Independence (probability theory)

Independence is a fundamental notion in probability theory, as in statistics and the theory of stochastic processes.

Two events are **independent**, **statistically independent**, or **stochastically independent**[1] if the occurrence of one does not affect the probability of occurrence of the other (equivalently, does not affect the odds). Similarly, two random variables are independent if the realization of one does not affect the probability distribution of the other.

When dealing with collections of more than two events, a weak and a strong notion of independence need to be distinguished. The events are called pairwise independent if any two events in the collection are independent of each other, while saying that the events are **mutually independent** (or **collectively independent**) intuitively means that each event is independent of any combination of other events in the collection. A similar notion exists for collections of random variables.

The name "mutual independence" (same as "collective independence") seems the outcome of a pedagogical choice, merely to distinguish the stronger notion from "pairwise independence" which is a weaker notion. In the advanced literature of probability theory, statistics, and stochastic processes, the stronger notion is simply named **independence** with no modifier. It is stronger since independence implies pairwise independence, but not the other way around.

Contents

Definition

For events

Two events

Two events A and B are **independent** (often written as $A \perp B$ or $A \perp\!\!\!\perp B$) if and only if their joint probability equals the product of their probabilities:[2]:p. 29[3]:p. 10

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B) \qquad \text{(Eq. 1)}$$

Why this defines independence is made clear by rewriting with conditional probabilities:

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

and similarly

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B) \iff \mathrm{P}(B) = \mathrm{P}(B \mid A).$$

Thus, the occurrence of B does not affect the probability of A, and vice versa. Although the derived expressions may seem more intuitive, they are not the preferred definition, as the conditional probabilities may be undefined if $\mathrm{P}(A)$ or $\mathrm{P}(B)$ are 0. Furthermore, the preferred definition makes clear by symmetry that when A is independent of B, B is also independent of A.

Log probability and information content

Stated in terms of log probability, two events are independent if and only if the log probability of the joint event is the sum of the log probability of the individual events:

$$\log \mathrm{P}(A \cap B) = \log \mathrm{P}(A) + \log \mathrm{P}(B)$$

In information theory, negative log probability is interpreted as information content, and thus two events are independent if and only if the information content of the combined event equals the sum of information content of the individual events:

$$\mathrm{I}(A \cap B) = \mathrm{I}(A) + \mathrm{I}(B)$$

See Information content § Additivity of independent events for details.

Odds

Stated in terms of odds, two events are independent if and only if the odds ratio of A and B is unity (1). Analogously with probability, this is equivalent to the conditional odds being equal to the unconditional odds:

$$O(A \mid B) = O(A) \text{ and } O(B \mid A) = O(B),$$

or to the odds of one event, given the other event, being the same as the odds of the event, given the other event not occurring:

$$O(A \mid B) = O(A \mid \neg B) \text{ and } O(B \mid A) = O(B \mid \neg A).$$

The odds ratio can be defined as

$$O(A \mid B) : O(A \mid \neg B),$$

or symmetrically for odds of B given A, and thus is 1 if and only if the events are independent.

More than two events

A finite set of events $\{A_i\}_{i=1}^{n}$ is **pairwise independent** if every pair of events is independent[4]—that is, if and only if for all distinct pairs of indices m, k,

$$\mathrm{P}(A_m \cap A_k) = \mathrm{P}(A_m)\mathrm{P}(A_k) \qquad \text{(Eq. 2)}$$

A finite set of events is **mutually independent** if every event is independent of any intersection of the other events[4][3]:p. 11—that is, if and only if for every $k \leq n$ and for every k-element subset of events $\{B_i\}_{i=1}^{k}$ of $\{A_i\}_{i=1}^{n}$,

$$\mathrm{P}\left(\bigcap_{i=1}^{k} B_i\right) = \prod_{i=1}^{k} \mathrm{P}(B_i) \qquad \text{(Eq. 3)}$$

This is called the *multiplication rule* for independent events. Note that it is not a single condition involving only the product of all the probabilities of all single events (see below for a counterexample); it must hold true for all subsets of events.

For more than two events, a mutually independent set of events is (by definition) pairwise independent; but the converse is not necessarily true (see below for a counterexample).[2]:p. 30

For real valued random variables

Two random variables

Two random variables X and Y are **independent** if and only if (iff) the elements of the π-system generated by them are independent; that is to say, for every x and y, the events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent events (as defined above in **Eq.1**). That is, X and Y with cumulative distribution functions $F_X(x)$ and $F_Y(y)$, are independent iff the combined random variable (X, Y) has a joint cumulative distribution function[3]:p. 15

$$F_{X,Y}(x, y) = F_X(x) F_Y(y) \quad \text{for all } x, y \qquad \text{(Eq. 4)}$$

or equivalently, if the probability densities $f_X(x)$ and $f_Y(y)$ and the joint probability density $f_{X,Y}(x, y)$ exist,

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) \quad \text{for all } x, y.$$

More than two random variables

A finite set of n random variables $\{X_1,\ldots,X_n\}$ is **pairwise independent** if and only if every pair of random variables is independent. Even if the set of random variables is pairwise independent, it is not necessarily mutually independent as defined next.

A finite set of n random variables $\{X_1,\ldots,X_n\}$ is **mutually independent** if and only if for any sequence of numbers $\{x_1,\ldots,x_n\}$, the events $\{X_1 \leq x_1\},\ldots,\{X_n \leq x_n\}$ are mutually independent events (as defined above in **Eq.3**). This is equivalent to the following condition on the joint cumulative distribution function $F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$. A finite set of n random variables $\{X_1,\ldots,X_n\}$ is **mutually independent** if and only if[3]:p. 16

$$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = F_{X_1}(x_1)\cdot\ldots\cdot F_{X_n}(x_n) \quad \text{for all } x_1,\ldots,x_n \qquad \text{(Eq. 5)}$$

Notice that it is not necessary here to require that the probability distribution factorizes for all possible $k-$element subsets as in the case for n events. This is not required because e.g. $F_{X_1,X_2,X_3}(x_1,x_2,x_3) = F_{X_1}(x_1)\cdot F_{X_2}(x_2)\cdot F_{X_3}(x_3)$ implies $F_{X_1,X_3}(x_1,x_3) = F_{X_1}(x_1)\cdot F_{X_3}(x_3)$.

The measure-theoretically inclined may prefer to substitute events $\{X \in A\}$ for events $\{X \leq x\}$ in the above definition, where A is any Borel set. That definition is exactly equivalent to the one above when the values of the random variables are real numbers. It has the advantage of working also for complex-valued random variables or for random variables taking values in any measurable space (which includes topological spaces endowed by appropriate σ-algebras).

For real valued random vectors

Two random vectors $\mathbf{X} = (X_1,\ldots,X_m)^T$ and $\mathbf{Y} = (Y_1,\ldots,Y_n)^T$ are called **independent** if[5]:p. 187

$$F_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = F_{\mathbf{X}}(\mathbf{x})\cdot F_{\mathbf{Y}}(\mathbf{y}) \quad \text{for all } \mathbf{x},\mathbf{y} \qquad \text{(Eq. 6)}$$

where $F_{\mathbf{X}}(\mathbf{x})$ and $F_{\mathbf{Y}}(\mathbf{y})$ denote the cumulative distribution functions of $\mathbf{X}$ and $\mathbf{Y}$ and $F_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})$ denotes their joint cumulative distribution function. Independence of $\mathbf{X}$ and $\mathbf{Y}$ is often denoted by $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$. Written component-wise, $\mathbf{X}$ and $\mathbf{Y}$ are called independent if

$$F_{X_1,\ldots,X_m,Y_1,\ldots,Y_n}(x_1,\ldots,x_m,y_1,\ldots,y_n) = F_{X_1,\ldots,X_m}(x_1,\ldots,x_m)\cdot F_{Y_1,\ldots,Y_n}(y_1,\ldots,y_n) \quad \text{for all } x_1,\ldots,x_m,y_1,\ldots,y_n.$$

For stochastic processes

For one stochastic process

The definition of independence may be extended from random vectors to a stochastic process. Thereby it is required for an independent stochastic process that the random variables obtained by sampling the process at any n times $t_1,\ldots,t_n$ are independent random variables for any n.[6]:p. 163

Formally, a stochastic process $\{X_t\}_{t\in\mathcal{T}}$ is called independent, if and only if for all $n \in \mathbb{N}$ and for all $t_1,\ldots,t_n \in \mathcal{T}$

$$F_{X_{t_1},\ldots,X_{t_n}}(x_1,\ldots,x_n) = F_{X_{t_1}}(x_1)\cdot\ldots\cdot F_{X_{t_n}}(x_n) \quad \text{for all } x_1,\ldots,x_n \qquad \text{(Eq. 7)}$$

where $F_{X_{t_1},\ldots,X_{t_n}}(x_1,\ldots,x_n) = \mathrm{P}(X(t_1) \leq x_1,\ldots,X(t_n) \leq x_n)$. Independence of a stochastic process is a property *within* a stochastic process, not between two stochastic processes.

For two stochastic processes

Independence of two stochastic processes is a property between two stochastic processes $\{X_t\}_{t\in\mathcal{T}}$ and $\{Y_t\}_{t\in\mathcal{T}}$ that are defined on the same probability space $(\Omega,\mathcal{F},P)$. Formally, two stochastic processes $\{X_t\}_{t\in\mathcal{T}}$ and $\{Y_t\}_{t\in\mathcal{T}}$ are said to be independent if for all $n \in \mathbb{N}$ and for all $t_1,\ldots,t_n \in \mathcal{T}$, the random vectors $(X(t_1),\ldots,X(t_n))$ and $(Y(t_1),\ldots,Y(t_n))$ are independent,[7]:p. 515 i.e. if

$$F_{X_{t_1},\ldots,X_{t_n},Y_{t_1},\ldots,Y_{t_n}}(x_1,\ldots,x_n,y_1,\ldots,y_n) = F_{X_{t_1},\ldots,X_{t_n}}(x_1,\ldots,x_n)\cdot F_{Y_{t_1},\ldots,Y_{t_n}}(y_1,\ldots,y_n) \quad \text{for all } x_1,\ldots,x_n$$

Independent σ-algebras

The definitions above (**Eq.1** and **Eq.2**) are both generalized by the following definition of independence for σ-algebras. Let $(\Omega,\Sigma,\mathrm{P})$ be a probability space and let $\mathcal{A}$ and $\mathcal{B}$ be two sub-σ-algebras of Σ. $\mathcal{A}$ and $\mathcal{B}$ are said to be **independent** if, whenever $A \in \mathcal{A}$ and $B \in \mathcal{B}$,

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

Likewise, a finite family of σ-algebras $(\tau_i)_{i\in I}$, where I is an index set, is said to be independent if and only if

$$\forall (A_i)_{i\in I} \in \prod\nolimits_{i\in I}\tau_i \ :\ \mathrm{P}\left(\bigcap\nolimits_{i\in I}A_i\right) = \prod\nolimits_{i\in I}\mathrm{P}(A_i)$$

and an infinite family of σ-algebras is said to be independent if all its finite subfamilies are independent.

The new definition relates to the previous ones very directly:

- Two events are independent (in the old sense) if and only if the σ-algebras that they generate are independent (in the new sense). The σ-algebra generated by an event $E \in \Sigma$ is, by definition,

$$\sigma(\{E\}) = \{\emptyset, E, \Omega\setminus E, \Omega\}.$$

- Two random variables X and Y defined over Ω are independent (in the old sense) if and only if the σ-algebras that they generate are independent (in the new sense). The σ-algebra generated by a random variable X taking values in some measurable space S consists, by definition, of all subsets of Ω of the form $X^{-1}(U)$, where U is any measurable subset of S.

Using this definition, it is easy to show that if X and Y are random variables and Y is constant, then X and Y are independent, since the σ-algebra generated by a constant random variable is the trivial σ-algebra $\{\varnothing, \Omega\}$. Probability zero events cannot affect independence so independence also holds if Y is only Pr-almost surely constant.

Properties

Self-independence

Note that an event is independent of itself if and only if

$$\mathrm{P}(A) = \mathrm{P}(A \cap A) = \mathrm{P}(A) \cdot \mathrm{P}(A) \Leftrightarrow \mathrm{P}(A) = 0 \text{ or } \mathrm{P}(A) = 1.$$

Thus an event is independent of itself if and only if it almost surely occurs or its complement almost surely occurs; this fact is useful when proving zero–one laws.[8]

Expectation and covariance

If X and Y are independent random variables, then the expectation operator E has the property

$$\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y],$$

and the covariance $\mathrm{cov}[X, Y]$ is zero, as follows from

$$\mathrm{cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y].$$

The converse does not hold: if two random variables have a covariance of 0 they still may be not independent. See uncorrelated.

Similarly for two stochastic processes $\{X_t\}_{t\in\mathcal{T}}$ and $\{Y_t\}_{t\in\mathcal{T}}$: If they are independent, then they are uncorrelated.[9]:p. 151

Characteristic function

Two random variables X and Y are independent if and only if the characteristic function of the random vector (X, Y) satisfies

$$\varphi_{(X,Y)}(t, s) = \varphi_X(t) \cdot \varphi_Y(s).$$

In particular the characteristic function of their sum is the product of their marginal characteristic functions:

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t),$$

though the reverse implication is not true. Random variables that satisfy the latter condition are called subindependent.

Examples

Rolling dice

The event of getting a 6 the first time a die is rolled and the event of getting a 6 the second time are *independent*. By contrast, the event of getting a 6 the first time a die is rolled and the event that the sum of the numbers seen on the first and second trial is 8 are *not* independent.

Drawing cards

If two cards are drawn *with* replacement from a deck of cards, the event of drawing a red card on the first trial and that of drawing a red card on the second trial are *independent*. By contrast, if two cards are drawn *without* replacement from a deck of cards, the event of drawing a red card on the first trial and that of drawing a red card on the second trial are *not* independent, because a deck that has had a red card removed has proportionately fewer red cards.

Pairwise and mutual independence

Consider the two probability spaces shown. In both cases, $\mathrm{P}(A) = \mathrm{P}(B) = 1/2$ and $\mathrm{P}(C) = 1/4$. The random variables in the first space are pairwise independent because $\mathrm{P}(A|B) = \mathrm{P}(A|C) = 1/2 = \mathrm{P}(A)$, $\mathrm{P}(B|A) = \mathrm{P}(B|C) = 1/2 = \mathrm{P}(B)$, and $\mathrm{P}(C|A) = \mathrm{P}(C|B) = 1/4 = \mathrm{P}(C)$; but the three random variables are not mutually independent. The random variables in the second space are both pairwise independent and mutually independent. To illustrate the difference, consider conditioning on two events. In the pairwise independent case, although any one event is independent of each of the other two individually, it is not independent of the intersection of the other two:

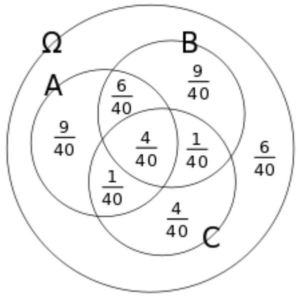
$$\mathrm{P}(A|BC) = \frac{\frac{4}{40}}{\frac{4}{40} + \frac{1}{40}} = \tfrac{4}{5} \neq \mathrm{P}(A)$$

$$\mathrm{P}(B|AC) = \frac{\frac{4}{40}}{\frac{4}{40} + \frac{1}{40}} = \tfrac{4}{5} \neq \mathrm{P}(B)$$

$$\mathrm{P}(C|AB) = \frac{\frac{4}{40}}{\frac{4}{40} + \frac{6}{40}} = \frac{2}{5} \neq \mathrm{P}(C)$$

In the mutually independent case, however,

$$\mathrm{P}(A|BC) = \frac{\frac{1}{16}}{\frac{1}{16} + \frac{1}{16}} = \frac{1}{2} = \mathrm{P}(A)$$
$$\mathrm{P}(B|AC) = \frac{\frac{1}{16}}{\frac{1}{16} + \frac{1}{16}} = \frac{1}{2} = \mathrm{P}(B)$$
$$\mathrm{P}(C|AB) = \frac{\frac{1}{16}}{\frac{1}{16} + \frac{3}{16}} = \frac{1}{4} = \mathrm{P}(C)$$



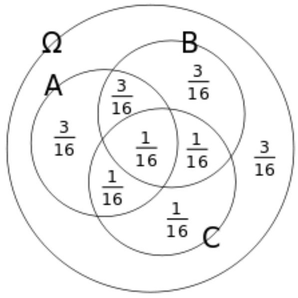
Pairwise independent, but not mutually independent, events.

Mutual independence

It is possible to create a three-event example in which

$$\mathrm{P}(A \cap B \cap C) = \mathrm{P}(A)\mathrm{P}(B)\mathrm{P}(C),$$

and yet no two of the three events are pairwise independent (and hence the set of events are not mutually independent).[10] This example shows that mutual independence involves requirements on the products of probabilities of all combinations of events, not just the single events as in this example.



Mutually independent events.

Conditional independence

For events

The events A and B are conditionally independent given an event C when

$$\mathrm{P}(A \cap B \mid C) = \mathrm{P}(A \mid C) \cdot \mathrm{P}(B \mid C).$$

For random variables

Intuitively, two random variables X and Y are conditionally independent given Z if, once Z is known, the value of Y does not add any additional information about X. For instance, two measurements X and Y of the same underlying quantity Z are not independent, but they are **conditionally independent given** Z (unless the errors in the two measurements are somehow connected).

The formal definition of conditional independence is based on the idea of conditional distributions. If X, Y, and Z are discrete random variables, then we define X and Y to be *conditionally independent given* Z if

$$\mathrm{P}(X \leq x, Y \leq y \mid Z = z) = \mathrm{P}(X \leq x \mid Z = z) \cdot \mathrm{P}(Y \leq y \mid Z = z)$$

for all x, y and z such that $\mathrm{P}(Z = z) > 0$. On the other hand, if the random variables are continuous and have a joint probability density function $f_{XYZ}(x, y, z)$, then X and Y are conditionally independent given Z if

$$f_{XY|Z}(x, y|z) = f_{X|Z}(x|z) \cdot f_{Y|Z}(y|z)$$

for all real numbers x, y and z such that $f_Z(z) > 0$.

If discrete X and Y are conditionally independent given Z, then

$$\mathrm{P}(X = x|Y = y, Z = z) = \mathrm{P}(X = x|Z = z)$$

for any x, y and z with $\mathrm{P}(Z = z) > 0$. That is, the conditional distribution for X given Y and Z is the same as that given Z alone. A similar equation holds for the conditional probability density functions in the continuous case.

Independence can be seen as a special kind of conditional independence, since probability can be seen as a kind of conditional probability given no events.

See also

- Copula (statistics)
- Independent and identically distributed random variables
- Mutually exclusive events
- Pairwise independent events
- Subindependence
- Conditional independence
- Normally distributed and uncorrelated does not imply independent
- Mean dependence

References

1. Russell, Stuart; Norvig, Peter (2002). Artificial Intelligence: A Modern Approach (https://archive.org/details/artificialintell00ru

ss_726). Prentice Hall. p. 478 (https://archive.org/details/artificialintell00russ_726/page/n506). ISBN 0-13-790395-2.
2. Florescu, Ionut (2014). Probability and Stochastic Processes. Wiley. ISBN 978-0-470-62455-5.
3. Gallager, Robert G. (2013). Stochastic Processes Theory for Applications. Cambridge University Press. ISBN 978-1-107-03975-9.
4. Feller, W (1971). "Stochastic Independence". An Introduction to Probability Theory and Its Applications. Wiley.
5. Papoulis, Athanasios (1991). Probability, Random Variables and Stochastic Porcesses. MCGraw Hill. ISBN 0-07-048477-5.
6. Hwei, Piao (1997). Theory and Problems of Probability, Random Variables, and Random Processes (https://archive.org/details/schaumsoutlineof00hsuh). McGraw-Hill. ISBN 0-07-030644-3.
7. Amos Lapidoth (8 February 2017). A Foundation in Digital Communication (https://books.google.com/books?id=6oTuDQAAQBAJ&q=independence). Cambridge University Press. ISBN 978-1-107-17732-1.
8. Durrett, Richard (1996). Probability: theory and examples (Second ed.). page 62
9. Park,Kun Il (2018). Fundamentals of Probability and Stochastic Processes with Applications to Communications. Springer. ISBN 978-3-319-68074-3.
0. George, Glyn, "Testing for the independence of three events," Mathematical Gazette 88, November 2004, 568. PDF (http://www.engr.mun.ca/~ggeorge/MathGaz04.pdf)

External links

- Media related to Statistical dependence at Wikimedia Commons