



## Could someone explain conditional independence?

Asked 9 years, 1 month ago   Active 1 year, 1 month ago   Viewed 52k times



90



86



My understanding right now is that an example of conditional independence would be:

If two people live in the same city, the probability that person **A** gets home in time for dinner, and the probability that person **B** gets home in time for dinner are independent; that is, we wouldn't expect one to have an affect on the other. But if a snow storm hits the city and introduces a probability **C** that traffic will be at a stand still, you would expect that the probability of both **A** getting home in time for dinner and **B** getting home in time for dinner, would change.

If this is a correct understanding, I guess I still don't understand what exactly conditional independence *is*, or what it does for us (why does it have a separate name, as opposed to just compounded probabilities), and if this isn't a correct understanding, could someone please provide an example with an explanation?

probability

independence

edited Nov 12 '15 at 14:43



BCLC

1

asked Feb 21 '11 at 18:06



Ryan

1,291

2

10

17

### 5 Answers

Active

Oldest

Votes



102



The scenario you describe provides a good example for conditional independence, though you haven't quite described it as such. As [the Wikipedia article](#) puts it,

$R$  and  $B$  are conditionally independent [given  $Y$ ] if and only if, given knowledge of whether  $Y$  occurs, knowledge of whether  $R$  occurs provides no information on the likelihood of  $B$  occurring, and knowledge of whether  $B$  occurs provides no information on the likelihood of  $R$  occurring.



In this case,  $R$  and  $B$  are the events of persons **A** and **B** getting home in time for dinner, and  $Y$  is the event of a snow storm hitting the city. Certainly the probabilities of  $R$  and  $B$  will depend on whether  $Y$  occurs. However, just as it's plausible to assume that if these two people have nothing to do with each other their probabilities of getting home in time are independent, it's also plausible to assume that, while they will both have a lower probability of getting home in time if a snow storm hits, these lower probabilities will nevertheless still be independent of each other. That is, if you already know that a snow storm is raging and I tell you that person **A** is getting home late, that gives you no new information about whether person **B** is getting home late. You're getting information on that from the fact that there's a snow storm, but given that fact, the fact that **A** is getting home late doesn't make it more or less likely that **B** is getting home late, too. So

conditional independence is the same as normal independence, but restricted to the case where you know that a certain condition is or isn't fulfilled. Not only can you not find out about **A** by finding out about **B** in general (normal independence), but you also can't do so under the condition that there's a snow storm (conditional independence).

An example of events that are independent but not conditionally independent would be: You randomly sample two people **A** and **B** from a large population and consider the probabilities that they will get home in time. Without any further knowledge, you might plausibly assume that these probabilities are independent. Now you introduce event  $Y$ , which occurs if the two people live in the same neighbourhood (however that might be defined). If you know that  $Y$  occurred and I tell you that **A** is getting home late, then that would tend to increase the probability that **B** is also getting home late, since they live in the same neighbourhood and any traffic-related causes of **A** getting home late might also delay **B**. So in this case the probabilities of **A** and **B** getting home in time are not conditionally independent given  $Y$ , since once you know that  $Y$  occurred, you are able to gain information about the probability of **B** getting home in time by finding out whether **A** is getting home in time.

Strictly speaking, this scenario only works if there's always the same amount of traffic delay in the city overall and it just moves to different neighbourhoods. If that's not the case, then it wouldn't be correct to assume independence between the two probabilities, since the fact that one of the two is getting home late would already make it somewhat likelier that there's heavy traffic in the city in general, even without knowing that they live in the same neighbourhood.

To give a precise example: Say you roll a blue die and a red die. The two results are independent of each other. Now you tell me that the blue result isn't a 6 and the red result isn't a 1. You've given me new information, but that hasn't affected the independence of the results. By taking a look at the blue die, I can't gain any knowledge about the red die; after I look at the blue die I will still have a probability of  $1/5$  for each number on the red die except 1. So the probabilities for the results are conditionally independent given the information you've given me. But if instead you tell me that the sum of the two results is even, this allows me to learn a lot about the red die by looking at the blue die. For instance, if I see a 3 on the blue die, the red die can only be 1, 3 or 5. So in this case the probabilities for the results are not conditionally independent given this other information that you've given me. This also underscores that conditional independence is always relative to the given condition -- in this case, the results of the dice rolls are conditionally independent with respect to the event "the blue result is not 6 and the red result is not 1", but they're not conditionally independent with respect to the event "the sum of the results is even".

edited Mar 8 '18 at 10:40



Georgy  
105 4

answered Feb 21 '11 at 18:26



joriki  
196k 11 215 406

21 This is an excellent answer and very helpful! Thank you! – Ryan Feb 21 '11 at 19:23

I don't understand how the first two examples are qualitatively different, or to be more precise, how **A** is conditionally independent of **B** given  $Y$  in the snowstorm case. If I know that a snow storm has occurred, and then I know that **A** did not make it home on time, am I not likely to increase the probability that **B** does not make it home on time, since clearly the storm is having an effect on everyone in the city. And vice-versa if **A** did make it home on time, does not the probability that **B** make it home on time increase? – zenna Oct 4 '12 at 5:45

1 @zenna What if **B** works and lives in the same building? – user13888 Nov 12 '13 at 13:48

Thank you for explicitly clarifying that: "This also underscores that conditional independence is always relative to the given condition"! That was very helpful! – Diego Sep 7 '16 at 17:32

Hi joriki, your reply is excellent and I vote up. But I disagree with the example of "a blue die and a red die", under the condition of knowing their sum is even, and if we know value of one die, we know there are only 3 (other than 6) possible values of the other die, it is 100% correct, but how does it links to independence of two die? I think it (know sum is even) does not break the assumption of two die are independent (for independent I mean the two die makes choices of their values independently)? Maybe we have different definition of what means independent here? – Lin Ma Sep 21 '16 at 6:18



10



The example you've given (the snowstorm) is usually given as a case where you might *think* two events might be truly independent (since they take totally different routes home), i.e.

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

However in this case they are not truly independent, they are "only" conditionally independent given the snowstorm i.e.

$$p(A|B, Z) = p(A|Z).$$

A clearer example paraphrased from [Norman Fenton's website](#): if Alice (A) and Bob (B) both flip the same coin, but that coin might be biased, *we cannot say*

$$p(A = H|B = H) = p(A = H)$$

(i.e. that they are independent) because if we see Bob flips heads, it is more likely to be biased towards heads, and hence the left probability should be higher. However if we denote Z as the event "the coin is biased towards heads", then

$$p(A = H|B = H, Z) = p(A = H|Z)$$

we can remove Bob from the equation because we know the coin is biased. Given the fact that the coin is biased, the two flips are conditionally independent.

This is the common form of conditional independence, you have events that are not statistically independent, but they are conditionally independent.

It is possible for something to be statistically independent and not conditionally independent. To borrow from [Wikipedia](#): if  $A$  and  $B$  both take the value 0 or 1 with 0.5 probability, and  $C$  denotes the product of the values of  $A$  and  $B$  ( $C = A \times B$ ), then  $A$  and  $B$  are independent:

$$p(A = 0|B = 0) = p(A = 0) = 0.5$$

but they are not conditionally independent given  $C$ :

$$p(A = 0|B = 0, C = 0) = 0.5 \neq \frac{2}{3} = p(A = 0|C = 0)$$

edited Mar 9 '16 at 10:55

answered Mar 8 '16 at 9:41



Andrew Chinery

111 1 3

- 2 Hi Andrew. Welcome to Math.SE. Thanks for you answer. This question was asked and answered 4 years ago. If you click on Questions above you can try your hand at answering some newer questions or unanswered ones. – Ian Miller Mar 8 '16 at 9:59

- 7 Hi Ian, thank you. I found this question through a Google search, and thought the answers were hard to follow, and more importantly there was not a single example of the actual statistical definitions on the page. If I found it through a Google search, others will too, so I felt like it was worth adding just in case they did. Not sure why that warrants getting voted down as I believe it's standard practice on most stackexchange sites, but there you go! – [Andrew Chinery](#) Mar 8 '16 at 18:59

I'm not sure why you got a down vote either. I just noticed your post as it was listed in the First Posts section for Math.SE and wasn't sure why you'd chosen to reply to such an old post but your explanation makes sense. I guess people understanding of the question and answers will vary based on their particular backgrounds so additional answers are always good. Thanks. (I've given you a +1 so you don't feel discouraged here at Math.SE and because I don't see anything wrong with your answer.) Maybe remove the first part referring to a previous answer as it and the comment to it are now gone. – [Ian Miller](#) Mar 9 '16 at 4:54

Thanks Ian! I will keep an eye out for newer questions on the off chance I can help :). – [Andrew Chinery](#) Mar 9 '16 at 10:56



Other answers have provided great responses elaborating on the intuitive meaning of conditional dependence. Here, I won't add to that; instead I want to address your question about "what it does for us," focusing on computational implications.



There are three events/propositions/random variables in play,  $A$ ,  $B$ , and  $C$ . They have a joint probability,  $P(A, B, C)$ . In general, a joint probability for three events can be factored in many different ways:

$$\begin{aligned}
 P(A, B, C) &= P(A)P(B, C|A) \\
 &= P(A)P(B|A)P(C|A, B) = P(A)P(C|A)P(B|A, C) \\
 &= P(B)P(A, C|B) \\
 &= P(B)P(A|B)P(C|A, B) = P(B)P(C|B)P(A|B, C) \\
 &= P(C)P(A, B|C) \\
 &= P(C)P(A|C)P(B|A, C) = P(C)P(B|C)P(A|B, C)
 \end{aligned}$$

Something to notice here is that *every expression on the RHS includes a factor with three variables*.

Now suppose our information about the problem tells us that  $A$  and  $B$  are **conditionally independent given**  $C$ . A conventional notation for this is:

$$A \perp\!\!\!\perp B \mid C,$$

which means (among other implications),

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

This means that the last of the many expressions I displayed for  $P(A, B, C)$  above can be written,

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

From a computational perspective, the key thing to note is that conditional dependence here means **we can write the 3-variable function  $P(A, B, C)$  in terms of 1-variable and 2-variable functions**. In a nutshell, conditional independence means that joint distributions are simpler than they might have been. When there are *lots* of variables, conditional independence

can imply *grand* simplifications of joint probabilities. And if (as is often the case) you have to sum or integrate over some of the variables, conditional independence can let you pull some factors through a sum/integral, simplifying the summand/integrand.

This can be very important for computational implementation of Bayesian inference. When we want to quantify how strongly some observed data,  $D$ , support rival hypotheses  $H_i$  (with  $i$  a label distinguishing the hypotheses), you are probably used to seeing Bayes's theorem (BT) in its "posterior  $\propto$  prior times likelihood" form:

$$P(H_i|D) = \frac{P(H_i)P(D|H_i)}{P(D)},$$

where the terms in the numerator are the prior probability for  $H_i$  and the sampling (or conditional predictive) probability for  $D$  (aka, the likelihood for  $H_i$ ), and the term in the denominator is the prior predictive probability for  $D$  (aka the marginal likelihood, since it is the marginal of  $P(D, H_i)$ ). But recall that  $P(H_i, D) = P(H_i)P(D|H_i)$  (in fact, one typically derives BT using this, and equating it to the alternative factorization). So BT can be written as

$$P(H_i|D) = \frac{P(H_i, D)}{P(D)},$$

or, in words,

$$\text{Posterior} = \frac{\text{Joint for everything}}{\text{Marginal for observations}}.$$

In models with complex dependence structures, this turns out to be the easiest way to think of modeling: The modeler expresses the joint probability for the data and all hypotheses (possibly including latent parameters for things you don't know but need to know in order to predict the data). From the joint, you compute the marginal for the data, to normalize the joint to give you the posterior (you may not even need to do this, e.g., if you use MCMC methods that don't depend on normalization constants).

Now you can see the value of conditional independence. Since the starting point of computation is the joint for everything, anything you can do to simplify the expression for the joint (and its sums/integrals) can be a great help to computation. Probabilistic programming languages (e.g., BUGS, JAGS, and to some degree Stan) use graphical representations of conditional dependence assumptions to organize and simplify computations.

edited Feb 23 '19 at 17:41

answered Mar 22 '18 at 19:52



Tom Lored

141 1 4

---

For me @joriki's (accepted) answer gives an excellent general overview, but for the practical understanding and application, this answer is also excellent! – Victoria Stuart Feb 7 '19 at 22:25

---

@pyrrhic Thanks for the careful eye, and fixing the typo! – Tom Lored Feb 23 '19 at 17:41

---



I will compass here to exemplify a case not yet discussed: an example of **dependence but conditional independence**. Please let me know of improvements or errors.

2

In keeping with your OP, where you defined person  $A$  and person  $B$  to live in the same city, let

$Pr(A)$  = probability that person A gets home in time for dinner and

$Pr(B)$  = the probability that person B gets home in time for dinner.



Here, I assume that persons  $A$  &  $B$  live in the same neighbourhood.

Let  $E$  be the event of a mandatory city-wide evacuation due to the menace of a falling satellite.

In line with user joriki's 2nd paragraph,  $A$  and  $B$  (the events, not their probabilities) are conceivably dependent.

However, given/on the condition of  $E$ , neither person  $A$  nor person  $B$  will probably be coming home tonight. If the satellite crashes before dinnertime, then mayhap. If it still hasn't crashed, then probably not. In both cases, any information on person A's time of return reveals nothing about person B's. In other words,  $Pr(A|E)$  and  $Pr(B|E)$  will be conditionally independent.

edited Oct 16 '18 at 10:52



amWhy

197k

29

240

453

answered Nov 12 '13 at 13:41



Greek - Area 51

Proposal

6,199

8

73

118

- 
- 4 The assertions that " $Pr(A)$  and  $Pr(B)$  are conceivably dependent" and that " $Pr(A|E)$  and  $Pr(B|E)$  will be conditionally independent" have no meaning I would be aware of. For example,  $Pr(A)$  and  $Pr(B)$  are **numbers**, as such they are neither dependent nor independent. – Did Nov 12 '13 at 13:55
- 

0

The events would be conditionally independent, for example, if A always comes home by car and B commutes by subway/metro train. Assuming subway is not affected by a snow storm but road transport is, it can be said that they are conditionally independent.

answered Nov 9 '13 at 5:17



kumar k

33

3

