



Probability space

In probability theory, a **probability space** or a **probability triple** (Ω, \mathcal{F}, P) is a mathematical construct that provides a formal model of a random process or "experiment". For example, one can define a probability space which models the throwing of a die.

A probability space consists of three elements:^{[1][2]}

1. A **sample space**, Ω , which is the set of all possible outcomes.
2. An **event space**, which is a set of events, \mathcal{F} , an event being a set of outcomes in the sample space.
3. A **probability function**, P , which assigns, to each event in the event space, a probability, which is a number between 0 and 1 (inclusive).

In order to provide a sensible model of probability, these elements must satisfy a number of axioms, detailed in this article.

In the example of the throw of a standard die, we would take the sample space to be $\{1, 2, 3, 4, 5, 6\}$. For the event space, we could simply use the set of all subsets of the sample space, which would then contain simple events such as $\{5\}$ ("the die lands on 5"), as well as complex events such as $\{2, 4, 6\}$ ("the die lands on an even number"). Finally, for the probability function, we would map each event to the number of outcomes in that event divided by 6 – so for example, $\{5\}$ would be mapped to $1/6$, and $\{2, 4, 6\}$ would be mapped to $3/6 = 1/2$.

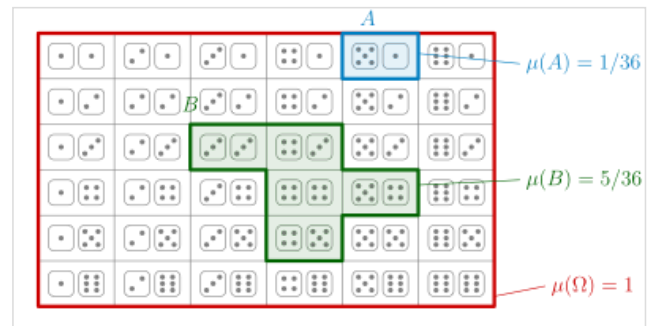
When an experiment is conducted, we imagine that "nature" "selects" a single outcome, ω , from the sample space Ω . All the events in the event space \mathcal{F} that contain the selected outcome ω are said to "have occurred". This "selection" happens in such a way that if the experiment were repeated many times, the number of occurrences of each event, as a fraction of the total number of experiments, would most likely tend towards the probability assigned to that event by the probability function P .

The Soviet mathematician Andrey Kolmogorov introduced the notion of probability space, together with other axioms of probability, in the 1930s. In modern probability theory there are a number of alternative approaches for axiomatization – for example, algebra of random variables.

Introduction

A probability space is a mathematical triplet (Ω, \mathcal{F}, P) that presents a model for a particular class of real-world situations. As with other models, its author ultimately defines which elements Ω , \mathcal{F} , and P will contain.

- The sample space Ω is the set of all possible outcomes. An outcome is the result of a single execution of the model. Outcomes may be states of nature, possibilities, experimental results and the like. Every instance of the real-world situation (or run of the experiment) must produce exactly one outcome. If outcomes of different runs of an experiment differ in any way that matters, they are distinct outcomes. Which differences matter depends on the kind of analysis we want to do. This leads to different choices of sample space.
- The σ -algebra \mathcal{F} is a collection of all the events we would like to consider. This collection may or may not include each of the elementary events. Here, an "event" is a set of zero or more outcomes; that is, a subset of the sample space. An event is considered to have "happened" during an experiment when the outcome of the latter is an element of the event. Since the same outcome may be a member of many events, it is possible for many events to have happened given a single outcome. For example, when the trial consists of throwing two dice, the set of all outcomes with a sum of 7 pips may constitute an event, whereas outcomes with an odd number of pips may constitute another event. If the outcome is the element of the elementary event of two pips on the first die and five on the second, then both of the events, "7 pips" and "odd number of pips", are said to have happened.
- The probability measure P is a set function returning an event's probability. A probability is a real number between zero (impossible events have probability zero, though probability-zero events are not necessarily impossible) and one (the event happens almost surely, with almost total certainty). Thus P is a function $P : \mathcal{F} \rightarrow [0, 1]$. The probability measure function must satisfy two simple requirements: First, the probability of a countable union of mutually exclusive events must be equal to the countable sum of the probabilities of each of these events. For example, the probability of the union of the mutually exclusive events **Head** and **Tail** in the random experiment of one coin toss, $P(\text{Head} \cup \text{Tail})$, is the sum of probability for



Probability space for throwing a die twice in succession: The sample space Ω consists of all 36 possible outcomes; three different events (colored polygons) are shown, with their respective probabilities (assuming a discrete uniform distribution).

Head and the probability for **Tail**, $P(\text{Head}) + P(\text{Tail})$. Second, the probability of the sample space Ω must be equal to 1 (which accounts for the fact that, given an execution of the model, some outcome must occur). In the previous example the probability of the set of outcomes $P(\{\text{Head}, \text{Tail}\})$ must be equal to one, because it is entirely certain that the outcome will be either **Head** or **Tail** (the model neglects any other possibility) in a single coin toss.

Not every subset of the sample space Ω must necessarily be considered an event: some of the subsets are simply not of interest, others cannot be "measured". This is not so obvious in a case like a coin toss. In a different example, one could consider javelin throw lengths, where the events typically are intervals like "between 60 and 65 meters" and unions of such intervals, but not sets like the "irrational numbers between 60 and 65 meters".

Definition

In short, a probability space is a measure space such that the measure of the whole space is equal to one.

The expanded definition is the following: a probability space is a triple (Ω, \mathcal{F}, P) consisting of:

- the sample space Ω – an arbitrary non-empty set,
- the σ -algebra $\mathcal{F} \subseteq 2^\Omega$ (also called σ -field) – a set of subsets of Ω , called events, such that:
 - \mathcal{F} contains the sample space: $\Omega \in \mathcal{F}$,
 - \mathcal{F} is closed under complements: if $A \in \mathcal{F}$, then also $(\Omega \setminus A) \in \mathcal{F}$,
 - \mathcal{F} is closed under countable unions: if $A_i \in \mathcal{F}$ for $i = 1, 2, \dots$, then also $(\bigcup_{i=1}^\infty A_i) \in \mathcal{F}$
 - The corollary from the previous two properties and De Morgan's law is that \mathcal{F} is also closed under countable intersections: if $A_i \in \mathcal{F}$ for $i = 1, 2, \dots$, then also $(\bigcap_{i=1}^\infty A_i) \in \mathcal{F}$
- the probability measure $P : \mathcal{F} \rightarrow [0, 1]$ – a function on \mathcal{F} such that:
 - P is countably additive (also called σ -additive): if $\{A_i\}_{i=1}^\infty \subseteq \mathcal{F}$ is a countable collection of pairwise disjoint sets, then $P(\bigcup_{i=1}^\infty A_i) = \sum_{i=1}^\infty P(A_i)$,
 - the measure of the entire sample space is equal to one: $P(\Omega) = 1$.

Discrete case

Discrete probability theory needs only at most countable sample spaces Ω . Probabilities can be ascribed to points of Ω by the probability mass function $p : \Omega \rightarrow [0, 1]$ such that $\sum_{\omega \in \Omega} p(\omega) = 1$. All subsets of Ω can be treated as events (thus, $\mathcal{F} = 2^\Omega$ is the power set). The probability measure takes the simple form

$$P(A) = \sum_{\omega \in A} p(\omega) \quad \text{for all } A \subseteq \Omega. \quad (*)$$

The greatest σ -algebra $\mathcal{F} = 2^\Omega$ describes the complete information. In general, a σ -algebra $\mathcal{F} \subseteq 2^\Omega$ corresponds to a finite or countable partition $\Omega = B_1 \cup B_2 \cup \dots$, the general form of an event $A \in \mathcal{F}$ being $A = B_{k_1} \cup B_{k_2} \cup \dots$. See also the examples.

The case $p(\omega) = 0$ is permitted by the definition, but rarely used, since such ω can safely be excluded from the sample space.

General case

If Ω is uncountable, still, it may happen that $P(\omega) \neq 0$ for some ω ; such ω are called atoms. They are an at most countable (maybe empty) set, whose probability is the sum of probabilities of all atoms. If this sum is equal to 1 then all other points can safely be excluded from the sample space, returning us to the discrete case. Otherwise, if the sum of probabilities of all atoms is between 0 and 1, then the probability space decomposes into a discrete (atomic) part (maybe empty) and a non-atomic part.

Non-atomic case

If $P(\omega) = 0$ for all $\omega \in \Omega$ (in this case, Ω must be uncountable, because otherwise $P(\Omega) = 1$ could not be satisfied), then equation $(*)$ fails: the probability of a set is not necessarily the sum over the probabilities of its elements, as summation is only defined for countable numbers of elements. This makes the probability space theory much more technical. A formulation stronger than summation, measure theory is applicable. Initially the probabilities are ascribed to some "generator" sets (see the examples). Then

a limiting procedure allows assigning probabilities to sets that are limits of sequences of generator sets, or limits of limits, and so on. All these sets are the σ -algebra \mathcal{F} . For technical details see Carathéodory's extension theorem. Sets belonging to \mathcal{F} are called measurable. In general they are much more complicated than generator sets, but much better than non-measurable sets.

Complete probability space

A probability space (Ω, \mathcal{F}, P) is said to be a complete probability space if for all $B \in \mathcal{F}$ with $P(B) = 0$ and all $A \subset B$ one has $A \in \mathcal{F}$. Often, the study of probability spaces is restricted to complete probability spaces.

Examples

Discrete examples

Example 1

If the experiment consists of just one flip of a fair coin, then the outcome is either heads or tails: $\Omega = \{\mathbf{H}, \mathbf{T}\}$. The σ -algebra $\mathcal{F} = 2^\Omega$ contains $2^2 = 4$ events, namely: $\{\mathbf{H}\}$ ("heads"), $\{\mathbf{T}\}$ ("tails"), $\{\}$ ("neither heads nor tails"), and $\{\mathbf{H}, \mathbf{T}\}$ ("either heads or tails"); in other words, $\mathcal{F} = \{\{\}, \{\mathbf{H}\}, \{\mathbf{T}\}, \{\mathbf{H}, \mathbf{T}\}\}$. There is a fifty percent chance of tossing heads and fifty percent for tails, so the probability measure in this example is $P(\{\}) = 0$, $P(\{\mathbf{H}\}) = 0.5$, $P(\{\mathbf{T}\}) = 0.5$, $P(\{\mathbf{H}, \mathbf{T}\}) = 1$.

Example 2

The fair coin is tossed three times. There are 8 possible outcomes: $\Omega = \{\text{HHH}, \text{HHT}, \text{HTH}, \text{HTT}, \text{THH}, \text{THT}, \text{TTH}, \text{TTT}\}$ (here "HTH" for example means that first time the coin landed heads, the second time tails, and the last time heads again). The complete information is described by the σ -algebra $\mathcal{F} = 2^\Omega$ of $2^8 = 256$ events, where each of the events is a subset of Ω .

Alice knows the outcome of the second toss only. Thus her incomplete information is described by the partition $\Omega = A_1 \sqcup A_2 = \{\text{HHH}, \text{HHT}, \text{THH}, \text{THT}\} \sqcup \{\text{HTH}, \text{HTT}, \text{TTH}, \text{TTT}\}$, where \sqcup is the *disjoint union*, and the corresponding σ -algebra $\mathcal{F}_{\text{Alice}} = \{\{\}, A_1, A_2, \Omega\}$. Bryan knows only the total number of tails. His partition contains four parts: $\Omega = B_0 \sqcup B_1 \sqcup B_2 \sqcup B_3 = \{\text{HHH}\} \sqcup \{\text{HHT}, \text{HTH}, \text{THH}\} \sqcup \{\text{TTH}, \text{THT}, \text{HTT}\} \sqcup \{\text{TTT}\}$; accordingly, his σ -algebra $\mathcal{F}_{\text{Bryan}}$ contains $2^4 = 16$ events.

The two σ -algebras are incomparable: neither $\mathcal{F}_{\text{Alice}} \subseteq \mathcal{F}_{\text{Bryan}}$ nor $\mathcal{F}_{\text{Bryan}} \subseteq \mathcal{F}_{\text{Alice}}$; both are sub- σ -algebras of 2^Ω .

Example 3

If 100 voters are to be drawn randomly from among all voters in California and asked whom they will vote for governor, then the set of all sequences of 100 Californian voters would be the sample space Ω . We assume that sampling without replacement is used: only sequences of 100 *different* voters are allowed. For simplicity an ordered sample is considered, that is a sequence (Alice, Bryan) is different from (Bryan, Alice). We also take for granted that each potential voter knows exactly his/her future choice, that is he/she does not choose randomly.

Alice knows only whether or not Arnold Schwarzenegger has received at least 60 votes. Her incomplete information is described by the σ -algebra $\mathcal{F}_{\text{Alice}}$ that contains: (1) the set of all sequences in Ω where at least 60 people vote for Schwarzenegger; (2) the set of all sequences where fewer than 60 vote for Schwarzenegger; (3) the whole sample space Ω ; and (4) the empty set \emptyset .

Bryan knows the exact number of voters who are going to vote for Schwarzenegger. His incomplete information is described by the corresponding partition $\Omega = B_0 \sqcup B_1 \sqcup \dots \sqcup B_{100}$ and the σ -algebra $\mathcal{F}_{\text{Bryan}}$ consists of 2^{101} events.

In this case, Alice's σ -algebra is a subset of Bryan's: $\mathcal{F}_{\text{Alice}} \subset \mathcal{F}_{\text{Bryan}}$. Bryan's σ -algebra is in turn a subset of the much larger "complete information" σ -algebra 2^Ω consisting of $2^{n(n-1)\cdots(n-99)}$ events, where n is the number of all potential voters in California.

Non-atomic examples

Example 4

A number between 0 and 1 is chosen at random, uniformly. Here $\Omega = [0,1]$, \mathcal{F} is the σ -algebra of Borel sets on Ω , and P is the Lebesgue measure on $[0,1]$.

In this case, the open intervals of the form (a,b) , where $0 < a < b < 1$, could be taken as the generator sets. Each such set can be ascribed the probability of $P((a,b)) = (b - a)$, which generates the Lebesgue measure on $[0,1]$, and the Borel σ -algebra on Ω .

Example 5

A fair coin is tossed endlessly. Here one can take $\Omega = \{0,1\}^\infty$, the set of all infinite sequences of numbers 0 and 1. Cylinder sets $\{(x_1, x_2, \dots) \in \Omega : x_1 = a_1, \dots, x_n = a_n\}$ may be used as the generator sets. Each such set describes an event in which the first n tosses have resulted in a fixed sequence (a_1, \dots, a_n) , and the rest of the sequence may be arbitrary. Each such event can be naturally given the probability of 2^{-n} .

These two non-atomic examples are closely related: a sequence $(x_1, x_2, \dots) \in \{0,1\}^\infty$ leads to the number $2^{-1}x_1 + 2^{-2}x_2 + \dots \in [0,1]$. This is not a one-to-one correspondence between $\{0,1\}^\infty$ and $[0,1]$ however: it is an isomorphism modulo zero, which allows for treating the two probability spaces as two forms of the same probability space. In fact, all non-pathological non-atomic probability spaces are the same in this sense. They are so-called standard probability spaces. Basic applications of probability spaces are insensitive to standardness. However, non-discrete conditioning is easy and natural on standard probability spaces, otherwise it becomes obscure.

Related concepts

Probability distribution

Any probability distribution defines a probability measure.

Random variables

A random variable X is a measurable function $X: \Omega \rightarrow S$ from the sample space Ω to another measurable space S called the *state space*.

If $A \subset S$, the notation $\Pr(X \in A)$ is a commonly used shorthand for $\Pr(\{\omega \in \Omega : X(\omega) \in A\})$.

Defining the events in terms of the sample space

If Ω is countable, we almost always define \mathcal{F} as the power set of Ω , i.e. $\mathcal{F} = 2^\Omega$ which is trivially a σ -algebra and the biggest one we can create using Ω . We can therefore omit \mathcal{F} and just write (Ω, P) to define the probability space.

On the other hand, if Ω is uncountable and we use $\mathcal{F} = 2^\Omega$ we get into trouble defining our probability measure P because \mathcal{F} is too "large", i.e. there will often be sets to which it will be impossible to assign a unique measure. In this case, we have to use a smaller σ -algebra \mathcal{F} , for example the Borel algebra of Ω , which is the smallest σ -algebra that makes all open sets measurable.

Conditional probability

Kolmogorov's definition of probability spaces gives rise to the natural concept of conditional probability. Every set A with non-zero probability (that is, $P(A) > 0$) defines another probability measure

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)}$$

on the space. This is usually pronounced as the "probability of B given A ".

For any event A such that $P(A) > 0$, the function Q defined by $Q(B) = P(B \mid A)$ for all events B is itself a probability measure.

Independence

Two events, A and B are said to be independent if $P(A \cap B) = P(A) P(B)$.

Two random variables, X and Y , are said to be independent if any event defined in terms of X is independent of any event defined in terms of Y . Formally, they generate independent σ -algebras, where two σ -algebras G and H , which are subsets of F are said to be independent if any element of G is independent of any element of H .

Mutual exclusivity

Two events, A and B are said to be mutually exclusive or *disjoint* if the occurrence of one implies the non-occurrence of the other, i.e., their intersection is empty. This is a stronger condition than the probability of their intersection being zero.

If A and B are disjoint events, then $P(A \cup B) = P(A) + P(B)$. This extends to a (finite or countably infinite) sequence of events. However, the probability of the union of an uncountable set of events is not the sum of their probabilities. For example, if Z is a normally distributed random variable, then $P(Z = x)$ is 0 for any x , but $P(Z \in \mathbf{R}) = 1$.

The event $A \cap B$ is referred to as " A and B ", and the event $A \cup B$ as " A or B ".

See also

- Space (mathematics)
- Measure space
- Fuzzy measure theory
- Filtered probability space
- Talagrand's concentration inequality

References

1. Loève, Michel. Probability Theory, Vol 1. New York: D. Van Nostrand Company, 1955.
2. Stroock, D. W. (1999). Probability theory: an analytic view. Cambridge University Press.

Bibliography

- Pierre Simon de Laplace (1812) *Analytical Theory of Probability*
The first major treatise blending calculus with probability theory, originally in French: *Théorie Analytique des Probabilités*.
- Andrei Nikolajevich Kolmogorov (1950) *Foundations of the Theory of Probability*
The modern measure-theoretic foundation of probability theory; the original German version (*Grundbegriffe der Wahrscheinlichkeitrechnung*) appeared in 1933.
- Harold Jeffreys (1939) *The Theory of Probability*
An empiricist, Bayesian approach to the foundations of probability theory.
- Edward Nelson (1987) *Radically Elementary Probability Theory*
Foundations of probability theory based on nonstandard analysis. Downloadable.
<http://www.math.princeton.edu/~nelson/books.html>
- Patrick Billingsley: *Probability and Measure*, John Wiley and Sons, New York, Toronto, London, 1979.
- Henk Tijms (2004) *Understanding Probability*
A lively introduction to probability theory for the beginner, Cambridge Univ. Press.
- David Williams (1991) *Probability with martingales*
An undergraduate introduction to measure-theoretic probability, Cambridge Univ. Press.

- Gut, Allan (2005). *Probability: A Graduate Course*. Springer. ISBN 0-387-22833-0.

External links

- Sazonov, V.V. (2001) [1994], "Probability space" (https://www.encyclopediaofmath.org/index.php?title=Probability_space), *Encyclopedia of Mathematics*, EMS Press
 - Animation (<https://www.youtube.com/watch?v=9eaOxgT5ys0>) demonstrating probability space of dice
 - Virtual Laboratories in Probability and Statistics (<http://www.math.uah.edu/stat/>) (principal author Kyle Siegrist), especially, Probability Spaces (<http://www.math.uah.edu/stat/prob>)
 - Citizendium (http://en.citizendium.org/wiki/Probability_space)
 - Complete probability space (http://www.encyclopediaofmath.org/index.php/Probability_space)
 - Weisstein, Eric W. "Probability space" (<https://mathworld.wolfram.com/ProbabilitySpace.html>). *MathWorld*.
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Probability_space&oldid=1201879276"

▪