilities* (extended version). New York: Dover (original work published in 1812).
- Laplace, P. S. de (1995). *Théorie analytique des probabilités*, 2nd ed. Paris: Courcier (original work published in 1814).
- Mises, R. v. (1919). Grundlagen der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 5, 52-99.
- Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Transactions of the Royal Statistical Society*, 97, 558-625.
- Neyman, J., & Pearson, E. S. (1967). On the use and interpretation of certain test criteria for purposes of statistical inference. Part I and II. *Biometrika* 20A, 175-240; 263-294 (original work published in 1928)
- Popper, K. R. (1959). The propensity interpretation of probability. *British Journal of the Philosophy of Science*, 10, 25-42.
- Popper, K. R. (1962). *Logic of scientific discovery*. London: Routledge (original work published in 1935).
- Porter, T. (1986). *The rise of statistical thinking, 1820-1900*. Princeton, NJ: Princeton University Press.
- Savage, L. J. (Ed.) (1962). *The foundation of statistical inference*. London: Methuen.
- Schnorr, C.P. (1971). *Zufälligkeit und Wahrscheinlichkeit. Eine algorithmische Begründung der Wahrscheinlichkeitstheorie*. Berlin-New York: Springer.
- Seidenfeld, T. (1979). *Philosophical problems of statistical inference – Learning from R. A. Fisher*. Dordrecht: D. Reidel.
- Stegmüller, W. (1973). *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie*, Vol. 4. Berlin: Springer.
- Steinbring, H. (1980). Zur Entwicklung des Wahrscheinlichkeitsbegriffs. Bielefeld: IDM.
- Steinbring, H. (1991). The theoretical nature of probability in the classroom. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters* (pp. 135-168). Dordrecht: Kluwer.
- Venn, J. (1866/1962). *The logic of chance*. Reprinted. New York: Chelsea.

Manfred Borovcnik

Professor at the Department of Statistics, University of Klagenfurt. Master in Mathematics,
Doctor in Statistics. Member of the Ethics Commission of the Federal State of Carinthia.

Elected member of the International Statistical Institute. Past vice president of the
International Association for Statistics Education. Co-editor of *Stochastik in der Schule* and
Co-editor of *Statistics Education Research Journal*. Research interest: Statistics and
probability education E-mail: manfred.borovcnik@auu.at