http://cacm.acm.org/blogs/blog-cacm/138907-john-mccarthy/fulltext
The Chain Rule

\[ P(X_1, X_2, \ldots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \ldots \]

- Trivial decomposition:
  \[ P(\text{Traffic, Rain, Umbrella}) = P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain, Traffic}) \]

- With assumption of conditional independence:
  \[ P(\text{Traffic, Rain, Umbrella}) = P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain}) \]

- Bayes’ nets / graphical models help us express conditional independence assumptions
Conditional Independence

• Reminder: independence
  – X and Y are independent if

\[ \forall x, y \ P(x, y) = P(x)P(y) \implies X \independent Y \]

  – X and Y are conditionally independent given Z

\[ \forall x, y, z \ P(x, y|z) = P(x|z)P(y|z) \implies X \independent Y|Z \]

  – (Conditional) independence is a property of a distribution
Independence in a BN

- Important question about a BN:
  - Are two nodes independent given certain evidence?
  - If yes, can prove using algebra (tedious in general)
  - If no, can prove with a counter example
  - Example:

    ![Diagram: X → Y → Z]

    - Question: are X and Z necessarily independent?
      - Answer: no. Example: low pressure causes rain, which causes traffic.
      - X can influence Z, Z can influence X (via Y)
      - Addendum: they could be independent: how?
Causal Chains

• This configuration is a “causal chain”

\[ P(x, y, z) = P(x)P(y|x)P(z|y) \]

– Is X independent of Z given Y?

\[ P(z|x, y) = \frac{P(x, y, z)}{P(x, y)} = \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)} = P(z|y) \]  

Yes!

– Evidence along the chain “blocks” the influence

X: Low pressure
Y: Rain
Z: Traffic
Common Cause

• Another basic configuration: two effects of the same cause
  – Are X and Z independent?
  – Are X and Z independent given Y?

\[
P(z|x, y) = \frac{P(x, y, z)}{P(x, y)} = \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)}
\]

\[
= P(z|y) \quad Yes!
\]

– Observing the cause blocks influence between effects.

Y: Project due
X: Newsgroup busy
Z: Lab full
Common Effect

- Last configuration: two causes of one effect (v-structures)
  - Are X and Z independent?
    - Yes: the ballgame and the rain cause traffic, but they are not correlated
    - Still need to prove they must be (try it!)
  - Are X and Z independent given Y?
    - No: seeing traffic puts the rain and the ballgame in competition as explanation?
  - This is backwards from the other cases
    - Observing an effect activates influence between possible causes.
The General Case

• Any complex example can be analyzed using these three canonical cases

• General question: in a given BN, are two variables independent (given evidence)?

• Solution: analyze the graph
Reachability

• Recipe: shade evidence nodes

• Attempt 1: if two nodes are connected by an undirected path not blocked by a shaded node, they are conditionally independent

• Almost works, but not quite
  – Where does it break?
  – Answer: the v-structure at T doesn’t count as a link in a path unless “active”
Reachability (D-Separation)

• Question: Are X and Y conditionally independent given evidence vars \( \{Z\} \)?
  – Yes, if X and Y “separated” by Z
  – Look for active paths from X to Y
  – No active paths = independence!

• A path is active if each triple is active:
  – Causal chain \( A \rightarrow B \rightarrow C \) where B is unobserved (either direction)
  – Common cause \( A \leftarrow B \rightarrow C \) where B is unobserved
  – Common effect (aka v-structure) \( A \rightarrow B \leftarrow C \) where B or one of its descendents is observed

• All it takes to block a path is a single inactive segment
Example

\[ R \perp B \quad \text{Yes} \]
\[ R \perp B | T \]
\[ R \perp B | T' \]
Example

• Variables:
  – R: Raining
  – T: Traffic
  – D: Roof drips
  – S: I’m sad

• Questions:
  \[ T \perp D \]
  \[ T \perp D|R \]
  \[ T \perp D|R, S \] Yes
Causality?

- When Bayes’ nets reflect the true causal patterns:
  - Often simpler (nodes have fewer parents)
  - Often easier to think about
  - Often easier to elicit from experts

- BNs need not actually be causal
  - Sometimes no causal net exists over the domain
  - E.g. consider the variables Traffic and Drips
  - End up with arrows that reflect correlation, not causation

- What do the arrows really mean?
  - Topology may happen to encode causal structure
  - Topology only guaranteed to encode conditional independence
Example: Traffic

- Basic traffic net
- Let’s multiply out the joint

\[ P(R) \]
- \( r \): \( 1/4 \)
- \( \neg r \): \( 3/4 \)

\[ P(T|R) \]
- \( r \): \( t \): \( 3/4 \)
- \( r \): \( \neg t \): \( 1/4 \)
- \( \neg r \): \( t \): \( 3/4 \)
- \( \neg r \): \( \neg t \): \( 1/4 \)

\[ P(T, R) \]
- \( r \): \( t \): \( 3/16 \)
- \( r \): \( \neg t \): \( 1/16 \)
- \( \neg r \): \( t \): \( 6/16 \)
- \( \neg r \): \( \neg t \): \( 6/16 \)
Example: Reverse Traffic

• Reverse causality?

\[
\begin{array}{c|c|c}
\text{t} & 9/16 \\
\text{t} & 7/16 \\
\text{r} & 1/3 \\
\text{r} & 2/3 \\
\text{r} & 1/7 \\
\text{r} & 6/7 \\
\hline
\text{r} & 3/16 \\
\text{r} & 1/16 \\
\text{r} & 6/16 \\
\text{r} & 6/16 \\
\end{array}
\]
Example: Coins

• Extra arcs don’t prevent representing independence, just allow non-independence

- Adding unneeded arcs isn’t wrong, it’s just inefficient
Changing Bayes’ Net Structure

• The same joint distribution can be encoded in many different Bayes’ nets
  – Causal structure tends to be the simplest

• Analysis question: given some edges, what other edges do you need to add?
  – One answer: fully connect the graph
  – Better answer: don’t make any false conditional independence assumptions
Example: Alternate Alarm

If we reverse the edges, we make different conditional independence assumptions.

To capture the same joint distribution, we have to add more edges to the graph.
Summary

• Bayes nets compactly encode joint distributions

• Guaranteed independencies of distributions can be deduced from BN graph structure

• D-separation gives precise conditional independence guarantees from graph alone

• A Bayes’ net’s joint distribution may have further (conditional) independence that is not detectable until you inspect its specific distribution