WIKIPEDIA

Conditional probability

In probability theory, **conditional probability** is a measure of the probability of an event given that (by assumption, presumption, assertion or evidence) another event has occurred.[1] If the event of interest is A and the event B is known or assumed to have occurred, "the conditional probability of A given B", or "the probability of A under the condition B", is usually written as $P(A|B)$, or sometimes $P_B(A)$. For example, the probability that any given person has a cough on any given day may be only 5%. But if we know or assume that the person has a cold, then they are much more likely to be coughing. The conditional probability of coughing given that you have a cold might be a much higher 75%.

The concept of conditional probability is one of the most fundamental and one of the most important concepts in probability theory.[2] But conditional probabilities can be quite slippery and require careful interpretation.[3] For example, there need not be a causal or temporal relationship between A and B.

$P(A|B)$ may or may not be equal to $P(A)$ (the unconditional probability of A). If $P(A|B) = P(A)$, then events A and B are said to be independent: in such a case, having knowledge about either event does not change our knowledge about the other event. Also, in general, $P(A|B)$ (the conditional probability of A given B) is not equal to $P(B|A)$. For example, if you have dengue you might have a 90% chance of testing positive for dengue. In this case what is being measured is that if event B ("having dengue") has occurred, the probability of A (*test is positive*) given that B (*having dengue*) occurred is 90%: that is, $P(A|B) = 90\%$. Alternatively, if you test positive for dengue you may have only a 15% chance of actually having dengue because most people do not have dengue and the false positive rate for the test may be high. In this case what is being measured is the probability of the event B (*having dengue*) given that the event A (*test is positive*) has occurred: $P(B|A) = 15\%$. Falsely equating the two probabilities causes various errors of reasoning such as the base rate fallacy. Conditional probabilities can be correctly reversed using Bayes' theorem.

Contents

Definition

Conditioning on an event

Kolmogorov definition

Given two events A and B, from the sigma-field of a probability space, with $P(B) > 0$, the conditional probability of A given B is defined as the quotient of the probability of the joint of events A and B, and the probability of B[4]:

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

This may be visualized as restricting the sample space to B. The logic behind this equation is that if the outcomes are restricted to B, this set serves as the new sample space.

Note that this is a definition but not a theoretical result. We just denote the quantity $\frac{P(A \cap B)}{P(B)}$ as $P(A|B)$ and call it the conditional probability of A given B.

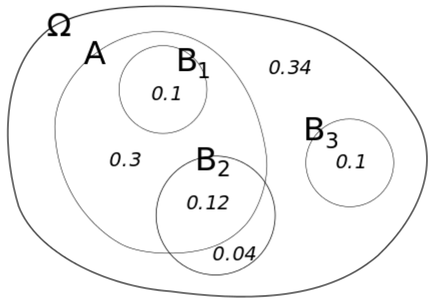
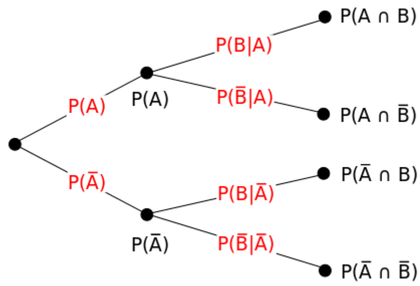


Illustration of conditional probabilities with an Euler diagram. The unconditional probability $P(A)$ = 0.52. However, the conditional probability $P(A|B_1)$ = 1, $P(A|B_2)$ = 0.75, and $P(A|B_3)$ = 0.

As an axiom of probability

Some authors, such as de Finetti, prefer to introduce conditional probability as an axiom of probability:

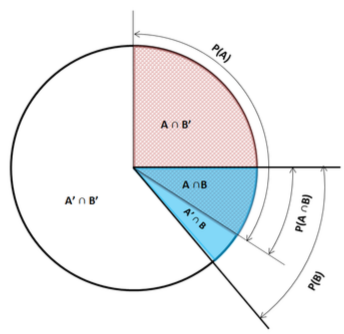
$$P(A \cap B) = P(A|B)P(B)$$



On a tree diagram, branch probabilities are conditional on the event associated with the parent node.

Although mathematically equivalent, this may be preferred philosophically; under major probability interpretations such as the subjective theory, conditional probability is considered a primitive entity. Further, this "multiplication axiom" introduces a symmetry with the summation axiom for mutually exclusive events:[5]

$$P(A \cup B) = P(A) + P(B) - \cancelto{0}{P(A \cap B)}$$



Venn Pie Chart describing conditional probabilities

As the probability of a conditional event

Conditional probability can be defined as the probability of a conditional event A_B[6]. Assuming that the experiment underlying the events A and B is repeated, the Goodman-Nguyen-van Fraassen conditional event can be defined as

$$A_B = \bigcup_{i \geq 1} \left(\bigcap_{j<i} \overline{B_j}, A_i B_i \right)$$

It can be shown that

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

which meets the Kolmogorov definition of conditional probability. Note that the equation $P(A_B) = P(A \cap B)/P(B)$ is a theoretical result and not a definition. The definition via conditional events can be understood directly in terms of the Kolmogorov axioms and is particularly close to the Kolmogorov interpretation of probability in terms of experimental data. For example, conditional events can be repeated themselves leading to a generalized notion of conditional event $A_{B(n)}$. It can be shown[6] that the sequence $(A_{B(n)})_{n \geq 1}$ is i.i.d., which yields a strong law of large numbers for conditional probability:

$$P(\lim_{n \longrightarrow \infty} \overline{A_B^n} = P(A|B)) = 100\%$$

Measure-theoretic definition

If *P*(*B*) = 0, then according to the simple definition, *P*(*A*|*B*) is undefined. However, it is possible to define a conditional probability with respect to a σ-algebra of such events (such as those arising from a continuous random variable).

For example, if *X* and *Y* are non-degenerate and jointly continuous random variables with density $f_{X,Y}(x, y)$ then, if *B* has positive measure,

$$P(X \in A \mid Y \in B) = \frac{\int_{y \in B} \int_{x \in A} f_{X,Y}(x,y)\, dx\, dy}{\int_{y \in B} \int_{x \in \mathbb{R}} f_{X,Y}(x,y)\, dx\, dy}.$$

The case where *B* has zero measure is problematic. For the case that $B = \{y_0\}$, representing a single point, the conditional probability could be defined as

$$P(X \in A \mid Y = y_0) = \frac{\int_{x \in A} f_{X,Y}(x, y_0)\, dx}{\int_{x \in \mathbb{R}} f_{X,Y}(x, y_0)\, dx},$$

however this approach leads to the Borel–Kolmogorov paradox. The more general case of zero measure is even more problematic, as can be seen by noting that the limit, as all δy_i approach zero, of

$$P(X \in A \mid Y \in \cup_i [y_i, y_i + \delta y_i]) \approxeq \frac{\sum_i \int_{x \in A} f_{X,Y}(x, y_i)\, dx\, \delta y_i}{\sum_i \int_{x \in \mathbb{R}} f_{X,Y}(x, y_i)\, dx\, \delta y_i},$$

depends on their relationship as they approach zero. See conditional expectation for more information.

Conditioning on a random variable

Conditioning on an event may be generalized to conditioning on a random variable. Let *X* be a random variable; we assume for the sake of presentation that *X* is discrete, that is, *X* takes on only finitely many values *x*. Let *A* be an event. The conditional probability of *A* given *X* is defined as the random variable, written *P*(*A*|*X*)*, that takes on the value*

$$P(A \mid X = x)$$

whenever

$$X = x.$$

More formally,

$$P(A|X)(\omega) = P(A \mid X = X(\omega)).$$

The conditional probability *P*(*A*|*X*) is a function of *X*: e.g., if the function *g* is defined as

$$g(x) = P(A \mid X = x),$$

then

$$P(A|X) = g \circ X.$$

Note that $P(A|X)$ and X are now both random variables. From the law of total probability, the expected value of $P(A|X)$ is equal to the unconditional probability of A.

Partial conditional probability

The partial conditional probability $P(A|B_1 \equiv b_1 \cdots B_m \equiv b_m)$ is about the probability of event A given that each of the condition events B_i has occurred to a degree b_i (degree of belief, degree of experience) that might be different from 100%. Frequentistically, partial conditional probability makes sense, if the conditions are tested in experiment repetitions of appropriate length n [7]. Such n-bounded partial conditional probability can be defined as the conditionally expected average occurrence of event A in testbeds of length n that adhere to all of the probability specifications $B_i \equiv b_i$, i.e.:

$$P^n(A|B_1 \equiv b_1 \cdots B_m \equiv b_m) = E(\overline{A^n}|\overline{B_1^n} = b_1 \cdots \overline{B_m^n} = b_m)$$ [7]

Based on that, partial conditional probability can be defined as

$$P(A|B_1 \equiv b_1 \cdots B_m \equiv b_m) = \lim_{n \longrightarrow \infty} P^n(A|B_1 \equiv b_1 \cdots B_m \equiv b_m),$$

where $b_i n \in \mathbb{N}$ [7]

Jeffrey conditionalization [8] [9] is a special case of partial conditional probability in which the condition events must form a partition:

$$P(A|B_1 \equiv b_1 \cdots B_m \equiv b_m) = \sum_{i=1}^{m} b_i P(A|B_i)$$

Example

Suppose that somebody secretly rolls two fair six-sided dice, and we must predict the outcome (the sum of the two upward faces).

- Let *D1* be the value rolled on die 1.
- Let *D2* be the value rolled on die 2.

What is the probability that* D1 *= 2?

Table 1 shows the sample space of 36 outcomes.

Clearly, *D1* = 2 in exactly 6 of the 36 outcomes; thus $P(D1{=}2) = {}^{6}\!/_{36} = {}^{1}\!/_{6}$.

Table 1

+		D2					
		1	**2**	**3**	**4**	**5**	**6**
D1	**1**	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

What is the probability that* D1+D2 ≤ *5 ?

Table 2 shows that $D1+D2 \leq 5$ for exactly 10 of the same 36 outcomes, thus $P(D1+D2 \leq 5) = {}^{10}\!/_{36}$.

Table 2

+		D2					
		1	**2**	**3**	**4**	**5**	**6**
D1	**1**	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

What is the probability that* D1 = *2* given that D1+D2 ≤ *5 ?

Table 3 shows that for 3 of these 10 outcomes, $D1 = 2$.

Thus, the conditional probability $P(D1=2 \mid D1+D2 \leq 5) = {}^{3}\!/_{10} = 0.3$.

Table 3

+		D2					
		1	**2**	**3**	**4**	**5**	**6**
D1	**1**	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Here in the earlier notation for the definition of conditional probability, the conditioning event B is that $D1+D2 \leq 5$, and the event A is $D1 = 2$. We have $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{3/36}{10/36} = \frac{3}{10}$, as seen in the table.

Use in inference

In statistical inference, the conditional probability is an update of the probability of an event based on new information.[3] Incorporating the new information can be done as follows:[1]

- Let A, the event of interest, be in the sample space, say (X,P).
- The occurrence of the event A knowing that event B has or will have occurred, means the occurrence of A as it is restricted to B, i.e. $A \cap B$.
- Without the knowledge of the occurrence of B, the information about the occurrence of A would simply be $P(A)$
- The probability of A knowing that event B has or will have occurred, will be the probability of $A \cap B$ relative to $P(B)$, the probability that B has occurred.
- This results in $P(A|B) = P(A \cap B)/P(B)$ whenever $P(B)>0$ and 0 otherwise.

Note: This approach results in a probability measure that is consistent with the original probability measure and satisfies all the Kolmogorov axioms. This conditional probability measure also could have resulted by assuming that the relative magnitude of the probability of A with respect to X will be preserved with respect to B (cf. a Formal Derivation below).

The wording "evidence" or "information" is generally used in the Bayesian interpretation of probability. The conditioning event is interpreted as evidence for the conditioned event. That is, $P(A)$ is the probability of A before accounting for evidence E, and $P(A|E)$ is the probability of A after having accounted for evidence E or after having updated $P(A)$. This is consistent with the frequentist interpretation, which is the first definition given above.

Statistical independence

Events A and B are defined to be statistically independent if

$$P(A \cap B) = P(A)P(B).$$

If $P(B)$ is not zero, then this is equivalent to the statement that

$$P(A|B) = P(A).$$

Similarly, if $P(A)$ is not zero, then

$$P(B|A) = P(B)$$

is also equivalent. Although the derived forms may seem more intuitive, they are not the preferred definition as the conditional probabilities may be undefined, and the preferred definition is symmetrical in A and B.

Independent events vs. mutually exclusive events

The concepts of mutually independent events and mutually exclusive events are separate and distinct.

As noted, statistical independence means

$$P(A|B) = P(A) \quad \text{and} \quad P(B|A) = P(B)$$

provided that the probability for the conditioning event is not zero. However, events being mutually exclusive means that

$$P(A|B) = 0, \quad P(B|A) = 0, \quad \text{and} \quad P(A \cap B) = 0.$$

In fact, mutually exclusive events cannot be statistically independent, since knowing that one occurs gives information about the other (specifically, that it certainly does not occur).

Common fallacies

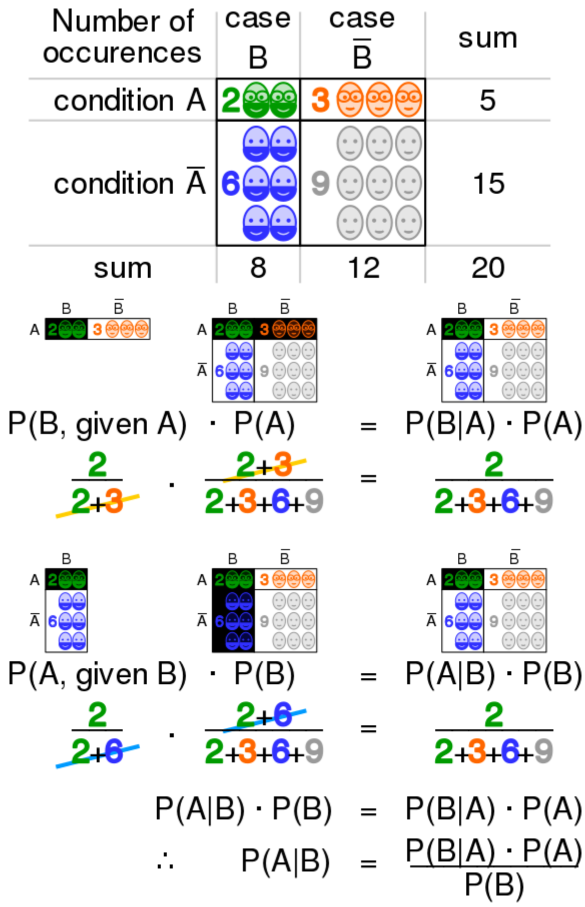
These fallacies should not be confused with Robert K. Shope's 1978 "conditional fallacy" (http://lesswrong.com/r/discussion/lw/9om/the_conditional_fallacy_in_contemporary_philosophy/), which deals with counterfactual examples that beg the question.

Assuming conditional probability is of similar size to its inverse

In general, it cannot be assumed that $P(A|B) \approx P(B|A)$. This can be an insidious error, even for those who are highly conversant with statistics.[10] The relationship between $P(A|B)$ and $P(B|A)$ is given by Bayes' theorem:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$
$$\Leftrightarrow \frac{P(B|A)}{P(A|B)} = \frac{P(B)}{P(A)}$$

That is, $P(A|B) \approx P(B|A)$ only if $P(B)/P(A) \approx 1$, or equivalently, $P(A) \approx P(B)$.



A geometric visualisation of Bayes' theorem. In the table, the values 2, 3, 6 and 9 give the relative weights of each corresponding condition and case. The figures denote the cells of the table involved in each metric, the probability being the fraction of each figure that is shaded. This shows that P(A|B) P(B) = P(B|A) P(A) i.e. P(A|B) = $\frac{P(B|A)\,P(A)}{P(B)}$. Similar reasoning can be used to show that P(Ā|B) = $\frac{P(B|\bar{A})\,P(\bar{A})}{P(B)}$ etc.

Assuming marginal and conditional probabilities are of similar size

In general, it cannot be assumed that $P(A) \approx P(A|B)$. These probabilities are linked through the law of total probability:

$$P(A) = \sum_n P(A \cap B_n) = \sum_n P(A|B_n)P(B_n).$$

where the events (B_n) form a countable partition of Ω.

This fallacy may arise through selection bias.[11] For example, in the context of a medical claim, let S_C be the event that a sequela (chronic disease) S occurs as a consequence of circumstance (acute condition) C. Let H be the event that an individual seeks medical help. Suppose that in most cases, C does not cause S so $P(S_C)$ is low. Suppose also that medical attention is only sought if S has occurred due to C. From experience of patients, a doctor may therefore erroneously conclude that $P(S_C)$ is high. The actual probability observed by the doctor is $P(S_C|H)$.

Over- or under-weighting priors

Not taking prior probability into account partially or completely is called *base rate neglect*. The reverse, insufficient adjustment from the prior probability is *conservatism*.

Formal derivation

Formally, $P(A|B)$ is defined as the probability of A according to a new probability function on the sample space, such that outcomes not in B have probability 0 and that it is consistent with all original probability measures.[12][13]

Let Ω be a sample space with elementary events $\{\omega\}$. Suppose we are told the event $B \subseteq \Omega$ has occurred. A new probability distribution (denoted by the conditional notation) is to be assigned on $\{\omega\}$ to reflect this. For events in B, it is reasonable to assume that the relative magnitudes of the probabilities will be preserved. For some constant scale factor α, the new distribution will therefore satisfy:

1. $\omega \in B : P(\omega|B) = \alpha P(\omega)$
2. $\omega \notin B : P(\omega|B) = 0$
3. $\sum_{\omega \in \Omega} P(\omega|B) = 1.$

Substituting 1 and 2 into 3 to select α:

$$\begin{aligned} 1 &= \sum_{\omega \in \Omega} P(\omega|B) \\ &= \sum_{\omega \in B} P(\omega|B) + \cancelto{0}{\sum_{\omega \notin B} P(\omega|B)} \\ &= \alpha \sum_{\omega \in B} P(\omega) \\ &= \alpha \cdot P(B) \\ \Rightarrow \alpha &= \frac{1}{P(B)} \end{aligned}$$

So the new probability distribution is

1. $\omega \in B : P(\omega|B) = \frac{P(\omega)}{P(B)}$
2. $\omega \notin B : P(\omega|B) = 0$

Now for a general event A,

$$\begin{aligned} P(A|B) &= \sum_{\omega \in A \cap B} P(\omega|B) + \cancelto{0}{\sum_{\omega \in A \cap B^c} P(\omega|B)} \\ &= \sum_{\omega \in A \cap B} \frac{P(\omega)}{P(B)} \\ &= \frac{P(A \cap B)}{P(B)} \end{aligned}$$

See also

- Borel–Kolmogorov paradox
- Chain rule (probability)

- Class membership probabilities
- Conditional probability distribution
- Conditioning (probability)
- Joint probability distribution
- Monty Hall problem
- Posterior probability

References

1. Gut, Allan (2013). *Probability: A Graduate Course* (Second ed.). New York, NY: Springer. ISBN 978-1-4614-4707-8.
2. Ross, Sheldon (2010). *A First Course in Probability* (8th ed.). Pearson Prentice Hall. ISBN 978-0-13-603313-4.
3. Casella, George; Berger, Roger L. (2002). *Statistical Inference*. Duxbury Press. ISBN 0-534-24312-6.
4. Kolmogorov, Andrey (1956), *Foundations of the Theory of Probability*, Chelsea
5. Gillies, Donald (2000); "Philosophical Theories of Probability"; Routledge; Chapter 4 "The subjective theory"
6. Draheim, Dirk (2017). "An Operational Semantics of Conditional Probabilities that Fully Adheres to Kolmogorov's Explication of Probability Theory". doi:10.13140/RG.2.2.10050.48323/3 (https://doi.org/10.13140%2FRG.2.2.10050.48323%2F3).
7. Draheim, Dirk (2017). "Generalized Jeffrey Conditionalization (A Frequentist Semantics of Partial Conditionalization)" (http://fpc.formcharts.org). Springer. Retrieved December 19, 2017.
8. Jeffrey, Richard C. (1983), *The Logic of Decision, 2nd edition*, University of Chicago Press
9. "Bayesian Epistemology" (https://plato.stanford.edu/entries/epistemology-bayesian/). Stanford Encyclopedia of Philosophy. 2017. Retrieved December 29, 2017.
10. Paulos, J.A. (1988) *Innumeracy: Mathematical Illiteracy and its Consequences*, Hill and Wang. ISBN 0-8090-7447-8 (p. 63 *et seq.*)
11. Thomas Bruss, F; Der Wyatt Earp Effekt; Spektrum der Wissenschaft; March 2007
12. George Casella and Roger L. Berger (1990), *Statistical Inference*, Duxbury Press, ISBN 0-534-11958-1 (p. 18 *et seq.*)
13. Grinstead and Snell's Introduction to Probability (http://math.dartmouth.edu/~prob/prob/prob.pdf), p. 134

External links

- Weisstein, Eric W. "Conditional Probability" (http://mathworld.wolfram.com/ConditionalProbability.html). *MathWorld*.
- F. Thomas Bruss Der Wyatt-Earp-Effekt oder die betörende Macht kleiner Wahrscheinlichkeiten (in German), Spektrum der Wissenschaft (German Edition of Scientific American), Vol 2, 110–113, (2007).
- Visual explanation of conditional probability (http://setosa.io/conditional/)