# Appendix: confirmation measures and the difficulty about conjunction

Marcello Di Bello and Rafal Urbaniak

## Contents



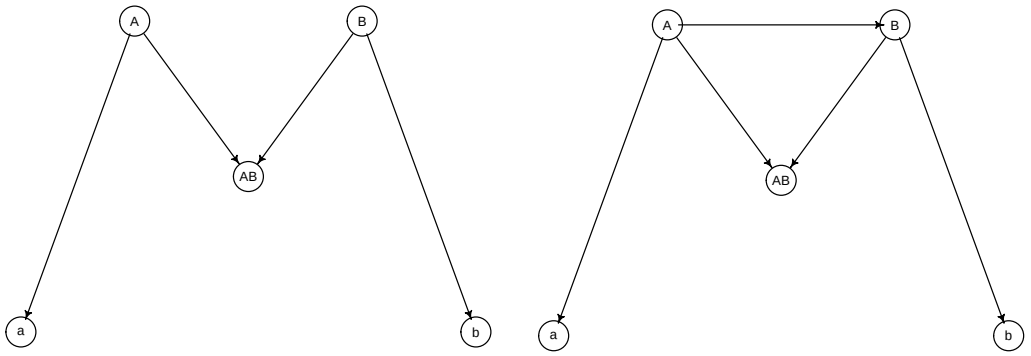
paragraph of flow

istic-independence}

### Bayesian networks and probabilistic independence

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 for a direct dependence between the hypotheses in the second network. In all the Bayesian networks that will be discussed in this appendix, the CPT for the conjunction trivially mirrors the one for conjunction, as in Table 1.




(a) DAG1 (b) DAG2

fig:conjunctionBNs}

Figure 1: Two DAGs for the conjunction problem.

| AB | A | B | 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 nondescentants (including ancestors), given its parents. If this is the case for a given probabilistic measure $\mathsf{P}()$ and a given DAG, we say that $\mathsf{P}()$ is compatible with $\mathsf{G}$, 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 $\mathsf{Z}$—$X \perp\!\!\!\perp_d Y|\mathsf{Z}$ — iff for every undirected path from $X$ to $Y$ there is a node $Z'$ on the path such that either:

- $Z' \in \mathsf{Z}$ and there is a **serial** connection, $\rightarrow Z' \rightarrow$, on the path (**pipe**),
- $Z' \in \mathsf{Z}$ and there is a **diverging** connection, $\leftarrow Z' \rightarrow$, on the path (**fork**),
- There is a **converging** connection $\rightarrow Z' \leftarrow$ on the path (in which case $Z'$ is a **collider**), and neither $Z'$ nor its descendants are in $\mathsf{Z}$.

**Alicja:** R: add graph of three types next to each other using daggity.

Finally, two sets of nodes, $\mathsf{X}$ and $\mathsf{Y}$, are d-separated given $\mathsf{Z}$ if every node in $\mathsf{X}$ is d-separated from every node in $\mathsf{Y}$ given $\mathsf{Z}$. 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, $G$ has 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 an 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 suspect 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 $G \rightarrow B \leftarrow O$ 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, DAG1 entails 31 d-separations, while DAG2 entails 22 of them. Attention should be paid to the notation. Whe we talk about DAGs, 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 $\mathsf{P}(B = B_i) = \mathsf{P}(B = B_i | a = a_i)$.

{tab:indepBNS}

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

Table 2: Some of d-separations entailed by DAG1 and DAG2 in the conjunction problem. One minimal testable implication (with the smallest possible conditioning set) is returned per missing edge of the graph.

Refer to the code

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 DAG under consideration.

| | | | |
|---|---|---|---|
| {eq:indAB} | $A \perp\!\!\!\perp B$ | DAG1 | (1) |
| {eq:indab} | $b \perp\!\!\!\perp a$ | DAG1 | (2) |
| {eq:I1} | $A \perp\!\!\!\perp b \mid a$ | DAG1 | (3) |
| {eq:I2} | $B \perp\!\!\!\perp a \wedge A \mid b$ | DAG1 | (4) |
| {eq:I3} | $a \perp\!\!\!\perp b \mid A \wedge B$ | DAG1 , DAG2 | (5) |
| {eq:I3a} | $a \perp\!\!\!\perp b \mid A$ | DAG1 , DAG2 | (6) |
| {eq:I3b} | $a \perp\!\!\!\perp b \mid \neg A$ | DAG1 , DAG2 | (7) |
| {eq:I4} | $a \perp\!\!\!\perp B \mid A$ | DAG1 , DAG2 | (8) |
| {eq:I4a} | $a \perp\!\!\!\perp B \mid \neg A$ | DAG1 , DAG2 | (9) |
| {eq:I4b} | $a \perp\!\!\!\perp \neg B \mid A$ | DAG1 , DAG2 | (10) |
| {eq:I4c} | $a \perp\!\!\!\perp \neg B \mid \neg A$ | DAG1 , DAG2 | (11) |
| {eq:I5} | $b \perp\!\!\!\perp A \wedge a \mid B$ | DAG1 , DAG2 | (12) |
| {eq:I5a} | $b \perp\!\!\!\perp \neg A \wedge a \mid B$ | DAG1 , DAG2 | (13) |
| {eq:I5b} | $b \perp\!\!\!\perp A \wedge a \mid \neg B$ | DAG1 , DAG2 | (14) |
| {eq:I5c} | $b \perp\!\!\!\perp \neg A \wedge a \mid \neg B$ | DAG1 , DAG2 | (15) |
| {eq:I6} | $b \perp\!\!\!\perp a \mid B$ | DAG1 , DAG2 | (16) |

Some caveats. In (3) the conditioning on $a$ is not essential, because it's not on the path between the nodes: the key reason why the independence remains upon this conditioning is that there is an unconditioned collider on the path. Still, we need this independence in the proof later on. In (5) what we are conditioning on is nodes $A$ and $B$ jointly. Technically, independence conditional on the conjunction node $AB$ does not fall out of the d-separations present in the network—it follows given that $AB$ and $A, B$ are connected deterministically: fixing $AB$ to true fixes both $A$ and $B$ to true.

In what follows, we will provide theorems where we managed to obtain them. However, in some cases analytic calculations are somewhat unmanageable, and so we decided to inspect such issues by means of a **simulation**. For each DAG under consideration, we generated 100k random Bayesian networks (with probabilities sampled from the Uniform(0,1) distribution) which share that DAG. For each of these networks, we calculated all the relevant probabilities, Bayes factors, and likelihood ratios. With the output of such calculations, various questions can be asked and the answers visualized. Even for cases in which we obtained the relevant proofs, the results of a simulation provide further insights, for instance, in terms of relative frequencies of various facts.

## Bayes factor: claims and proofs

-claims-and-proofs}

We will start with the Bayes factor $\mathsf{P}(E|H)/\mathsf{P}(E)$ and its relation to conjunction. Let's abbreviate:

$$BF_A = \frac{\mathsf{P}(a|A)}{\mathsf{P}(a)},$$
$$BF_B = \frac{\mathsf{P}(b|B)}{\mathsf{P}(b)}$$

**Fact 1.** *If the independence assumptions* (1), (2), (8) *and* (12) *hold (all of which are entailed by DAG1), then* $BF_{AB} = BF_A \times BF_B$.

{fac:BFindep}

*Proof.*

$$\frac{\mathsf{P}(a \wedge b|A \wedge B)}{\mathsf{P}(a \wedge b)} = \frac{\mathsf{P}(A \wedge B \wedge a \wedge b)}{\mathsf{P}(A \wedge B)} \Big/ \mathsf{P}(a \wedge b) \qquad \text{(conditional probability)}$$
$$= \frac{\mathsf{P}(A) \times \mathsf{P}(B|A) \times \mathsf{P}(a|A \wedge B) \times \mathsf{P}(b|A \wedge B \wedge a)}{\mathsf{P}(A \wedge B)} \Big/ \mathsf{P}(a \wedge b) \qquad \text{(chain rule)}$$

Now, let's deploy the respective independence assumptions, as follows:

$$
\begin{aligned}
&= \frac{\mathsf{P}(A) \times \overbrace{\mathsf{P}(B|A)}^{\mathsf{P}(B) \text{ by (1)}} \times \overbrace{\mathsf{P}(a|A \wedge B)}^{\mathsf{P}(a|A) \text{ by (8)}} \times \overbrace{\mathsf{P}(b|A \wedge B \wedge a)}^{\mathsf{P}(b|B) \text{ by (12)}}}{\underbrace{\mathsf{P}(A \wedge B)}_{\mathsf{P}(A)\times\mathsf{P}(B) \text{ by (1)}}} \Big/ \underbrace{\mathsf{P}(a \wedge b)}_{\mathsf{P}(a)\times\mathsf{P}(b) \text{ by (2)}} \\
&= \frac{\mathsf{P}(A) \times \mathsf{P}(B) \times \mathsf{P}(a|A) \times \mathsf{P}(b|B)}{\mathsf{P}(A) \times \mathsf{P}(B)} \Big/ \mathsf{P}(a) \times \mathsf{P}(b) \\
&= \frac{\mathsf{P}(a|A)}{\mathsf{P}(a)} \times \frac{\mathsf{P}(b|B)}{\mathsf{P}(b)} \\
&= BF_A \times BF_B
\end{aligned}
$$

□

This fact has the following straigthforward consequences which simply follow from the simple fact that if $a = b \times c$ and $b, c > 1$, then $a > \max(b,c)$ and that if $b, c < 1$, then $a < \min(b,c)$.

{cor:BFind2} **Corollary 1.1.** *If the independence assumptions* (1), (2), (8) *and* (12) *hold, and $BF_A$ and $BF_B$ are both greater than 1, then $BF_{AB}$ is greater than one. In fact, $BF_{AB}$ is then greater than* $\max(BF_A, BF_B)$.

{cor:BFind3} **Corollary 1.2.** *If the independence assumptions* (1), (2), (8) *and* (12) *hold, and $BF_A$ and $BF_B$ are both strictly less than 1, then $BF_{AB}$ is less than* $\min(BF_A, BF_B)$.

Now, does a similar claim if we drop the independence assumption specific to DAG1? A claim somewhat weaker than Fact 1 can be proven without employing independencies entailed by DAG1, relying only on some independencies entailed by DAG2. Let's abbreviate:

$$
\begin{aligned}
BF_B^{'} &= \frac{\mathsf{P}(b|B)}{\mathsf{P}(b|a)} \\
BF_A^{'} &= \frac{\mathsf{P}(a|A)}{\mathsf{P}(a|b)}
\end{aligned}
$$

{fac:BFdep} **Fact 2.** *If* (8) *and* (12) *hold (and they do in BNs based on DAG 2), then $BF_{AB} = BF_A \times BF_B^{'} = BF_B \times BF_A^{'}$.*

*Proof.* We start with the definition of conditional probability and the chain rule, as in the proof of Fact 1, but now we use fewer independencies (all of them entailed by DAG2), as we use the chain rule instead in a few cases.

$$
\begin{aligned}
\frac{\mathsf{P}(a \wedge b|A \wedge B)}{\mathsf{P}(a \wedge b)} &= \frac{\mathsf{P}(A) \times \mathsf{P}(B|A) \times \overbrace{\mathsf{P}(a|A \wedge B)}^{\mathsf{P}(a|A) \text{ by (8)}} \times \overbrace{\mathsf{P}(b|A \wedge B \wedge a)}^{\mathsf{P}(b|B) \text{ by (12)}}}{\underbrace{\mathsf{P}(A \wedge B)}_{\mathsf{P}(A)\times\mathsf{P}(B|A) \text{ by the chain rule}}} \Big/ \underbrace{\mathsf{P}(a \wedge b)}_{\mathsf{P}(a)\times\mathsf{P}(b|a) \text{ by the chain rule}} \\
&= \frac{\mathsf{P}(a|A)}{\mathsf{P}(a)} \times \frac{\mathsf{P}(b|B)}{\mathsf{P}(b|a)} \\
&= BF_A \times BF_B^{'}
\end{aligned}
$$

If instead of obtaining $\mathsf{P}(a)\mathsf{P}(b|a)$ in the denominator we deploy the chain rule differently, resulting in $\mathsf{P}(b)\mathsf{P}(a|b)$, we end up with:

$$
\begin{aligned}
&= \frac{\mathsf{P}(a|A)}{\mathsf{P}(a|b)} \times \frac{\mathsf{P}(b|B)}{\mathsf{P}(b)} \\
&= BF_A^{'} \times BF_B
\end{aligned}
$$

□

Now, to obtain a corollary analogous to Corollary 1.1, there are at lest two paths. One starts with an additional disjunctive assumption that either $\mathsf{P}(b|a) \leq \mathsf{P}(b)$ or $\mathsf{P}(a|b) \leq \mathsf{P}(a)$.

{cor:BFweaker}

**Corollary 2.1.** *Suppose* (8) *and* (12) *hold and* $BF_{BA}, BF_B > 1$. *Then if either* $\mathsf{P}(b|a) \leq \mathsf{P}(b)$ *or* $\mathsf{P}(a|b) \leq \mathsf{P}(a)$, *we have* $BF_{AB} \geq BF_A, BF_B$.

*Proof.* Notice that by Fact 2 we have that on the assumptions of the corollary:

$$\begin{aligned} BF_{AB} &= BF_A \times BF_B^{'} \\ &= BF_B \times BF_A^{'} \end{aligned}$$

If we now can show that either $BF_B^{'} \geq 1$, or $BF_A^{'} \geq 1$, the reasoning we used before applies: the product of two numbers greater than one is not less than either of them. Consider $BF_B$, which we assumed to be greater than 1. If we can show $BF_B^{'} \geq BF_B$, we're done. But this holds on one of the disjuncts assumed in the corollary, $\mathsf{P}(b|a) \leq \mathsf{P}(b)$, as then we have:

$$\begin{aligned} \frac{\mathsf{P}(b|B)}{\mathsf{P}(b|a)} &\geq \frac{\mathsf{P}(b|B)}{\mathsf{P}(b)} \\ BF_B^{'} &\geq BF_B \\ BF_{AB} = BF_A \times BF_B^{'} &\geq BF_A \times BF_B \end{aligned}$$

Similarly, if the other disjunct holds, we run an analogous argument, this time focusing on $BF_A^{'}$. □

R: added this paragraph and another collorary, check

However, the problem with this assumption is that if it holds, this means that one item of evidence makes the other item of evidence at least as surprising as it originally was. This, at least intuitively, might not happen in a legal context. If $A$ and $B$ are elements of a crime, say that the driver was drunk and that they caused harm by erratic driving, evidence for one hypothesis, say the blood alcohol level test, in fact makes the evidence for the other hypothesis—witnesses attesting to the erratic driving, the presence of the harm, and so on, more likely.

Another approach uses a conjunction of assumptions.

{cor:BFweaker2}

**Corollary 2.2.** *Suppose* (8) *and* (12) *hold and* $BF_{BA}, BF_B > 1$. *Then if both* $\mathsf{P}(a|b) \leq \mathsf{P}(a|A)$ *and* $\mathsf{P}(b|a) \leq \mathsf{P}(b|B)$, *we have* $BF_{AB} \geq BF_A, BF_B$.

*Proof.* Assume the first conjunct holds. Then $\frac{\mathsf{P}(a|A)}{\mathsf{P}(a|b)} \geq 1$ and so:

$$BF_{AB} = BF_A^{'} \times BF_B \geq BF_B$$

The argument for the other comparison is analogous. □

While this is a conjunction, not a disjunction, the assumption still seems more plausible in usual legal contexts. If, say, $a$ is used as evidence for $A$, we often expect $A$ and $a$ to be fairly strongly connected, that is, we expect $\mathsf{P}(a|A)$ to be rather high, while the connection between different pieces of evidence for different hypotheses, intuitively, is not expected to be as strong.

## Bayes factor: simulations

factor-simulations}

First, let's look at simulations based on DAG1. We can illustrate the distribution of Bayes factors in the two separate scenarios used as assumptions of the above corollaries.

For the DAG corresponding to DAG1, the simulated frequency of cases in which $BF_{AB} < BF_A, BF_B$ is 25% (which is twice higher than for the likelihood ratio), and the structure of such cases is visualized in Figure 3.

The picture doesn't change when we move to DAG2 (Figure 4). One reason this is interesting is that this suggests that the additional assumption we used in the proof of Corollary 2.1 was not needed. We

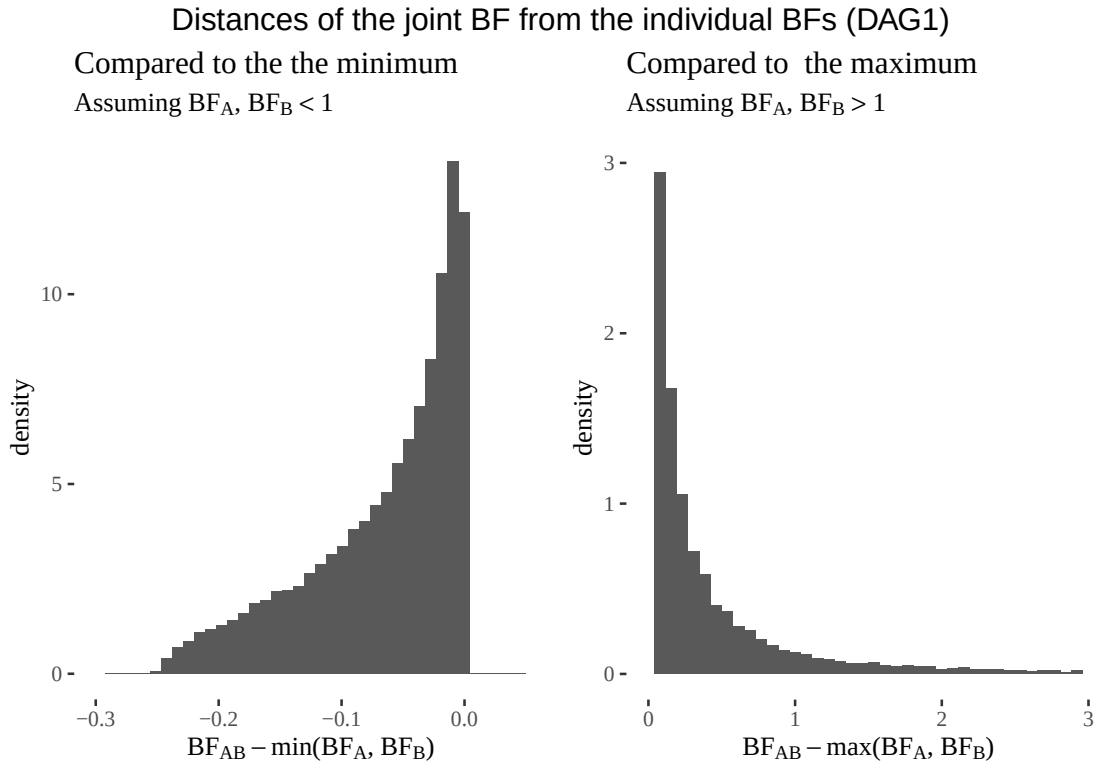



Figure 2: Distances of the joint Bayes factor from maxima and minima of individual Bayes factors, depending on whether the individual support levels are both positive or both negative. Simulation based on 100k Bayesian networks build over the DAG of DAG1.

have failed to derive the claim without it, though. What does change, however, is that the joint BF is no longer the result of multiplying the individual BFs. Nevertheless, the values are fairly close (Figure 5).

Now, what is striking is that while Corollaries 2.1 and 2.2 required additional assumptions to prove that the joint BF will be at least as high as the individual BFs, this assumption is not explicitly build into the simulation, and nevertheless the consequent comes out satisfied. In fact, even more can be said. In the simulation $\mathsf{P}(b|a) > \mathsf{P}(b)$ around 50% of the time, and the same holds for $\mathsf{P}(b|a) > \mathsf{P}(b)$, so the assumptions of Corollary 2.1 fail.

So do the assumptions of Corollary 2.2: $\mathsf{P}(a|b) > \mathsf{P}(a|A)$ around 50% of the time and the same holds for $\mathsf{P}(b|a) > \mathsf{P}(b|B)$, the disjunction of these claims also fails 50% of the time, and all these frequencies do not change even if we restrict attention to cases in which the individual BFs are above 1. It is not generally the case in DAG 2 that $BF_A^{'} \geq BF_A$ or $BF_B^{'} \geq BF_B$. So what validates the results?

What does in fact hold in all BNs build on DAG 2, however, is the assumptions of Corollary 2.2: it falls out of the DAG 2 setup and the assumption that the individual Bayes factor is greater than 1 (this assumption is essential, the claim fails without it) that $\mathsf{P}(a|b) \leq \mathsf{P}(a|A)$ and $\mathsf{P}(b|a) \leq \mathsf{P}(b|B)$, and so, still both $BF_A^{'}$ and $BF_B^{'}$ are above one if the individual Bayes factors are above one, even though the former might be lower the latter (respectively). We will turn to proving this very soon.

How about distribution? Note that in principle a failure of distribution can occur whenever the joint BF is strictly greater than at least one of the individual BFs. The distribution of such cases for both DAGs can be inspected in Figure 6.

## Corollary 2.2 and DAG 2

{corollary-and}

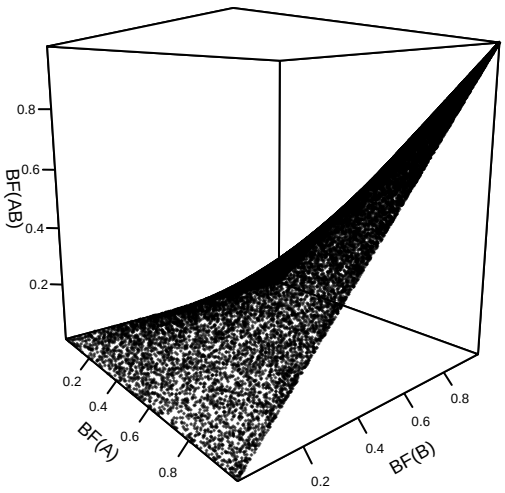
Let us start with the following lemma.

{lem:BFLR}

**Lemma 1.** *For any probabilistic measure* P*, if* $BF_A > 1$*, then* $LR_A > 1$.

**Alicja:** R: this subsection is new, check statements and proofs in detail.

R: this subsection is new, check statements and proofs in detail.




{fig:BFfails}

Figure 3: Ca. 25k cases (out of simulated 100k) in which the joint BF is below each of the individual BFs.

*Proof.* We start with our assumption.

$$
\begin{aligned}
1 &\leq \frac{\mathsf{P}(a|A)}{\mathsf{P}(a)} && (BF_A \geq 1)\\
\mathsf{P}(A) &\leq \frac{\mathsf{P}(a|A)}{\mathsf{P}(a)}\mathsf{P}(A) && \text{(algebraic manipulation)}\\
\mathsf{P}(A) &\leq \mathsf{P}(A|a) && \text{(Bayes' theorem)}\\
-\mathsf{P}(A) &\geq -\mathsf{P}(A|a) && \text{(algebraic manipulation)}\\
1-\mathsf{P}(A) &\geq 1-\mathsf{P}(A|a) && \text{(algebraic manipulation)}\\
1-\mathsf{P}(A) &\geq \mathsf{P}(\neg A|a) && \text{(algebraic manipulation)}\\
\mathsf{P}(a)\,(1-\mathsf{P}(A)) &\geq \mathsf{P}(a)\mathsf{P}(\neg A|a) && \text{(algebraic manipulation)}\\
\mathsf{P}(a) &\geq \frac{\mathsf{P}(a)\mathsf{P}(\neg A|a)}{\mathsf{P}(\neg A)} && \text{(algebraic manipulation, negation)}\\
\mathsf{P}(a) &\geq \mathsf{P}(a|\neg A) && \text{(conditional probability)}
\end{aligned}
$$

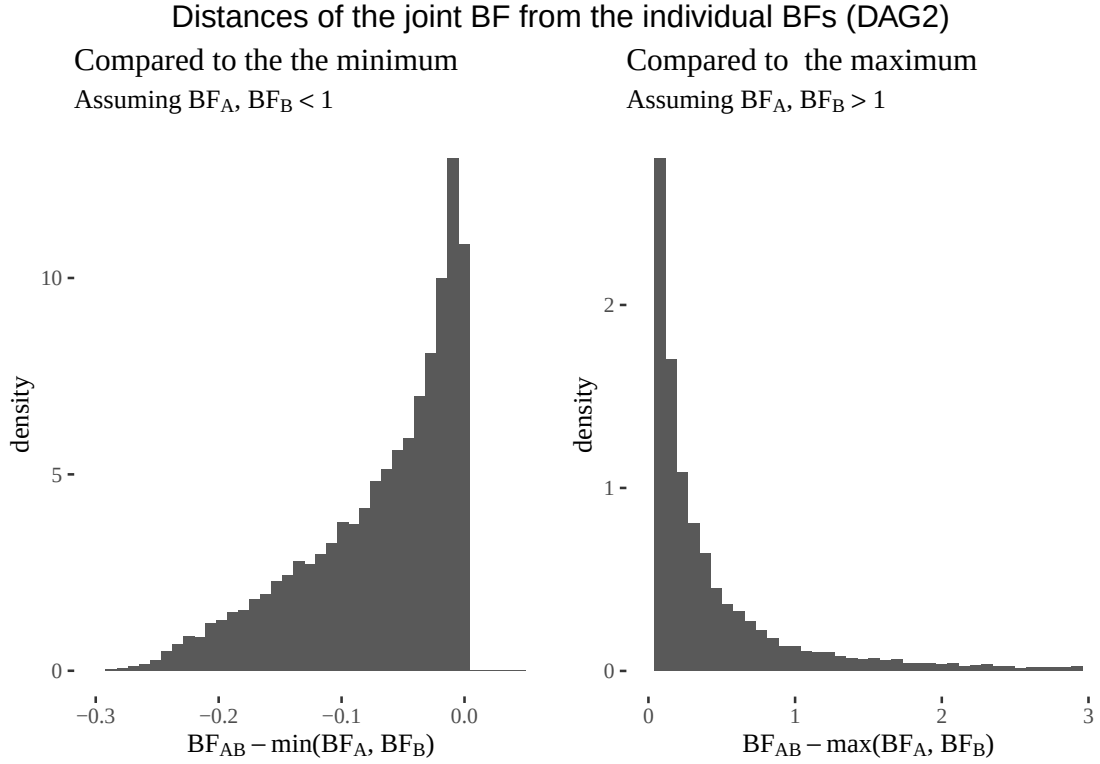
From this and our assumption that $\mathsf{P}(a|A) \geq \mathsf{P}(a)$ it follows that $\mathsf{P}(a|A) \geq \mathsf{P}(a|\neg A)$, that is, that $LR_A \geq 1$. □

Now the main claim.

**Fact 3.** *For any probabilistic measure* $\mathsf{P}$ *appropriate for DAG 2, if* $BF_A > 1$*, then* $\mathsf{P}(a|A) \geq \mathsf{P}(a|b)$.

*Proof.* First, we have:

$$
\begin{aligned}
\mathsf{P}(a|b) &= \mathsf{P}(a \wedge A|b) + \mathsf{P}(a \wedge \neg A|b) && \text{(total probability)}\\
&= \underbrace{\mathsf{P}(a|b \wedge A)}_{\mathsf{P}(a|A) \text{ by } (6)}\mathsf{P}(A|b) + \underbrace{\mathsf{P}(a|b \wedge \neg A)}_{\mathsf{P}(a|\neg A) \text{ by } (7)}\mathsf{P}(\neg A|b) && \text{(chain rule)}
\end{aligned}
$$




{fig:BFind2}

Figure 4: Distances of the joint Bayes factor from maxima and minima of individual Bayes factors, depending on whether the individual support levels are both positive or both negative. Simulation based on 100k Bayesian networks build over the DAG of DAG2.

Now let's introduce some abbreviations:

$$= \underbrace{\mathsf{P}(a|A)}_{k}\underbrace{\mathsf{P}(A|b)}_{x} + \underbrace{\mathsf{P}(a|\neg A)}_{t}\underbrace{\mathsf{P}(\neg A|b)}_{(1-x)}$$

Note that the assumption that $BF_A \geq 1$ entails, by Lemma 1, that $k \geq t$, and so $k - t \geq 0$. Also, since $x$ is a probability, we know $0 \leq x \leq 1$. This allows us to reason algebraically as follows:

$$\begin{aligned} k &\geq k \\ k &\geq t + (k-t) \\ k &\geq t + (k-t)x \\ k &\geq kx + t - tx \\ \mathsf{P}(a|A) = k &\geq kx + t(1-x) = \mathsf{P}(a|b) \end{aligned}$$

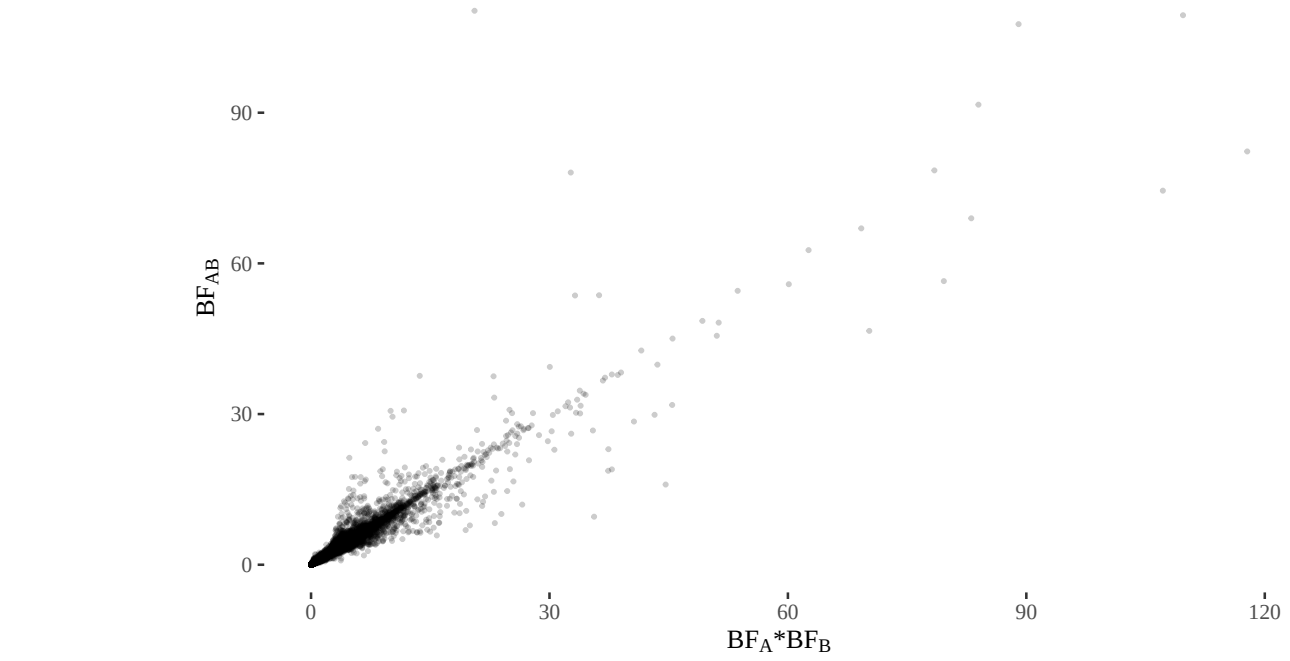
This completes the proof. □

### Likelihood ratio: claims and proofs

-claims-and-proofs}

Now, let's turn to the likelihood understood as:

$$\frac{\mathsf{P}(E|H)}{\mathsf{P}(E|\neg H)} = \frac{\textit{sensitivity}}{\textit{1- specificity}}$$




{fig:BFmulti}

Figure 5: In DAG2, the result of multiplying individual BFs does not equal the joing BF, but often is a good approximation thereof.

Let's introduce the following abbreviations:

$$LR_{AB} = \frac{\mathsf{P}(a \wedge b | a \wedge B)}{\mathsf{P}(a \wedge b | \neg(A \wedge B))}$$
$$LR_{A} = \frac{\mathsf{P}(a|A)}{\mathsf{P}(a|\neg A)}$$
$$LR_{B} = \frac{\mathsf{P}(b|B)}{\mathsf{P}(b|\neg B)}.$$

**Fact 4.** *If independence conditions* (8), (9), (10), (11), (12), (13), (14), *and* (15) *hold, then:*

$$LR_{AB} = \frac{\mathsf{P}(a|A) \times \mathsf{P}(b|B)}{\frac{\mathsf{P}(\neg A)\mathsf{P}(B|\neg A)\mathsf{P}(a|\neg A)\mathsf{P}(b|B)+\mathsf{P}(A)\mathsf{P}(\neg B|A)\mathsf{P}(a|A)\mathsf{P}(b|\neg B)+\mathsf{P}(\neg A)\mathsf{P}(\neg B|\neg A)\mathsf{P}(a|\neg A)\mathsf{P}(b|\neg B)}{\mathsf{P}(\neg A)\mathsf{P}(B|\neg A)+\mathsf{P}(A)\mathsf{P}(\neg B|A)+\mathsf{P}(\neg A)\mathsf{P}(\neg B|\neg A)}}$$

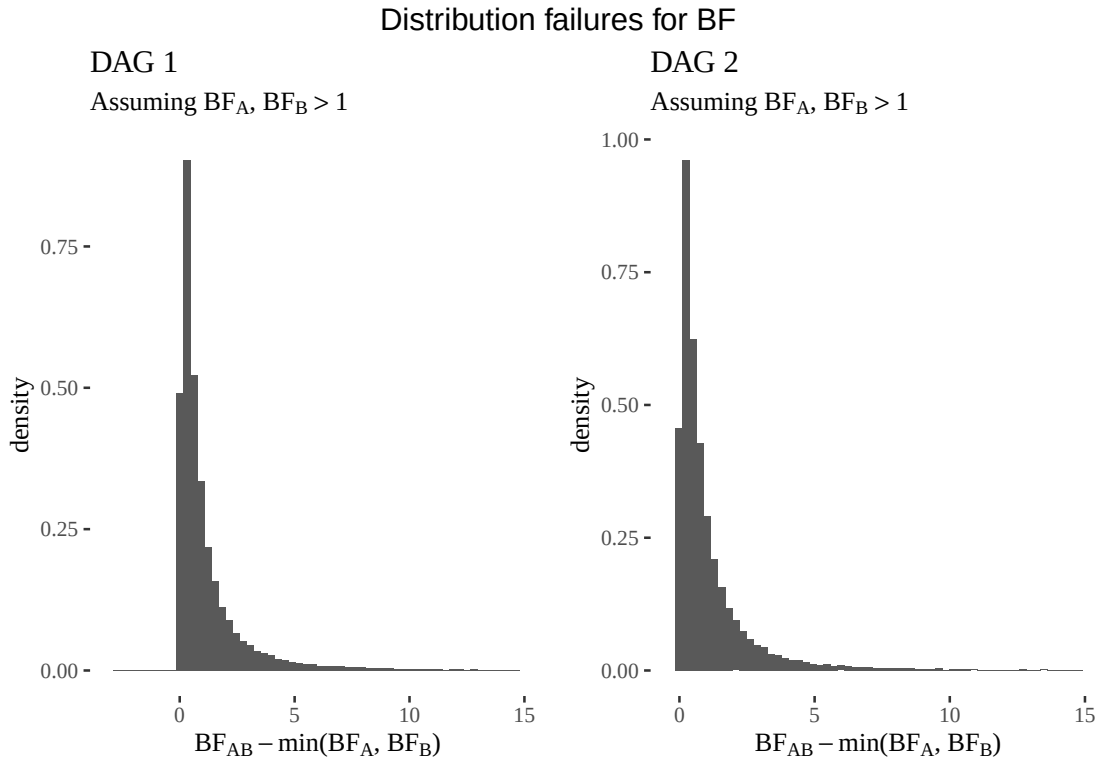
Note that these independence assumptions are entailed not only in DAG1, but also in DAG2.

*Proof.* Let's first compute the numerator of $LR_{AB}$:

$$\begin{aligned}
\mathsf{P}(a \wedge b | A \wedge B) &= \frac{\mathsf{P}(A \wedge B \wedge a \wedge b)}{\mathsf{P}(A \wedge B)} && \text{(conditional probability)} \\
&= \frac{\mathsf{P}(A) \times \mathsf{P}(B|A) \times \mathsf{P}(a|A \wedge B) \times \mathsf{P}(b|A \wedge B \wedge a)}{\mathsf{P}(A) \times \mathsf{P}(B|A)} && \text{(chain rule)}
\end{aligned}$$

We deploy the relevant independencies as follows:

$$\begin{aligned}
&= \frac{\mathsf{P}(A) \times \mathsf{P}(B|A) \times \overbrace{\mathsf{P}(a|A \wedge B)}^{\mathsf{P}(a|A) \text{ by } (8)} \times \overbrace{\mathsf{P}(b|A \wedge B \wedge a)}^{\mathsf{P}(b|B) \text{ by } (12)}}{\mathsf{P}(A) \times \mathsf{P}(B|A)} \\
&= \mathsf{P}(a|A) \times \mathsf{P}(b|B) && \text{(algebraic manipulation)}
\end{aligned}$$




{fig:BFdistr}

Figure 6: Distribution failure for the Bayes factor, DAG 1. The $x$ axis restricted to $(-3, 15)$ for visibility.

The denominator of $LR_{AB}$ is more complicated, mostly because of the conditioning on $\neg(A \wedge B)$.

$$\begin{aligned} \mathsf{P}(a \wedge b | \neg(A \wedge B)) &= \frac{\mathsf{P}(a \wedge b \wedge \neg(A \wedge B))}{\mathsf{P}(\neg(A \wedge B))} && \text{(conditional probability)} \\ &= \frac{\mathsf{P}(a \wedge b \wedge \neg A \wedge B) + \mathsf{P}(a \wedge b \wedge A \wedge \neg B) + \mathsf{P}(a \wedge b \wedge \neg A \wedge \neg B)}{\mathsf{P}(\neg A \wedge B) + \mathsf{P}(A \wedge \neg B) + \mathsf{P}(\neg A \wedge \neg B)} && \text{(logic \& additivity)} \end{aligned}$$

Now consider the first summand from the numerator:

$$\begin{aligned} \mathsf{P}(a \wedge b \wedge \neg A \wedge B) &= \mathsf{P}(\neg A)\mathsf{P}(B|\neg A)\mathsf{P}(a|\neg A \wedge B)\mathsf{P}(b|a \wedge \neg A \wedge B) && \text{(chain rule)} \\ &= \mathsf{P}(\neg A)\mathsf{P}(B|\neg A)\mathsf{P}(a|\neg A)\mathsf{P}(b|B) && \text{(independencies (9) and (13))} \end{aligned}$$

The simplification of the other two summanda is analogous (albeit with slightly different independence assumptions—(10) and (14) for the second one and (11) and (15) for the third. Once we plug these into the denominator formula we get:

$$\begin{aligned} \mathsf{P}(a \wedge b | \neg(A \wedge B)) &= \frac{\mathsf{P}(\neg A)\mathsf{P}(B|\neg A)\mathsf{P}(a|\neg A)\mathsf{P}(b|B) + \mathsf{P}(A)\mathsf{P}(\neg B|A)\mathsf{P}(a|A)\mathsf{P}(b|\neg B) + \mathsf{P}(\neg A)\mathsf{P}(\neg B|\neg A)\mathsf{P}(a|\neg A)\mathsf{P}(b|\neg B)}{\mathsf{P}(\neg A)\mathsf{P}(B|\neg A) + \mathsf{P}(A)\mathsf{P}(\neg B|A) + \mathsf{P}(\neg A)\mathsf{P}(\neg B|\neg A)} \\ &= \frac{\mathsf{P}(\neg A)\mathsf{P}(B|\neg A)\mathsf{P}(a|\neg A)\mathsf{P}(b|B) + \mathsf{P}(A)\mathsf{P}(\neg B|A)\mathsf{P}(a|A)\mathsf{P}(b|\neg B) + \mathsf{P}(\neg A)\mathsf{P}(\neg B|\neg A)\mathsf{P}(a|\neg A)\mathsf{P}(b|\neg B)}{\mathsf{P}(\neg A)\mathsf{P}(B|\neg A) + \mathsf{P}(A)\mathsf{P}(\neg B|A) + \mathsf{P}(\neg A)\mathsf{P}(\neg B|\neg A)} \end{aligned}$$

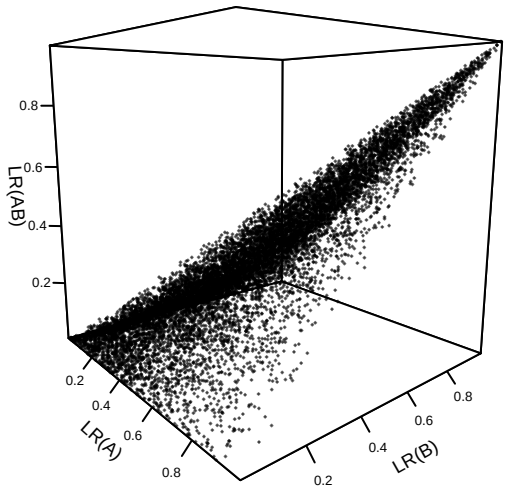
□

-ratio-simulations}

### Likelihod ratio: simulations

While the analytic approach turns out to be cumbersome, let's inspect the problem using simulations. First of all, there are cases in which the joint likelihood ratio are lower than each of the individual likelihood ratios. Their frequency is twice lower than the corresponding frequency for the Bayes factor (recall Figure 3). For the DAG corresponding to both DAGs, the simulated frequency of cases in which

$LR_{AB} < LR_A, LR_B$ is 12-12.5% and the distribution of such cases, somewhat of a different shape, is visualized in Figure 7 (the picture for DAG2 is very similar).

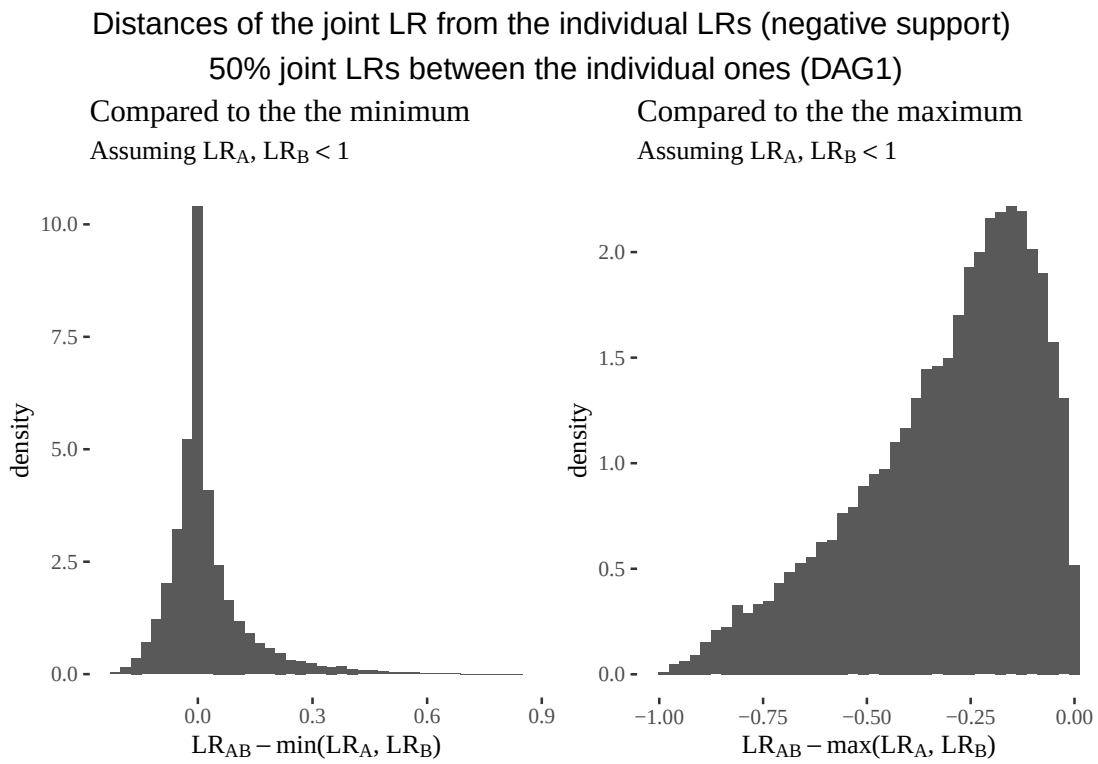



{fig:LRfails}

Figure 7: Ca. 25k cases (out of simulated 100k) in which the joint BF is below each of the individual BFs.

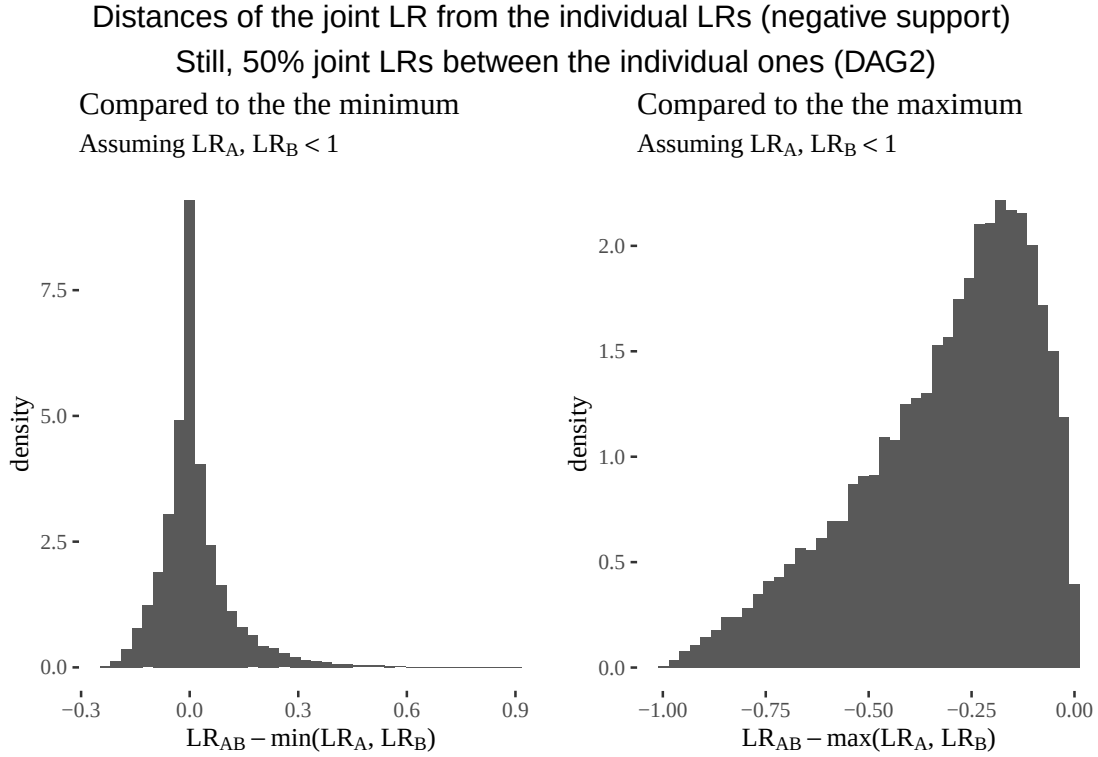
Interestingly, even if the individual likelihood ratios are $< 1$, the joint likelihood ratio can be higher than their minimum, but is never higher than their maximum (Figures 8 and 9).

Once the individual likelihood ratios are above 1, the joint likelihood ratio can be lower than the maximum, but is not lower than the minimum of the individual likelihood ratios (Figures 10 and 11).



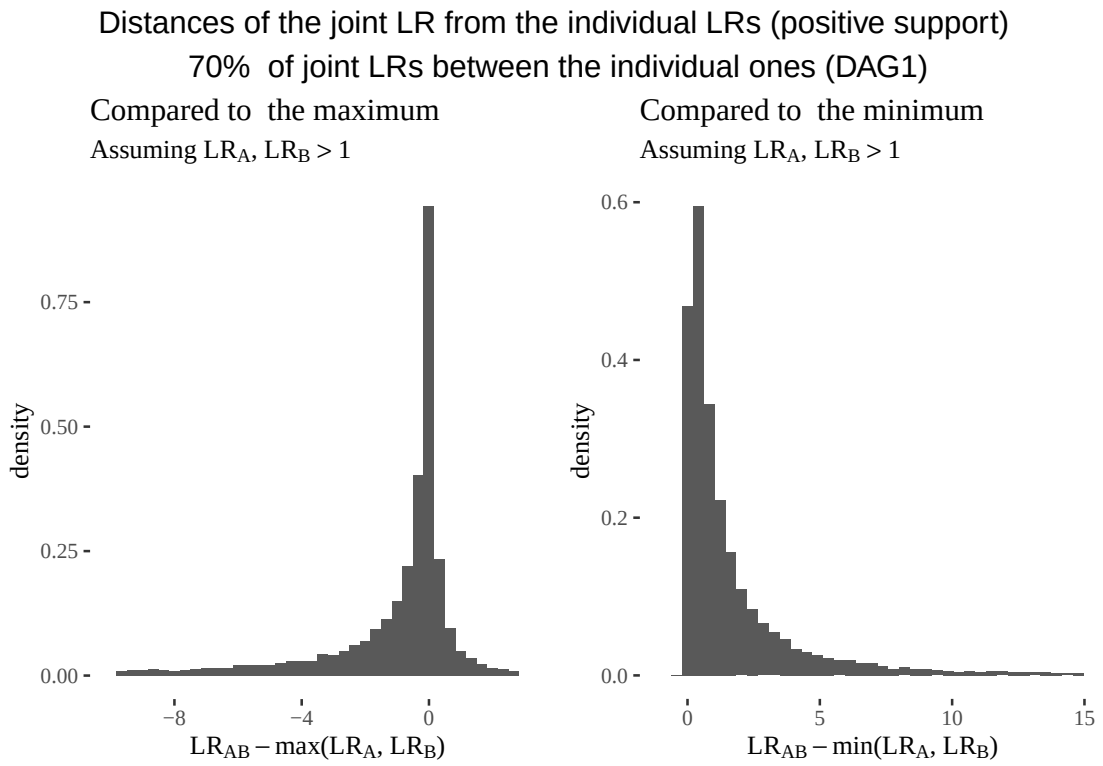

{fig:LRlowerPlot}

Figure 8: Distances of joint likelihood ratios for the minima and the maxima of the individual likelihood ratios if the individual likelihood ratios are below 1, DAG used in DAG1.




{fig:LRlowerPlot2}

Figure 9: Distances of joint likelihood ratios for the minima and the maxima of the individual likelihood ratios if the individual likelihood ratios are below 1, DAG used in DAG2.




{fig:LRabovePlot}

Figure 10: Distances of joint likelihood ratios for the minima and the maxima of the individual likelihood ratios if the individual likelihood ratios are above 1, DAG used in DAG1.

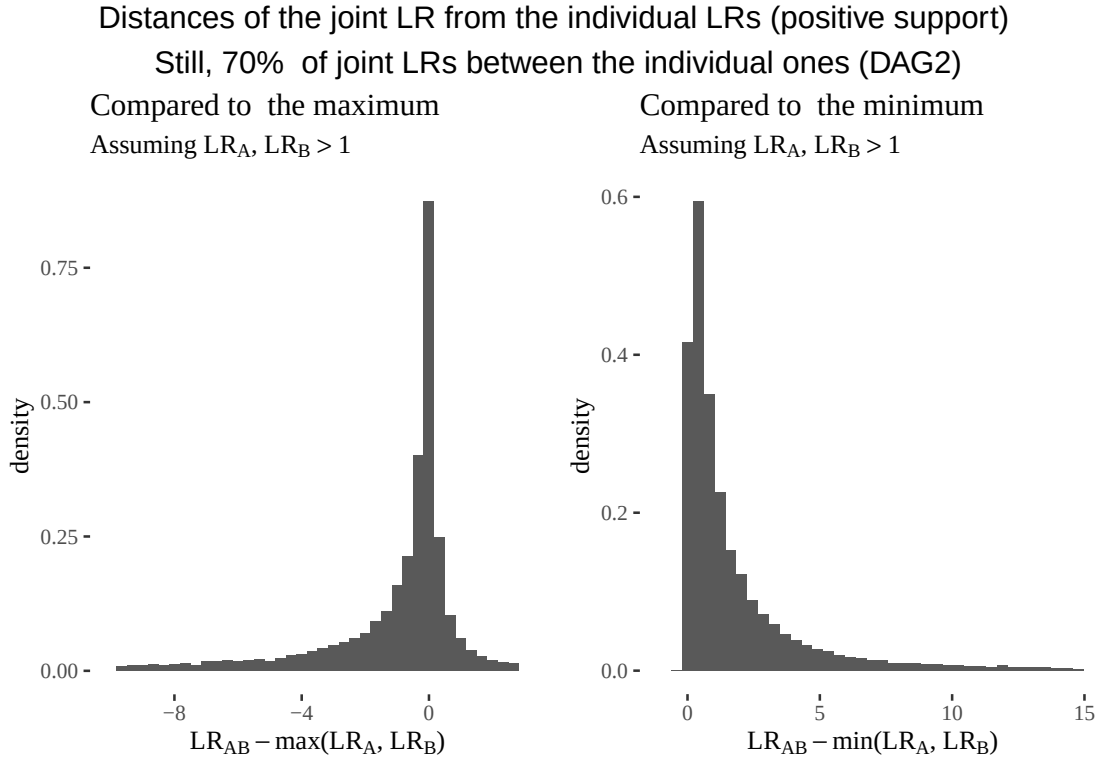



fig:LRabovePlotDep}

Figure 11: Distances of joint likelihood ratios for the minima and the maxima of the individual likelihood ratios if the individual likelihood ratios are above 1, DAG used in DAG2.