

Contract no: BAP098M-RG-15-Med

Probability Theory

J.M. Steele

Wharton

Probability theory is that part of mathematics that aims to provide insight into phenomena that depend on chance or on uncertainty. The most prevalent use of the theory comes through the frequentists' interpretation of probability in terms of the outcomes of repeated experiments, but probability is also used to provide a measure of subjective beliefs, especially as judged by one's willingness to place bets.

The roots of probability theory are not as ancient as those of many parts of mathematics, and only in the sixteenth and seventeenth centuries does one find the first glimmerings of the theory in the investigations made by Gerolamo Cardano, Pierre de Fermat, and Blaise Pascal into games of chance. Despite the luminous reputations of these famous mathematicians and philosophers, the subject of probability theory remained on the periphery of respectability, and for a long time development was halting and lugubrious. Through the first third of the twentieth century, the eighteenth century works of Jakob Bernoulli (see Bernoulli Family) and Abraham De Moivre continued to be viewed as the nearly definitive treatises of probability theory.

Still, even in the early days of the twentieth century when probability theory clearly suffered from the lack of a widely accepted foundation, there were profound developments, most notably Albert Einstein's use of Brownian motion in 1905 to provide the first determination of Avagadro's number [7]. Nevertheless, in 1933 when Andrey Nikolayevich Kolmogorov published his elegant succinct volume *Foundations of Probability Theory* [10], the mathematical world was hungry for such a treatment, and the subsequent development of probability theory was explosive.

1 Firm Foundation

Central to Kolmogorov's foundation for probability theory was his introduction of the triple $(\Omega, \mathcal{F}, \mathcal{P})$ that we now call a probability space, or sometimes the "probabilist's trinity". The triple's first element, Ω , is required only to be a set. The second element is a collection of subsets of Ω about which more will be said later. The third element is a function that assigns a real number to each of the elements of \mathcal{F} . This function is called a probability measure \mathcal{P} provided that it satisfies the three following axioms:

Axiom 1. For all $\mathcal{A} \in \mathcal{F}$ we have $\mathcal{P}(\mathcal{A}) \geq 0$.

Axiom 2. For any countable collection $\{\mathcal{A}_i \in \mathcal{F} : 1 \leq i < \infty\}$ for which $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$ for all $i \neq j$, we have

$$\mathcal{P}(\cup_{i=1}^{\infty} \mathcal{A}_i) = \sum_{i=1}^{\infty} \mathcal{P}(\mathcal{A}_i).$$

Axiom 3. $\mathcal{P}(\Omega) = 1$ Axioms 1 and 3 are quite bland. Axiom 1 only captures our understanding that probabilities of events are nonnegative numbers, and Axiom 3 just echoes our assumption that Ω is a sensible representation for the universe of all possible outcomes of the chance experiment being modeled. Only about Axiom 2 can there can be any quarrel, and at times arguments have been made for preferring a probability theory that only requires additivity of probabilities for finite collections of sets. Kolmogorov's decision to assume countable additivity is not the only possible choice, but it has been a fecund one that has proved to be appropriate in a wide variety of circumstances.

The mathematical benefit of Kolmogorov's second axiom is that it connects probability theory with the theory of measure as put forward by Borel, Lebesgue, Radon, and Fréchet in the early part of the twentieth century. It was in fact Fréchet who noted some 13 years after Lebesgue's famous 1902 thesis that the natural domain for a probability measure is a collection of sets that is closed under complementation and countable unions. Fréchet called such collections σ -algebras, and Kolmogorov required that the second term of his triple be just such a collection.

2 Basic Quantities of the Theory

To the practical mind, Kolmogorov's axiomatization of probability may seem only to defer the problem of construction of probability models that serve to inform us about the physical and social world, but by putting the elusive probability function \mathcal{P} on an axiomatic footing Kolmogorov did provide real assurance that one could

study probability as sensibly as one could study measure theory, analysis or algebra. In particular, one could proceed with the investigation of the objects that had been of concern from probability's earliest days.

One of the most fundamental notions of probability theory is the random variable, and in Kolmogorov's framework a random variable is nothing more than a function from $X : \Omega \rightarrow \mathfrak{R}$ with the property that for all t one has that the sets $\{\omega : X(\omega) \leq t\}$ are elements of the σ -algebra \mathcal{F} . With this definition we are on firm footing when we take the definition of the distribution function \mathcal{F} of X to be

$$\mathcal{F}(t) = \mathcal{P}(X \leq t),$$

because the set $\{\omega : X(\omega) \leq t\}$ is in the domain of the set function \mathcal{P} . In this framework the expectation $E(X)$ of the random variable X can be defined as the Lebesgue integral of X with regard to \mathcal{P} , or as the Riemann-Stieltjes integral with respect to \mathcal{F} , giving us

$$E(X) \int_{\Omega} X(\omega) d\mathcal{P}(\omega) = \int_{-\infty}^{\infty} x d\mathcal{F}(x).$$

The probability distribution function and the expectation operation provide us with the core language that is needed to express almost everything that one needs to say about individual random variables. For example, a basic measure of dispersion of a random variable is the variance, which one writes in terms of the expectation as

$$\text{var}(X) = E(X - \mu)^2,$$

where $\mu = E(X)$ and the standard deviation of X is defined to be the square root of the variance.

3 Central Role of Independence

With expectations and distributions we recapture much of the most basic language of probability theory, but the real power of probability theory only emerges with the introduction of the central notion of *independence* of events, algebras, and random variables. To begin that development, one first defines elements \mathcal{A} and \mathcal{B} of \mathcal{F} to be independent provided

$$\mathcal{P}(\mathcal{A} \cap \mathcal{B}) = \mathcal{P}(\mathcal{A})\mathcal{P}(\mathcal{B}).$$

This definition is then extended to sub- σ -algebras of \mathcal{A} and \mathcal{B} of \mathcal{F} by calling \mathcal{A} and \mathcal{B} independent provided \mathcal{A} and \mathcal{B} are independent for all $A \in \mathcal{A}$ and all $B \in \mathcal{B}$. Finally, random variables X and Y are independent if \mathcal{A} and \mathcal{B} independent when these are respectively the smallest σ -algebras containing all the sets $\{X \leq t\}$ and all the sets $\{Y \leq t\}$.

This definition of independence of random variables may look a little burdensome at first, but for many purposes it is much more convenient than the definition of independence that is sometimes given in elementary texts that call for the factor-

ization of the joint density of X and Y . In fact densities may not exist, but that is not the telling point. More to the heart of the matter is that with Kolmogorov's definition one clearly sees that the independence of X and Y implies the independence of $f(X)$ and $g(Y)$ for any monotone functions f and g , while this intuitive fact is cumbersome to check if one needs to verify a density factorization.

4 Theorems That Make the Theory

There are two theorems that live at the very heart of probability theory. The first is the law of large numbers, without which our most fundamental intuitions about the relationship of probability theory and the physical world would be at odds. The second is the central limit theorem, which is arguably the result that most clearly accounts for the practical utility of probability as a helpmate to statistics, as well as to the social and physical sciences.

Theorem 1 (Law of Large Numbers). If $\{X_i : 1 \leq i < \infty\}$ is a sequence of independent random variables, with the distribution function, \mathcal{F} , and if $E | X_i | < \infty$, then the event that the sequence

$$\frac{1}{n}\{X_1 + X_2 + \dots + X_n\}$$

converges to $E(X_1)$ has probability one.

Theorem 2 (Central Limit Theorem). If $\{X_i : 1 \leq i < \infty\}$ is a sequence of independent random variables with distribution function \mathcal{F} , $E(X_i) = \mu < \infty$, and

$\text{var}(X) = \sigma^2 < \infty$, then

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathcal{P}\left(\frac{1}{\alpha\sqrt{n}}\{X_1 + X_2 + \dots + X_n - n\mu\} \leq x\right) \\ = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^x e^{-\mu^2/2} du. \end{aligned}$$

5 Beyond Independent Random Variables

While the purest view of the aims and accomplishments of probability theory may be found in the study of sums of independent random variables, the applications of probability theory require the development of structures that also capture aspects of dependence. To give the simplest illustration of a such a system, we consider a set finite set $\mathcal{S} = \{1, 2, \dots, n\}$ which we will call the set of “states”, and a matrix $P = \{p_{ij}\}$, where all of the matrix entries satisfy $0 \leq p_{ij} \leq 1$ and where the row sums $p_{i1} + p_{i2} + \dots + p_{in}$ all equal one. We now consider a sequence of random variables X_n that are defined by sequential transitions according to the row of the matrix P . Specifically, if $X_n = i$, then X_{n+1} is determined by making a choice from the set \mathcal{S} in accordance with the probability masses (p_{ij}) . Such a sequence of random variables is called a Markov chain, and the theory of such sequences offers an important first step from the core theory of independent random variables. The index of the sequence $\{X_n : n \geq 0\}$ is usually viewed as “time” and an important extension of the notion of a Markov chain is that of a Markov Process where the index is taken to be the whole positive real line and the state space is permitted to be \mathfrak{R}^d (or even a more complex space). The most important such process is *Brownian*

motion.

Another direction for the development of probability theory that goes beyond independence is provided by the theory of *martingales*. On one level, martingales capture the notion of a fair gambling game, and although this view is interesting (and loyal to the origins of probability theory), the theory of martingales turns out to be an appropriate tool for many kinds of investigation (see *Counting Process Methods in Survival Analysis*). In particular, the theory of martingales provides the key to profound connections between the theory of Markov processes and the classical theory of harmonic functions.

References

- a) Adams, W.J. (1974). *The Life and Times of the Central Limit Theorem*, Kaedmon Press, New York.
- b) Chung, K.L. (1974). *Elementary Probability with Stochastic Processes*. Springer-Verlag, New York.
- c) David, F.N. (1962). *Games, Gods, and Gambling: The Origins and History of Probability from the Earliest Times to the Newtonian Era*. Griffin, London.
- d) Doob, Joseph L. (1994). The development of rigor in mathematical probability (1900-1950), in *Development of Mathematics 1900-1950*, J.-P. Pier, ed. Birkhauser-Verlag, Basel.
- e) Dudley, R.M. (1989). *Real Analysis and Probability*. Wadsworth-Brooks/Cole, Pacific Grove.

- f) Durren, R. (1991). *Probability: Theory and Examples*, Wadsworth-Brooks/Cole, Pacific Grove.
- g) Einstein, A. (1905). On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat (in German), *Ann. Pys. (Ser. 4)* 17, 549–560.
- h) Feller, W. (1968). *An Introduction to Probability and Its Applications*. Vol. I, 3rd Ed. Wiley,, New York.
- i) Kolmogorov, A.N. (1933). *Grundbegriffe der Wahrscheinlichkeitrechnung..* Springer-Verlag, Berlin. (English translation: N. Morrison (1956), *Foundations of the Theory of Probability*, Chelsea, New York.)
- j) Stigler, S.M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*. Harvard University Press, Cambridge, MA.

(See also Axioms of Probability; Foundations of Probability)

J.M. Steele