WIKIPEDIA

Probability axioms

The **Kolmogorov axioms** are the foundations of probability theory introduced by Andrey Kolmogorov in 1933.[1] These axioms remain central and have direct contributions to mathematics, the physical sciences, and real-world probability cases.[2] An alternative approach to formalising probability, favoured by some Bayesians, is given by Cox's theorem.[3]

Contents

Axioms

The assumptions as to setting up the axioms can be summarised as follows: Let (Ω, F, P) be a measure space with $P(E)$ being the probability of some event E, and $P(\Omega) = 1$. Then (Ω, F, P) is a probability space, with sample space Ω, event space F and probability measure P.[1]

First axiom

The probability of an event is a non-negative real number:

$$P(E) \in \mathbb{R}, P(E) \geq 0 \qquad \forall E \in F$$

where F is the event space. It follows that $P(E)$ is always finite, in contrast with more general measure theory. Theories which assign negative probability relax the first axiom.

Second axiom

This is the assumption of unit measure: that the probability that at least one of the elementary events in the entire sample space will occur is 1

$$P(\Omega) = 1.$$

Third axiom

This is the assumption of σ-additivity:

> Any countable sequence of disjoint sets (synonymous with *mutually exclusive* events) $E_1, E_2, \ldots$ satisfies
>
> $$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i).$$

Some authors consider merely finitely additive probability spaces, in which case one just needs an algebra of sets, rather than a σ-algebra.[4] Quasiprobability distributions in general relax the third axiom.

Consequences

From the Kolmogorov axioms, one can deduce other useful rules for studying probabilities. The proofs[5][6][7] of these rules are a very insightful procedure that illustrates the power of the third axiom, and its interaction with the remaining two axioms. Four of the immediate corollaries and their proofs are shown below:

Monotonicity

$$\quad \text{if} \quad A \subseteq B \quad \text{then} \quad P(A) \leq P(B).$$

If A is a subset of, or equal to B, then the probability of A is less than, or equal to the probability of B.

Proof of monotonicity[5]

In order to verify the monotonicity property, we set $E_1 = A$ and $E_2 = B \setminus A$, where $A \subseteq B$ and $E_i = \varnothing$ for $i \geq 3$. From the properties of the empty set ($\varnothing$), it is easy to see that the sets E_i are pairwise disjoint and $E_1 \cup E_2 \cup \cdots = B$. Hence, we obtain from the third axiom that

$$P(A) + P(B \setminus A) + \sum_{i=3}^{\infty} P(E_i) = P(B).$$

Since, by the first axiom, the left-hand side of this equation is a series of non-negative numbers, and since it converges to $P(B)$ which is finite, we obtain both $P(A) \leq P(B)$ and $P(\varnothing) = 0$.

The probability of the empty set

$$P(\varnothing) = 0.$$

In some cases, $\varnothing$ is not the only event with probability 0.

Proof of probability of the empty set

As shown in the previous proof, $P(\varnothing) = 0$. This statement can be proved by contradiction: if $P(\varnothing) = a, a > 0$ then the left hand side $[P(A) + P(B \setminus A) + \sum_{i=3}^{\infty} P(E_i)]$ is infinite;

$$\sum_{i=3}^{\infty} P(E_i) = \sum_{i=3}^{\infty} P(\varnothing) = \sum_{i=3}^{\infty} a = \begin{cases} 0 & \text{if } a = 0, \\ \infty & \text{if } a > 0. \end{cases}$$

If $a > 0$ we have a contradiction, because the left hand side is infinite while $P(B)$ must be finite (from the first axiom). Thus, $a = 0$. We have shown as a byproduct of the proof of monotonicity that $P(\varnothing) = 0$.

The complement rule

$$P(A^c) = P(\Omega \setminus A) = 1 - P(A)$$

Proof of the complement rule

Given A and A^c are mutually exclusive and that $A \cup A^c = \Omega$:

$P(A \cup A^c) = P(A) + P(A^c)$ ... *(by axiom 3)*

and, $P(A \cup A^c) = P(\Omega) = 1$ ... *(by axiom 2)*

$\Rightarrow P(A) + P(A^c) = 1$

$\therefore P(A^c) = 1 - P(A)$

The numeric bound

It immediately follows from the monotonicity property that

$$0 \leq P(E) \leq 1 \qquad \forall E \in F.$$

Proof of the numeric bound

Given the complement rule $P(E^c) = 1 - P(E)$ and *axiom 1* $P(E^c) \geq 0$:

$1 - P(E) \geq 0$

$\Rightarrow 1 \geq P(E)$

$\therefore 0 \leq P(E) \leq 1$

Further consequences

Another important property is:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

This is called the addition law of probability, or the sum rule. That is, the probability that an event in A *or* B will happen is the sum of the probability of an event in A and the probability of an event in B, minus the probability of an event that is in both A *and* B. The proof of this is as follows:

Firstly,

$$P(A \cup B) = P(A) + P(B \setminus A)$$ ... *(by Axiom 3)*

So,

$$P(A \cup B) = P(A) + P(B \setminus (A \cap B))$$ (by $B \setminus A = B \setminus (A \cap B)$).

Also,

$$P(B) = P(B \setminus (A \cap B)) + P(A \cap B)$$

and eliminating $P(B \setminus (A \cap B))$ from both equations gives us the desired result.

An extension of the addition law to any number of sets is the inclusion–exclusion principle.

Setting B to the complement A^c of A in the addition law gives

$$P\left(A^c\right) = P(\Omega \setminus A) = 1 - P(A)$$

That is, the probability that any event will *not* happen (or the event's complement) is 1 minus the probability that it will.

Simple example: coin toss

Consider a single coin-toss, and assume that the coin will either land heads (H) or tails (T) (but not both). No assumption is made as to whether the coin is fair.

We may define:

$$\Omega = \{H, T\}$$
$$F = \{\varnothing, \{H\}, \{T\}, \{H, T\}\}$$

Kolmogorov's axioms imply that:

$$P(\varnothing) = 0$$

The probability of *neither* heads *nor* tails, is 0.

$$P(\{H, T\}^c) = 0$$

The probability of *either* heads *or* tails, is 1.

$$P(\{H\}) + P(\{T\}) = 1$$

The sum of the probability of heads and the probability of tails, is 1.

See also

- Borel algebra
- Conditional probability – Probability of an event occurring, given that another event has already occurred
- Fully probabilistic design
- Intuitive statistics
- Quasiprobability
- Set theory – Branch of mathematics that studies sets
- σ-algebra

References

1. Kolmogorov, Andrey (1950) [1933]. *Foundations of the theory of probability* (https://archive.org/details/foundationsofthe00kolm). New York, USA: Chelsea Publishing Company.
2. Aldous, David. "What is the significance of the Kolmogorov axioms?" (https://www.stat.berkeley.edu/~aldous/Real_World/kolmogorov.html). *David Aldous*. Retrieved November 19, 2019.
3. Terenin Alexander; David Draper (2015). "Cox's Theorem and the Jaynesian Interpretation of Probability" (https://archive.org/details/arxiv-1507.06597). arXiv.org. arXiv:1507.06597 (https://arxiv.org/abs/1507.06597). Bibcode:2015arXiv150706597T (https://ui.adsabs.harvard.edu/abs/2015arXiv150706597T).
4. Hájek, Alan (August 28, 2019). "Interpretations of Probability" (https://plato.stanford.edu/entries/probability-interpret/#KolProCal). *Stanford Encyclopedia of Philosophy*. Retrieved November 17, 2019.
5. Ross, Sheldon M. (2014). *A first course in probability* (Ninth ed.). Upper Saddle River, New Jersey. pp. 27, 28. ISBN 978-0-321-79477-2. OCLC 827003384 (https://www.worldcat.org/oclc/827003384).
6. Gerard, David (December 9, 2017). "Proofs from axioms" (https://dcgerard.github.io/stat234/11_proofs_from_axioms.pdf) (PDF). Retrieved November 20, 2019.
7. Jackson, Bill (2010). "Probability (Lecture Notes - Week 3)" (http://www.maths.qmul.ac.uk/~bill/MTH4107/notesweek3_10.pdf) (PDF). *School of Mathematics, Queen Mary University of London*. Retrieved November 20, 2019.

Further reading

- DeGroot, Morris H. (1975). *Probability and Statistics* (https://archive.org/details/probabilitystati0000degr/page/12). Reading: Addison-Wesley. pp. 12–16 (https://archive.org/details/probabilitystati0000degr/page/12). ISBN 0-201-01503-X.
- McCord, James R.; Moroney, Richard M. (1964). "Axiomatic Probability" (https://archive.org/details/introductiontopr00mcco). *Introduction to Probability Theory*. New York: Macmillan. pp. 13–28 (https://archive.org/details/introductiontopr00mcco/page/13).
- Formal definition (https://web.archive.org/web/20130923121802/http://mws.cs.ru.nl/mwiki/prob_1.html#M2) of probability in the Mizar system, and the list of theorems (http://mmlquery.mizar.org/cgi-bin/mmlquery/emacs_search?input=(symbol+Probability+%7C+notation+%7C+constructor+%7C+occur+%7C+th)+ordered+by+number+of+ref) formally proved about it.