Foundations of Probability

Thomas Augustin, Marco E. G. V. Cattaneo

Department of Statistics, LMU Munich, Ludwigstraße 33, 80539 München, Germany

---

Introduction

Probability theory is that part of mathematics that is concerned with the description and modeling of random phenomena, or in a more general — but not unanimously accepted — sense, of any kind of uncertainty. Probability is assigned to random events, expressing their tendency to occur in a random experiment, or more generally to propositions, characterizing the degree of belief in their truth.

Probability is the fundamental concept underlying most statistical analyses that go beyond a mere description of the observed data. In statistical inference, where conclusions from a random sample have to be drawn about the properties of the underlying population, arguments based on probability allow to cope with the sampling error and therefore control the inference error, which is necessarily present in any generalization from a part to the whole. Statistical modeling aims at separating regularities (structure explainable by a model) from randomness. There, the sampling error and all the variation that is not explained by the chosen optimal model are comprised in an error probability as a residual category.

Different Interpretations and Their Consequences for Statistical Inference

The concept of probability has a very long history (see for instance Vallverdú, 2011). Originally, the term had a more philosophical meaning, describing the degree of certainty or the persuasive power of an argument. The beginnings of a more mathematical treatment of probability are related to considerations of symmetry in games of chance (see for example Hald, 2003). The scope of the theory was extended by Bernoulli (1713), who applied similar symmetry considerations in the study of epistemic probability in civil, moral, and economic problems. In this connection, he proved his "law of large numbers," which can be seen as the first theorem of mathematical statistics, and as a cornerstone of the *frequentist* interpretation of probability, which understands the probability of an event as the limit of its relative frequency in an infinite sequence of independent repetitions of a random experiment. Typically, the frequentist (or aleatoric) point of view is *objectivist* in the sense that it relates probability to random phenomena only and perceives probability as a property of the random experiment (e.g. rolling a dice) under consideration.

Email addresses: `augustin@stat.uni-muenchen.de` (Thomas Augustin), `cattaneo@stat.uni-muenchen.de` (Marco E. G. V. Cattaneo)

This is the unformatted version of: `doi:10.1007/978-3-642-04898-2_260`

In contrast, the second of the two most common interpretations (see for instance Peterson, 2011, for more details), the *subjective*, personalistic, or epistemic viewpoint, perceives probability as a property of the subject confronted with uncertainty. Consequently, here probability can be assigned to anything the very subject is uncertain about, and the question of whether or not there is an underlying random process vanishes. For the interpretation, in the tradition of Savage (1954) a fictive scenario is used where preferences between actions are described. In particular, the probability of an event is understood as the price at which the subject is indifferent between buying and selling a security paying 1 when the event occurs (and 0 otherwise).

The interpretation of probability predetermines to a considerable extent the choice of the statistical inference methods to learn the unknown parameters ϑ of a statistical model from the data. The frequentist perceives ϑ as an unknown but fixed quantity and seeks methods that are optimal under fictive infinite repetitions of the statistical experiment, while for the subjectivist it is straightforward to express her/his uncertainty about ϑ by a (*prior*) probability distribution, which is then, in the light of new data, updated by the so-called Bayes' rule to obtain the (*posterior*) probability distribution describing all her/his knowledge about ϑ (*Bayesian inference*).

Kolmorgorov's Axioms

While the interpretation of probability is quite important for statistical applications, the mathematical theory of probability can be developed almost independently of the interpretation of probability. The foundations of the modern theory of probability were laid by Kolmogorov (1933) in measure theory: probability is axiomatized as a normalized measure.

More specifically (see for instance Merkle, 2011, and Rudas, 2011, for more details), let Ω be the set of elementary events under consideration (Ω is usually called *sample space*). The events of interest are described as sets of elementary events: it is assumed that they build a σ-algebra $\mathcal{A}$ of subsets of Ω (i.e., $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ is nonempty and closed under complementation and countable union). A probability measure on $(\Omega, \mathcal{A})$ is a function $P : \mathcal{A} \to [0, 1]$ such that $P(\Omega) = 1$ and

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n) \tag{1}$$

for all sequences of pairwise disjoint events $E_1, E_2, \ldots \in \mathcal{A}$. When Ω is uncountable, a Borel σ-algebra is usually selected as the set $\mathcal{A}$ of events of interest, because the natural choice $\mathcal{A} = \mathcal{P}(\Omega)$ would place too strong limitations on the probability measure P, at least under the axiom of choice (see for example Solovay, 1970).

Kolmogorov supplemented his axioms by two further basic definitions: the definition of *independence* of events and the definition of *conditional probability* $P(A|B)$ (that is, the probability of event A given an event B).

From the axioms, fundamental theorems with a strong impact on statistics have been derived on the behavior of independent repetitions of a random experiment (see for instance Billingsley, 1995, and Schervish, 1995, for more details). They include different *laws of large numbers* (see above), the *central limit theorem*, distinguishing the Gaussian distribution as a standard distribution for analyzing large samples, and the *Glivenko–Cantelli theorem*, formulating convergence of the so-called empirical distribution function to its theoretical counterpart, which means, loosely speaking, that the true probability distribution can be rediscovered in a large sample and thus can be learned from data.

Current Discussion and Challenges

In statistical methodology, for a long time Kolmogorov's measure-theoretic axiomatization of probability theory remained almost undisputed: only countable additivity (1) was criticized by some proponents of the subjective interpretation of probability, such as De Finetti (1974–1975). If countable additivity is replaced by the weaker assumption of finite additivity (i.e., $P(E_1 \cup E_2) = P(E_1)+P(E_2)$ for all pairs of disjoint events $E_1, E_2 \in \mathcal{A}$), then it is always possible to assign a probability to any set of elementary events (that is, the natural choice $\mathcal{A} = \mathcal{P}(\Omega)$ does not pose problems anymore). However, without countable additivity many mathematical results of measure theory are not valid anymore.

In recent years, the traditional concept of probability has been questioned in a more fundamental way, especially from the subjectivist point of view. On the basis of severe problems encountered when trying to model uncertain expert knowledge in artificial intelligence, the role of probability as the exclusive methodology for handling uncertainty has been rejected (see for example the introduction of Klir and Wierman, 1999). It is argued that traditional probability is only a one-dimensional, too reductionistic view on the multidimensional phenomenon of uncertainty. Similar conclusions (see for instance Hsu et al., 2005) have been drawn in economic decision theory following Ellsberg's seminal experiments (Ellsberg, 1961), where the extent of ambiguity (or non-stochastic uncertainty) has been distinguished as a constitutive component of decision making.

Such insights have been the driving force for the development of the theory of *imprecise probability* (see for example Coolen et al., 2011, for a brief survey), comprising approaches that formalize the probability of an event A as an interval $[\underline{P}(A), \overline{P}(A)]$, with the difference between $\overline{P}(A)$ and $\underline{P}(A)$ expressing the extent of ambiguity. Here, $\underline{P}$ and $\overline{P}$ are non-additive set-functions, often called *lower* and *upper probabilities*. In particular, Walley (1991) has partially extended De Finetti's framework (De Finetti, 1974–1975) to a behavioral theory of imprecise probability, based on an interpretation of probability as possibly differing buying and selling prices, while Weichselberger (2001) has developed a theory of *interval-probability* by generalizing Kolmogorov's axioms.

References

Bernoulli, J. (1713). *Ars Conjectandi*. Thurneysen.

Billingsley, P. (1995). *Probability and Measure*. 3rd edition, Wiley.

Coolen, F.P.A., Troffaes, M.C.M., and Augustin, T. (2011). Imprecise probability. In: Lovric, M. (ed.), *International Encyclopedia of Statistical Science*, Springer, pp. 645–648.

De Finetti, B. (1974–1975). *Theory of Probability*. 2 volumes, Wiley.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *Q. J. Econ.* 75:643–669.

Hald, A. (2003). *A History of Probability and Statistics and Their Applications before 1750*. Wiley.

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., and Camerer, C.F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310:1680–1683.

Klir, G.J., and Wierman, M.J. (1999). *Uncertainty-Based Information*. Physica-Verlag.

Kolmogorov, A. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer.

Merkle, M. (2011). Measure theory in probability. In: Lovric, M. (ed.), *International Encyclopedia of Statistical Science*, Springer, pp. 793–795.

Peterson, M. (2011). Philosophy of probability. In: Lovric, M. (ed.), *International Encyclopedia of Statistical Science*, Springer, pp. 1065–1069.

Rudas, T. (2011). Probability theory: an outline. In: Lovric, M. (ed.), *International Encyclopedia of Statistical Science*, Springer, pp. 1123–1126.

Savage, L.J. (1954). *The Foundations of Statistics*. Wiley.

Schervish, M.J. (1995). *Theory of Statistics*. Springer.

Solovay, R.M. (1970). A model of set-theory in which every set of reals is Lebesgue measurable. *Ann. Math. (2)* 92:1–56.
Vallverdú, J. (2011). History of probability. In: Lovric, M. (ed.), *International Encyclopedia of Statistical Science*, Springer, pp. 1126–1128.
Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall.
Weichselberger, K. (2001). *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I*. Physica-Verlag.