Mutual Information and Kullback-Leibler Divergence

Questions

Q1: Which is a better measure to report - KL Divergence or Mutual Information?

Q2: Is it true that the mutual information of a variable to itself is 1?

Answers

Q1: Mutual Information vs KL Divergence

The Mutual Information between two variables X and Y is defined as follows: $$I(X,Y) =\sum_{x \in X}\sum_{y \in Y} p(x,y)\log_2 \frac{p(x,y)}{p(x)p(y)}$$ The KL Divergence allows comparing two probability distributions, P and Q $$D_{KL}(P({\cal X})\|Q({\cal X})) =\sum_{\cal X}P({\cal X})log_2\frac{P({\cal X})}{Q({\cal X})}$$ We use the KL Divergence in BayesiaLab for measuring the strength of a direct relationship between two variables. Here, P is the Bayesian network with the link, and Q is the one without the link. The Mutual Information can be rewritten as: $$I(x,y)=D_{KL}(p(x,y)\|p(x)p(y))$$

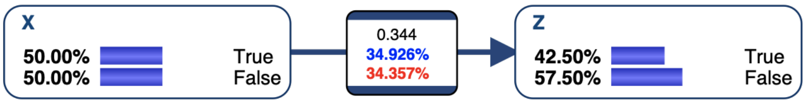
Therefore, Mutual Information (I) and KL Divergence are identical when there are no spouses (co-parents) implied in the measured relation.

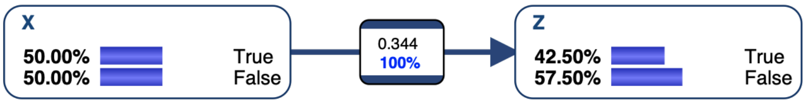
Example

Let's take the following network with two nodes *X* and *Z*.

X	True	False
True	75.000	25.000
False	10.000	90.000

The analysis of the relation with Mutual Information (**Validation Mode: Analysis | Visual | Arcs' Mutual Information**) and with KL (**Validation Mode: Analysis | Visual | Arc Force**) return the same value: 0.3436





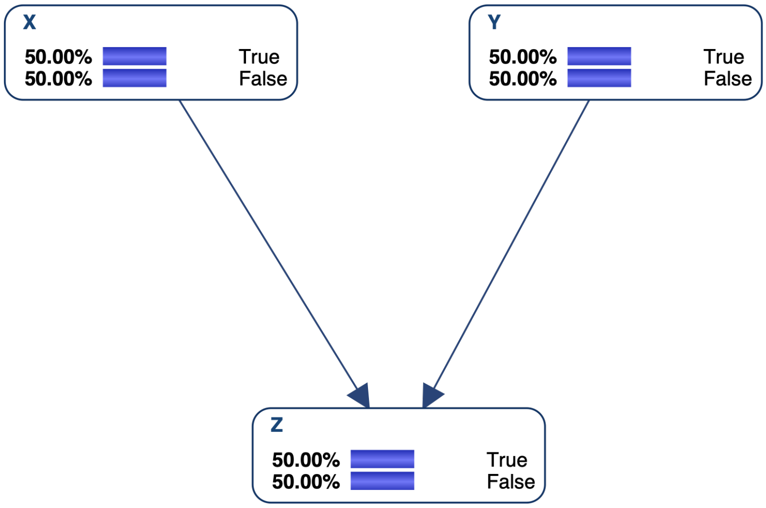
The percentage value in blue in the Mutual Information analysis corresponds to the Relative Mutual Information $$I_N(X,Z)=\frac{I(X,Z)}{H(Z)}$$ and the one in red corresponds to $$I_N(X,Z)=\frac{I(X,Z)}{H(X)}$$ where H() is the entropy defined as: $$H(X)=-\sum_{x\in X}p(x)

The percentage in blue in the Arc Force analysis is the relative weight of the link compared to the sum of all arc forces.

However, as soon as other variables are implied in the relation as co-parents, the KL Divergence will integrate them in the analysis, leading to a more precise result.

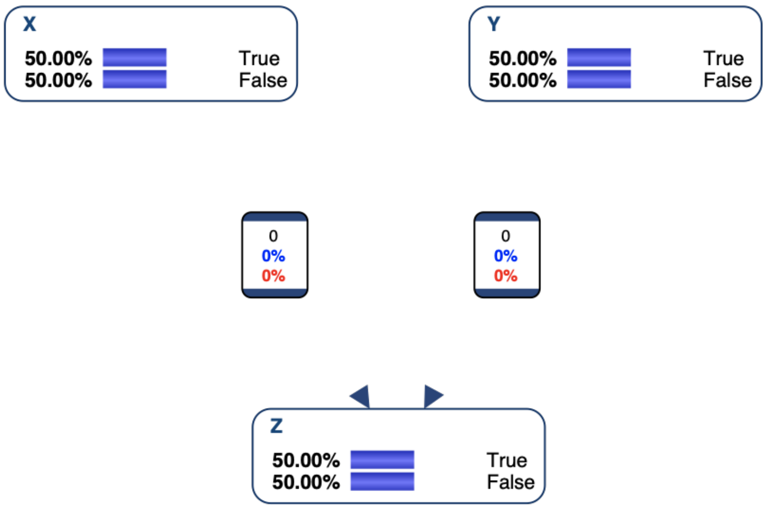
Example

Let's take the following deterministic example where *Z* is an Exclusive Or between *X* and *Y*, i.e. true when *X* and *Y* are different.

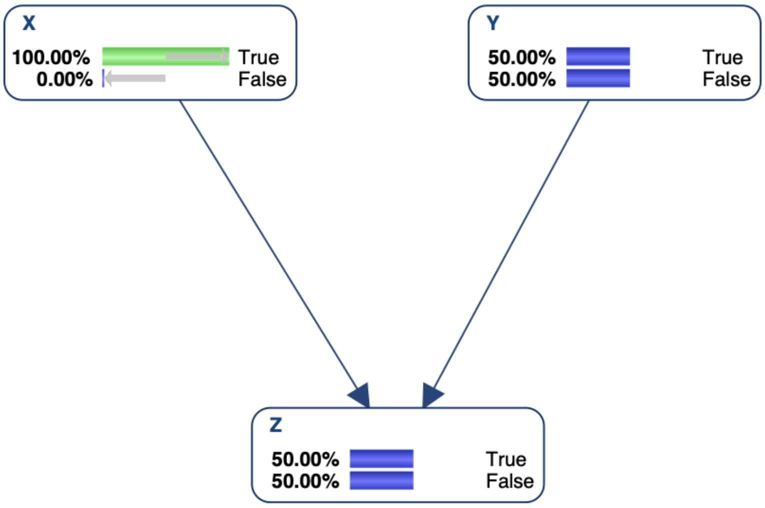


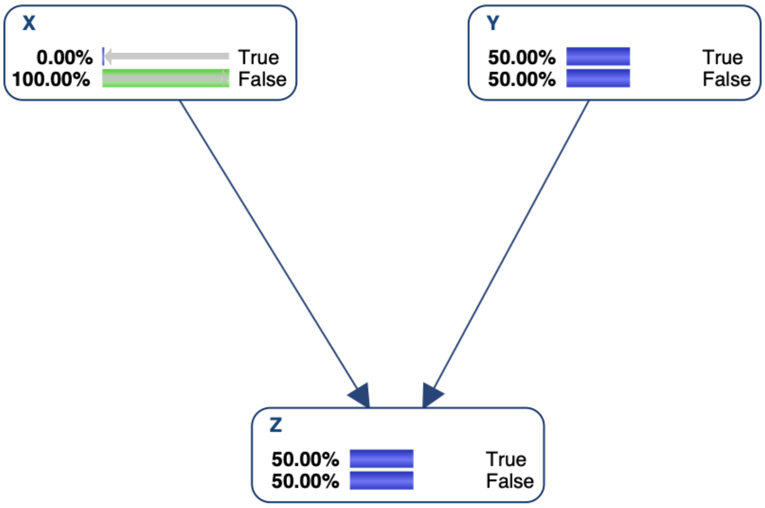
X	Y	Value
True	True	False
	False	True
False	True	True
	False	False

The analysis of the relations with Mutual Information (**Validation Mode: Analysis | Visual | Arcs' Mutual Information**) returns the following graph where the mutual information between *X* and *Z* and *Y* and *Z* are both null.

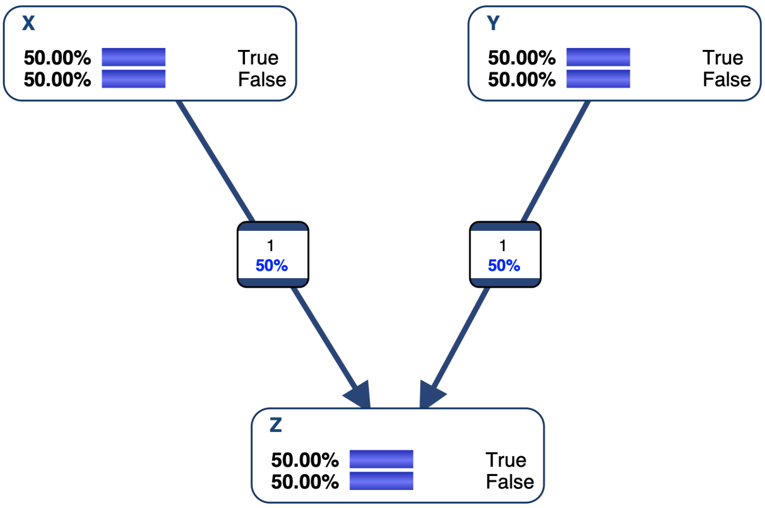


Indeed, X and Y do not have any impact on Z when they are analyzed separately.





On the other hand, the force of the arcs computed with KL (**Validation Mode: Analysis | Visual | Arc Force**) reflects perfectly the deterministic relation between *X* and *Y* on *Z*.



Q2: Relative Mutual Information

Two clones will have a Relative Mutual Information $I_N(X, X) = 1$ but not necessarily a Mutual Information $I(X, X)=1$. It depends on the value of the initial entropy $H(X)$. You will get it with a binary variable X that has a uniform marginal distribution.