Appendix: confirmation measures and the difficulty about conjunction

Marcello Di Bello and Rafal Urbaniak

One assumption often made in the formulation of the conjunction paradox is that claims *A* and *B* are probabilistically independent. This is not always the case—we have seen that the paradox does subside even if the two claims are dependent. However, two fairly natural set-ups for conjunctive hypotheses and evidence supporting them (Figure 1) indeed do have some independencies built in, and so it is also natural what can be said about the conjunction problem given these independencies. Moreover, in some context, we will be freely using the independence assumptions, as a counterexample to a general claim still remains one even if it satisfies an additional requirement, that is, independence conditions.

We will be considering the conjunction of two hypotheses, *A* and *B*, their respective pieces of evidence *a* and *b*, and their conjunction *AB*, in two set-ups, illustrated by the Bayesian networks shown in Figure 1. The key difference here is that we allow direct dependence between the hypotheses in the second network. In both Bayesian networks, the CPT for the conjunction trivially mirrors the one for conjunction, as in Table 1.
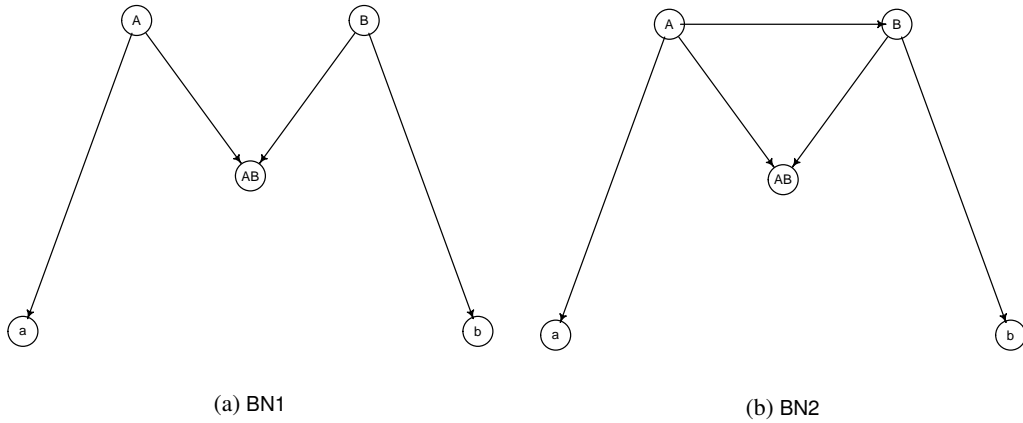

fig:conjunctionBNs}

(a) BN1

(b) BN2

Figure 1: Two Bayesian networks for the conjunction problem.

	A	B	
AB			Pr
1	1	1	1
0	1	1	0
1	0	1	0
0	0	1	1
1	1	0	0
0	1	0	1
1	0	0	0
0	0	0	1

ab:CPTconjunction2}

Table 1: Conditional probability table for the conjunction node.

Directed Acyclic Graphs (DAGs) are useful for representing graphically these relationships of independence. The edges, intuitively, are meant to capture direct influence between the nodes. The role that such direct influence plays is that in a Bayesian network built over a DAG any node is conditionally independent of its nondescendants (including ancestors), given its parents. If this is the case for a given probabilistic measure $P()$ and a given DAG, we say that $P()$ is compatible with G, and they can be put together to constitute a Bayesian network.

The graphical counterpart of probabilistic independence is the so-called **d-separation**, $\perp\!\!\!\perp_d$. We say that two nodes, X and Y, are d-separated given a set of nodes Z—$X \perp\!\!\!\perp_d Y | Z$ — iff for every undirected path from X to Y there is a node Z' on the path such that either:

- $Z' \in Z$ and there is a *serial* connection, $\rightarrow Z' \rightarrow$, on the path,
- $Z' \in Z$ and there is a diverging connection, $\leftarrow Z' \rightarrow$, on the path,
- There is a connection $\rightarrow Z' \leftarrow$ on the path, and neither Z' nor its descendants are in Z.

Finally, two sets of nodes, X and Y, are d-separated given Z if every node in X is d-separated from every node in Y given Z. With serial connection, for instance, if:

Node	Proposition
G	The suspect is guilty.
B	The blood stain comes from the suspect.
M	The crime scene stain and the suspect's blood share their DNA profile.

We naturally would like to have the connection $G \rightarrow B \rightarrow M$. If we don't know whether B holds, G seems to have an indirect impact on the probability of M. Yet, once we find out that B is true, we expect the profile match, and whether G holds has no further impact on the probability of M.

Take an example of a diverging connections. Say you have two coins, one fair, one biased. Conditional on which coin you have chosen, the results of subsequent tosses are independent. But if you don't know which coin you have chosen, the result of previous tosses give you some information about which coin it is, and this has impact on your estimate of the probability of heads in the next toss. Whether a coin is fair, F or not has an impact on the result of the first toss, $H1$, and on the result of the second toss, $H2$. So $H1 \leftarrow F \rightarrow H2$ seems to be appropriate. Now, on one hand, as long as we don't know whether F, $H1$ increases the probability of $H2$. On the other, once we know that F, though, $H1$ and $H2$ become independent, and so conditioning on the parent in a fork makes its childern independent (provided there is no other open path between them in the graph).

For converging connections, let G and B be as above, and let:

Node	Proposition
O	The crime scene stain comes from the offender.

Both *G* and *O* influence *B*. If he's guilty, it's more likely that the blood stain comes from him, and if the blood crime stain comes from the offender it is more likely to come from the suspect (for instance, more so than if it comes from the victim). Moreover, *G* and *O* seem independent – whether the suspect is guilty doesn't have any bearing on whether the stain comes from the offender. Thus, a converging connection $G \to B \leftarrow O$ seem appropriate. However, if you do find out that *B* is true, that the stain comes from the suspect, whether the crime stain comes from the offender becomes relevant for whether the suspect is guilty.

One important reason why d-separation matters is that it can be proven that if two sets of nodes are d-separated given a third one, then they are independent given the third one, for any probabilistic measures compatible with a given DAG. Interestingly, lack of d-separation doesn't entail dependence for any probabilistic measure compatible with a given DAG. Rather, it only allows for it: if nodes are d-separated, there is at least one probabilistic measure fitting the DAG according to which they are independent. So, at least, no false independence can be inferred from the DAG, and all the dependencies are built into it.

Now, getting back to the conjunction problem, the d-separations entailed by these networks differ (examples can be found in Table 2)—in fact, BN1 entails 31 d-separations, while BN2 entails 22 of them. Attention should be paid to the notation. In the above, variables represent nodes, and so each d-separation entails a probabilistic statement about all combination of the node states involved. For instance, assuming each node is binary with two possible states, 1 and 0, $B \perp\!\!\!\perp_d a$ entails that for any $B_i, a_i \in \{0,1\}$ we have $\mathsf{P}(B = B_i) = \mathsf{P}(B = B_i | a = a_i)$.

Bayesian network 1	Bayesian network 2
$A \perp\!\!\!\perp_d B$	$A \perp\!\!\!\perp_d b\|B$
$A \perp\!\!\!\perp_d b$	$AB \perp\!\!\!\perp_d a\|A$
$AB \perp\!\!\!\perp_d a\|A$	$AB \perp\!\!\!\perp_d b\|B$
$AB \perp\!\!\!\perp_d b\|B$	$B \perp\!\!\!\perp_d a\|A$
$B \perp\!\!\!\perp_d a$	$a \perp\!\!\!\perp_d b\|B$
$a \perp\!\!\!\perp_d b$	$a \perp\!\!\!\perp_d b\|A$

{tab:indepBNS}

Table 2: Some of d-separations entailed by BN1 and BN2 in the conjunction problem. One minimal testable implication (with the smallest possible conditioning set) is returned per missing edge of the graph.

In what follows, however, we will sometimes use a finer level of granularity, being very explicit on what independence assumptions are used in the derivations. In such contexts, we will be talking about states rather than nodes, and so when we present the derivation, $b \perp\!\!\!\perp A \wedge a | \neg B$ is a claim about events (or propositions) means the same as $\mathsf{P}(b = 1 | B = 0) = \mathsf{P}(b = 1 | A = 1, a = 1, B = 0)$. This distinction matters, as, first, independence conditional on $B = 0$ doesn't entail independence given $B = 1$ (for instance, your final grade might depend on how hard you work if the teacher is fair, but this might fail if the teacher is not fair), and, second, sometimes only some of the independencies entailed by a Bayesian network will be actually required for a given claim to hold, and we want to be explicit about such cases. We hope this slight ambiguity in notation will cause no confusion, as whether we talk about nodes or events will be clear from the context. So, moving to events, here is a list of independence claims used in the arguments that follow. We also marked whether they are entailed by the Bayesian networks under consideration.

{eq:indAB}	$A \perp\!\!\!\perp B$	BN1	(1)
{eq:indab}	$b \perp\!\!\!\perp a$	BN1	(2)
{eq:I1}	$A \perp\!\!\!\perp b \mid a$	BN1	(3)
{eq:I2}	$B \perp\!\!\!\perp a \wedge A \mid b$	BN1 , BN2	(4)
{eq:I3}	$a \perp\!\!\!\perp b \mid A \wedge B$	BN1 , BN2	(5)
{eq:I3a}	$a \perp\!\!\!\perp b \mid A$	BN1 , BN2	(6)
{eq:I4}	$a \perp\!\!\!\perp B \mid A$	BN1 , BN2	(7)
{eq:I4a}	$a \perp\!\!\!\perp B \mid \neg A$	BN1 , BN2	(8)
{eq:I4b}	$a \perp\!\!\!\perp \neg B \mid A$	BN1 , BN2	(9)
{eq:I4c}	$a \perp\!\!\!\perp \neg B \mid \neg A$	BN1 , BN2	(10)
{eq:I5}	$b \perp\!\!\!\perp A \wedge a \mid B$	BN1 , BN2	(11)
{eq:I5a}	$b \perp\!\!\!\perp \neg A \wedge a \mid B$	BN1 , BN2	(12)
{eq:I5b}	$b \perp\!\!\!\perp A \wedge a \mid \neg B$	BN1 , BN2	(13)
{eq:I5c}	$b \perp\!\!\!\perp \neg A \wedge a \mid \neg B$	BN1 , BN2	(14)
{eq:I6}	$b \perp\!\!\!\perp a \mid B$	BN1 , BN2	(15)

> Double check if we in fact use I3.

We will start with the Bayes factor $\mathrm{P}(E|H)/\mathrm{P}(E)$ and its relation to conjunction. First, given CONDITIONS, if separate BFs are greater than 1, then so is the joint BF.

Fact 1. *Suppose a and b positively support A and B, that is:*

$$BF_A = \frac{\mathrm{P}(a|A)}{\mathrm{P}(a)} > 1 \tag{16}$$

$$BF_B = \frac{\mathrm{P}(b|B)}{\mathrm{P}(b) = s_B} => 1. \tag{17}$$

Then, if CONDITIONS, it follows that:

$$BF_{AB} > 1 = \frac{\mathrm{P}(a \wedge b|A \wedge B)}{\mathrm{P}(a \wedge b)} > 1. \tag{18}$$

Proof.

$$\frac{\mathrm{P}(a \wedge b|A \wedge B)}{\mathrm{P}(a \wedge b)} = \frac{\mathrm{P}(A \wedge B|a \wedge b)}{\mathrm{P}(A \wedge B)} \qquad \text{(Bayes's theorem)}$$

$$= \frac{\frac{\mathrm{P}(A \wedge B \wedge a \wedge b)}{\mathrm{P}(a \wedge b)}}{\mathrm{P}(A \wedge B)} \qquad \text{(definition of conditional probability)}$$

$$= \frac{\frac{\mathrm{P}(A) \times \mathrm{P}(B|A) \times \mathrm{P}(a|A \wedge B) \times \mathrm{P}(b|A \wedge B \wedge a)}{\mathrm{P}(a) \times \mathrm{P}(b|a)}}{\mathrm{P}(A \wedge B)} \qquad \text{(chain rule)}$$

$$= \frac{\frac{\mathrm{P}(A) \times \mathrm{P}(B|A) \times \mathrm{P}(a|A \wedge B) \times \mathrm{P}(b|A \wedge B \wedge a)}{\mathrm{P}(a) \times \mathrm{P}(b|a)}}{\mathrm{P}(A) \times \mathrm{P}(B)} \qquad \text{(independencies in Figure ??)}$$

$$= \frac{\frac{\mathrm{P}(A) \times \mathrm{P}(B) \times \mathrm{P}(a|A) \times \mathrm{P}(b|B)}{\mathrm{P}(a) \times \mathrm{P}(b)}}{\mathrm{P}(A) \times \mathrm{P}(B)} \qquad \text{(independencies in Figure ??)}$$

$$= \frac{\mathrm{P}(a|A)}{\mathrm{P}(a)} \times \frac{\mathrm{P}(b|B)}{\mathrm{P}(b)} \qquad \text{(algebraic manipulation)}$$

$$BF_{AB} = BF_A \times BF_B$$

$\square$

The step marked by asterisk rests on the independence assumptions codified in Figure **??**, namely: (1), (7), and (11).]

$$\frac{\mathsf{P}(a \wedge b | A \wedge B)}{\mathsf{P}(a \wedge b)} = \frac{\mathsf{P}(a|A)}{\mathsf{P}(a)} \times \frac{\mathsf{P}(b|B)}{\mathsf{P}(b)}$$

$$BF_{AB} = BF_A \times BF_B$$

Alicja: M: Check reference to independence assumptions