MATHEMATICAL FOUNDATIONS AND INTERPRETATIONS OF PROBABILITY

Wiebe R. Pestman
University of Utrecht, the Netherlands

Keywords: probability space, Kolmogorov triplet, probability theory, conditional probability

Contents

Summary

This article explains how, by axiomatization, intuitive concepts of probability can be turned into a proper mathematical discipline. This is first discussed for probability experiments in which the list of possible outcomes, the so-called outcome space, is finite. As was pointed out in the early days of probability theory, arguments leaning on the assumption of equally probable outcomes can be very powerful in the analysis of such cases. Combinatorics, the art of sophisticated counting, thus enters the scene of probability in a natural way. Then, as a first step in generalizing the theory of finite probability experiments, it is explained how one can axiomatize experiments in which the list of all possible outcomes is countably infinite (which in mathematics is the lowest degree of infiniteness concerning the number of elements in a set). Finally, there is a discussion of what kind of problems mathematicians encounter in axiomatizing experiments in which the list of possible outcomes is uncountably infinite.

1. Introduction

Probability theory is the branch of mathematics dealing with experiments in which randomness plays a role: in other words, experiments in which it is impossible to determine the outcome of the experiment beforehand. In the following text, experiments of this kind will be called *probability experiments.* To arrive at some kind of a definition, one could say that a probability experiment is an experiment which, when repeated under exactly the same conditions, does not necessarily result in the same outcome. Standard examples of probability experiments are a toss of a coin or a throw of a dice. The list of all possible outcomes in a probability experiment is called the *outcome space* or *sample space.* This set is very often denoted by the Greek capital

Ω. The elements of Ω, that is to say the possible outcomes of the experiment, are then usually denoted by ω or by $\omega_1, \omega_2, \omega_3, \ldots$ if more than one outcome is to be dealt with. In the context of probability the subsets of Ω are looked upon as being *events*. This should be read as follows: if *E* is a subset of Ω then it is said that the event *E occurs* if the outcome ω of the experiment happens to be an element of *E*. So, briefly, the event *E* occurs if $\omega \in E$. By thus defining the concept of an event, probability theory is connecting itself directly to mathematical set theory.

Example 1: In a game of chance a dice is cast and the number of spots observed. Of course this is a probability experiment: when repeating the experiment, the observed number of spots is not necessarily the same. The possible outcomes are here:

$$\omega = 1, \quad \omega = 2, \quad \omega = 3, \quad \omega = 4, \quad \omega = 5, \quad \omega = 6 .$$

Grouping these outcomes together in a set gives the outcome space Ω of the experiment:

$$\Omega := \{1, 2, 3, 4, 5, 6\} .$$

An example of a subset of Ω is the set $E := \{2, 4, 6\}$. The corresponding event occurs if the outcome is 2, 4, or 6. Hence the subset *E* stands for the event that the outcome of the experiment is an even number of spots.

Given some subset *E* of Ω one can consider the complement of *E* relative to Ω, denoted by E^c. The subset E^c then corresponds to the event that *E* does not occur, that is, that the outcome of the experiment is not an element of *E*. When two subsets *E* and *F* of Ω are given, then one can form new subsets by taking for example their union $E \cup F$ or their intersection $E \cap F$. The union $E \cup F$ stands for the event that *E* or *F* occurs, where the conjunction 'or' is used in the non-exclusive way. The event $E \cap F$ stands for the event that *E* and *F* occur simultaneously. If *E* and *F* have an empty intersection, briefly if $E \cap F = \varnothing$, then the sets *E* and *F* are said to be *disjoint*. It is then impossible for the corresponding events to occur simultaneously. For that reason disjointness of *E* and *F* is in probabilistic language often expressed by saying that the events *E* and *F* mutually *exclude* each other.

Example 2: Throwing a dice twice the possible outcomes of the experiment can be captured by ordered pairs (i, j), where an outcome (i, j) stands for the event that the first throw shows i and the second j spots. According to this notation the outcome space Ω can be described as: $\Omega := \{(i, j) \mid i, j = 1, 2, 3, 4, 5, 6\}$. The subset $E := \{(4,6), (6,4), (5,5), (5,6), (6,5), (6,6)\}$ thus corresponds to the event that the total number of thrown spots is 10 or more. Hence the complement E^c of *E* stands for the event that the total number of spots is (strictly) less than 10. Let F be the event that the first throw shows six spots, with no specification on the second throw. Then *F* can be presented as a subset of Ω by setting $F := \{(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\}$. The union $E \cup F$ is now given by: $E \cup F = \{(4,6), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\}$.

It presents the event that the first throw shows six spots or that the total number of spots thrown is 10 or more. The intersection $E \cap F$ is here given by:

$$E \cap F = \{(6,4), (6,5), (6,6)\}.$$

This subset represents the event that the first throw shows six spots and that the second throw is such that it makes the total number of spots equal to 10 or more.

Given some probability experiment with outcome space Ω, let E be some fixed subset of Ω. Repeating the experiment N times under exactly the same conditions, one could count the number of times that event E occurs:

$$N(E) := \text{number of times that } E \text{ occurred after } N \text{ trials}. \tag{1}$$

Of course one always has

$$N(E) \leq N \tag{2}$$

The very trivial event $E = \Omega$ occurs in all of the trials and therefore

$$N(\Omega) = N \tag{3}$$

Moreover, as can easily be verified, one has

$$N(E \cup F) = N(E) + N(F) \quad \text{if } E \cap F = \varnothing \tag{4}$$

The rate at which the event E occurs when repeating unendingly can be expressed by the ratio $N(E)/N$, thereby taking N very large. This ratio is understood to be the *probability* that event E occurs and denoted by $\mathbb{P}(E)$. More precisely, the probability that the event E occurs could be thought to be 'defined' as

$$\mathbb{P}(E) := \lim_{N \to \infty} \frac{N(E)}{N} \tag{5}$$

If, for example, one has $\mathbb{P}(E) = 0.40$ then, when repeating unendingly, the event E can be expected to occur in 40 percent of the repetitions.

Starting from the provisory definition given in (5), the next three properties can directly be distilled from (2), (3) and (4).

$$0 \leq \mathbb{P}(E) \leq 1 \tag{6}$$

$$\mathbb{P}(\Omega) = 1 \tag{7}$$

$$\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) \quad \text{if } E \cap F = \varnothing \tag{8}$$

From the mathematical point of view, however, the way of defining $\mathbb{P}$ by means of (5) is quite unsatisfactory, there being no guarantee at all that the limit on the right side of (5) actually exists. It is not possible to deduce this from the given setup. In early days of mathematics, the Greek mathematician Euclid faced similar problems in his efforts to set up geometry as a deductive science. For example, Euclid found himself unable to prove that, given a straight line and a point (not necessarily on the line), there is exactly one straight line through this point that is parallel to the given straight line. To get around this kind of problems, Euclid defined straight lines as objects having some basic properties. These basic properties were called *axioms.* All other properties were thereupon to be deduced by logical reasoning from the given axioms. Thus geometry was born as a deductive science. In very much the same way, mathematicians have been axiomatizing the concept of probability. In the heuristic setup given above it is clear that, in defining the concept of probability, there must be some mechanism that assigns to every subset $E \subset \Omega$ a number $\mathbb{P}(E)$ between zero and one, representing the chance that event E will occur. Denoting the collection of all subsets of Ω by $\mathcal{P}(\Omega)$ one could express this more precisely as follows: in axiomatizing the concept of probability along the heuristic lines given previously, there must be some function of the form

$$\mathbb{P}:\mathcal{P}(\Omega)\to[0,1] \tag{9}$$

That is to say, a function

$$E \mapsto \mathbb{P}(E) \tag{10}$$

where $\mathbb{P}(E)$ is always a number between zero and one. This function should have certain properties. The heuristic setup given before suggests that it should at least have the properties (6), (7) and (8). Mathematicians have thus been trying to capture the concept of a probability experiment in the following abstract way.

Definition 1: A *probability experiment* or a *probability space* is a pair $(\Omega, \mathbb{P})$ where Ω is some set and where $\mathbb{P}:\mathcal{P}(\Omega)\to[0,1]$ is a function with the properties

(a) $\mathbb{P}(\Omega)=1$.

(a) (b) $\mathbb{P}(E\cup F)=\mathbb{P}(E)+\mathbb{P}(F)$ if $E\cap F=\varnothing$.

Functions defined on $\mathcal{P}(\Omega)$ are usually called *set functions.* Property (a) is often referred to by saying that the set function $\mathbb{P}$ is *normalized*. Property (b) is usually expressed by saying that $\mathbb{P}$ is an *additive* set function. Definition 1, as will be explained in the sections that follow, is not quite satisfactory in many cases. However, in cases where Ω is a finite set there are no problems at all. It should be noted that from Definition 1 a number of very useful properties can be deduced. First of all the trivial property that

$$\mathbb{P}(\varnothing)=0. \tag{11}$$

Moreover, the rule for taking complements can be deduced:

$$\mathbb{P}(E^c)=1-\mathbb{P}(E) \tag{12}$$

It can also be deduced from Definition 1 that

$$\mathbb{P}(E)\le\mathbb{P}(F) \quad \text{if } E\subset F. \tag{13}$$

For every two subset E and F, not necessarily disjoint, one has
$\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$. (14)

It follows from this that
$\mathbb{P}(E \cup F) \leq \mathbb{P}(E) + \mathbb{P}(F)$ for every pair of subsets E and F. (15)

By recurrence one can prove that for every finite sequence of mutually disjoint subsets $E_1, E_2, \ldots, E_n$ one has

$$\mathbb{P}(E_1 \cup E_2 \cup \cdots \cup E_n) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \cdots + \mathbb{P}(E_n). \tag{16}$$

When the $E_1, E_2, \ldots, E_n$ are not necessarily disjoint, then

$$\mathbb{P}(E_1 \cup E_2 \cup \cdots \cup E_n) \leq \mathbb{P}(E_1) + \mathbb{P}(E_2) + \cdots + \mathbb{P}(E_n) \tag{17}$$

which generalizes (15). For n arbitrary subsets $E_1, E_2, \ldots, E_n$ formula (14) can be generalized to

$$\mathbb{P}(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{i} \mathbb{P}(E_i) - \sum_{i<j} \mathbb{P}(E_i \cap E_j) + \sum_{i<j<k} \mathbb{P}(E_i \cap E_j \cap E_k) - \cdots \cdots + (-1)^{n-1} \mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n). \tag{18}$$

All these very basic laws of probability can be deduced from Definition 1.

-
-
-

Bibliography

Banach S. (1923). Sur le problème de la mesure. *Fundamenta Mathematicae* 4, 7–33. [This very interesting article treats the existence of rotation invariant additive set functions on a circle.]

Bauer H. (1996). *Probability Theory*. Berlin: Walter de Gruyter. 523 pp. [A thorough introduction to probability theory for mathematics students. Some knowledge of measure theory is a prerequisite.]

Chandra T.K. and Chatterjee D. (2001). *A First Course in Probability*. Pangbourne, UK: Alpha Science. 468 pp. [A fine modern introduction to probability. The development is systematic and rigorous, with many interesting examples illustrating the theory.]

Feller W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. 1, 2. New York: John Wiley & Sons. Vol I 528 pp., Vol II 704 pp. [Masterful introduction to the subject. In the first chapters of volume 1 foundations and interpretations of probability are discussed in some detail, thereby giving many interesting examples. The phenomenon that intuitive symmetrization in the theory of elementary particles is not always allowed is also discussed.]

Fine T.L. (1973). *Theories of Probability*. New York: Academic Press. 263 pp. [A must for anyone interested in axiomatization of probability.]

Halmos P.R. (1964). *Naïve Set Theory*. Princeton: Van Nostrand. 104 pp. [A very readable little book introducing the reader to the foundations of set theory.]

Hausdorff F. (1978). *Grundzüge der Mengenlehre*. New York: reprinted by Chelsea. 476 pp. [The first edition of this masterwork appeared in 1914! It is still a very useful introduction to set theory. The problems discussed in the last section of this article are treated in some detail in chapter 10.]

Kolmogorov A.N. (1956). *Foundations of the Theory of Probability*. New York: Chelsea. 84 pp. [Excellent introduction to the subject, written by the founder of modern probability theory himself.]

Kümmerer B. and Maassen H. (1998). Elements of Quantum Probability. *Quantum Probability Communications* 10, 73–100. [A clear introduction to the subject. Little mathematical foreknowledge required.]

Meyer P.A. (1993). *Quantum Probability for Probabilists. Lecture Notes in Mathematics No. 1538*. Berlin: Springer Verlag. 287 pp. [An introduction to the subject for mathematicians.]

Mises von R. and Geiringer H. (1964). *The Mathematical Theory of Probability and Statistics*. New York: Academic Press. 694 pp. [In this book more constructive alternatives to the Kolmogorov axiomatization are discussed.]

Pestman W.R. (in press). *Measure, Probability and Linear Space Techniques*. [Modern introduction to measure theory and its natural connection to probability theory. The problems encountered in the last section of this article are discussed in some detail.] (In press: draft edition available.)

Robinson R.M. (1947). On the decomposition of spheres. *Fundamenta Mathematicae* 34, 246–256). [In this article some pathological decompositions of spheres are discussed. These all draw on the axiom of choice in set theory and seem to contradict every form of intuition.]

Ventsel H. (1977). *Théorie des Probabilités*. Moscow : Edition Mir. 563 pp. [An excellent introduction to probability theory. The absolute basics of set theory are also explained; in a clear, pragmatic, and pleasant style, with very little mathematical foreknowledge required. Also available in the Russian language, in which it was originally written.]

Biographical Sketch

Wiebe R. Pestman studied mathematics, physics, and astronomy at the University of Groningen, where he also obtained his doctor's degree in mathematics. He then opted out of scientific life for a while, before eventually returning to mathematics and becoming lecturer at the University of Nijmegen. Pestman is currently a lecturer at the University of Utrecht. His mathematical interests are in functional analysis, probability and statistics, operator algebras, and harmonic analysis. He has published several works on various mathematical topics, including a textbook on mathematical statistics.