The causal structure could be different from the best posterior probability structure

Prof Jan Kruger
Unisa graduate school for business leadership - MIDRAND
Business school/ finance case

• Keynote address presented at Finance conferences
• Causality used in Investment selection
• Causality can explain Financial distress
• Causality can be used to forecast VaR
• Causality is a tool to explain phenomena
Simplified Bayesian Structure

\[ P(\text{Age, Occupation, Weather, Disease, Symptoms}) = P(\text{Symptoms} | \text{Disease})P(\text{Disease} | \text{Age, Occupation, Weather})P(\text{Age})P(\text{Occupation})P(\text{Symptoms}) \]

(Buntine, 1996)

Prof Jan Kruger
Krugejw@unisa.ac.za
Cooper and Herskovits, 1989

\[ p(D \mid S^h) = \prod_{i=1}^{n} \prod_{j=1}^{q} \frac{\Gamma(\alpha_{ij})}{\Gamma(N_{ij} + \alpha_{ij})} \prod_{k=1}^{r} \frac{\Gamma(N_{ijk} + \alpha_{ijk})}{\Gamma(\alpha_{ijk})} \]
Number of structures

\[ \frac{n(n-1)}{3^2} \]
Steps

• Select a model
• Create a frequency distribution
• Do an exhaustive search over structures
• To find the max posterior probability structure
• It must be the same as the structure that created the data.
P(A) = 0.6855

P(B|A=0) = 0.8279
p(B|A=1) = 0.9930

P(B|A=0) = 0.8279
p(B|A=1) = 0.9930

P(D|B=0,C=0) = 0.5651
p(D|B=0,C=1) = 0.6899
p(D|B=1,C=0) = 0.5924
p(D|B=1,C=1) = 0.7412

P(C|B=0) = 0.1586
p(C|B=1) = 0.3666

P(E|D=0) = 0.6632
p(E|D=1) = 0.9798
Best posterior model is not the causal model

- The unscaled log posterior probability for the correct model is -23755. The models with the best unscaled log posterior probability of -23750 are $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E$, $A \leftarrow B \rightarrow C \rightarrow D \rightarrow E$, $A \leftarrow B \leftarrow C \rightarrow D \rightarrow E$, $A \leftarrow B \leftarrow C \leftarrow D \rightarrow E$ and $A \leftarrow B \leftarrow C \leftarrow D \leftarrow E$. This gives a posterior probability that is $e^5$ times larger for the wrong model.
New norm

Path between nodes A-B-C

Norm $\rho_{AB}\rho_{BC}\rho_{AC}$
Given variables A, B and C and assume only linear relationships, then the norm of the path (A-B-C) = 1 if and only if A and C are conditionally independent given B or I(A, C|B)

If the norm of the path A-B-C equals one, then
\[
\frac{\rho_{AB}\rho_{BC}}{\rho_{AC}} = 1 \quad \text{or} \quad \rho_{AB}\rho_{BC} = \rho_{AC}
\]

Only linear relationships are considered,
\[
\rho_{AC,B} = \frac{\rho_{AC} - \rho_{AB}\rho_{BC}}{\sqrt{(1 - \rho_{AB}^2)(1 - \rho_{BC}^2)}} = 0
\]

therefore A and C are conditionally independent given B or I(A, C|B).

The converse: If A and C are conditionally independent given B, I(A, C|B) then \(\rho_{AB}\rho_{BC} = \rho_{AC}\).
Elimination heuristic

• Start with a complete graph containing all the variables.
• Eliminate arcs between variables that are independent (with a correlation coefficient not significantly different from zero).
• Eliminate the arcs where Conditional Independence Statements have been identified (where the norm equals one).
• This gives us the causal tree structure, and I tested that it is also the structure with the highest posterior probability.
Theorem 1

• If two variables, X and W, are correlated and the path U-X-W is a valid path, then the path with X and W interchanged, U-W-X will have a norm equal to the coefficient of determination (the square of the correlation coefficient) between X and W.
If \( u-x-w \) valid path then norm of \( u-w-x = \rho^2 \)

\[
\text{Norm}(U-X-W) = 1 \\
\text{Norm}(U-W-X) = \text{Coeff. Det. Between} \\
\text{X and W}
\]
If u-x-w valid path then norm of u-w-x = $\rho^2$

- **Proof**: A valid (star decomposable) path U-X-W means that
- or, $\rho_{UW} = \rho_{UX} \cdot \rho_{XW}$
- Therefore the norm for the path U-W-X is
- This proves the theorem.
- The following theorem can be used to check if we have found a valid path.

Prof Jan Kruger
Krugejw@unisa.ac.za
Theorem 2

- $X-Y-Z$ is a valid and only path between $X$ and $Z$ if and only if the product of the norms is equal to the coefficient of determination between the start and end nodes of this path.
- $\text{Norm}(X-Y-Z).\text{Norm}(Y-X-Z).\text{Norm}(X-Z-Y) = \rho^2_{xz}$. 
Proof

• Assume a valid and only path X-Y-Z. Then, for a valid and only path there is a tree structure between the variables and the norm will be one (in other words \( \rho_{XY} \rho_{XZ} = \rho_{YZ} \)).

• Therefore

\[
\frac{\rho_{XY}}{\rho_{XZ}} \cdot \frac{\rho_{XZ}}{\rho_{YZ}} \cdot \rho_{XY} \rho_{YZ} = \rho_{XY} \rho_{XZ} \rho_{YZ} = \rho_{YZ}^2
\]

• In the opposite direction: if the product of the norms is equal to the coefficient of determination, then for the path A-E-C this condition can be written as:

\[
\frac{\rho_{AE} \rho_{EC}}{\rho_{AC}} \cdot \frac{\rho_{AC} \rho_{CE}}{\rho_{AE}} \cdot \frac{\rho_{AE} \rho_{AC}}{\rho_{EC}} = \rho_{AC}^2
\]

• If the correlation coefficient between A and C is not zero, then

\[
\therefore \rho_{EC} \rho_{AE} \rho_{AC} = \rho_{AC}^2
\]

or A-E-C is a valid and only path.