Conditional Independence and Irrelevance

Philip Dawid

University of Cambridge

Royal Netherlands Academy of Arts and Sciences Colloquium
"DEPENDENCE LOGIC"

3–5 March 2014

Probabilistic Independence

- Base space Ω
  - discrete — for simplicity (only)
- Random variables $X, Y, Z, \dots$
  - functions on Ω
- Probability distribution P on Ω

Marginal/conditional distributions:

$$\begin{aligned} p(x) &:= P(X = x) \\ p(x \mid y) &:= P(X = x \mid Y = y) \\ &= p(x, y)/p(y) \end{aligned}$$

[defined when $p(y) > 0$]

Probabilistic Independence

Say X is independent of Y, and write $X \perp\!\!\!\perp_p Y \ [P]$, or $X \perp\!\!\!\perp Y$, if

$$
\begin{aligned}
p(x, y) &\equiv p(x)\, p(y) \\
p(x, y) &\equiv a(x)\, b(y)
\end{aligned}
$$

- "No interaction" between X and Y

$$
\begin{aligned}
p(x \mid y) &\equiv p(x) \\
p(x \mid y) &\equiv a(x)
\end{aligned}
$$

- Information about Y is *irrelevant* to uncertainty about X.

Conditional Independence

Say X is (conditionally) independent of Y, given Z, and write $X \perp\!\!\!\perp_p Y \mid Z \; [P]$, or $X \perp\!\!\!\perp Y \mid Z$, if

$$\begin{aligned} p(x, y \mid z) &\equiv p(x \mid z)\, p(y \mid z) \\ p(x, y \mid z) &\equiv a(x, z)\, b(y, z) \\ p(x, y, z) &\equiv p(x \mid z)\, p(y \mid z) p(z) \\ p(x, y, z) &\equiv p(x, z)\, p(y, z)/p(z) \\ p(x, y, z) &\equiv a(x, z)\, b(y, z) \end{aligned}$$

- Once Z is known, no further interaction between X and Y

$$\begin{aligned} p(x \mid y, z) &\equiv p(x \mid z) \\ p(x \mid y, z) &\equiv a(x, z) \end{aligned}$$

- Once Z is known, any further information about Y is *irrelevant* to uncertainty about X
- X depends on Z (and not on Y)
  - satisfies dependence logic???

Fundamental Properties [1, 3, 6]

Writing $W \preceq Y$ to mean that W is a function of Y:

$$\text{P1} \quad : \quad X \perp\!\!\!\perp Y \mid Z \quad \Rightarrow \quad Y \perp\!\!\!\perp X \mid Z$$

$$\text{P2} \quad : \quad X \perp\!\!\!\perp Y \mid X$$

$$\text{P3} \quad : \quad \left.\begin{array}{c} X \perp\!\!\!\perp Y \mid Z \\ \text{and} \\ W \preceq Y \end{array}\right\} \quad \Rightarrow \quad X \perp\!\!\!\perp W \mid Z$$

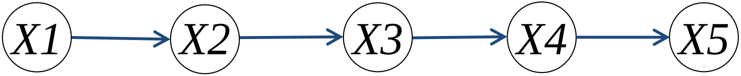
$$\text{P4} \quad : \quad \left.\begin{array}{c} X \perp\!\!\!\perp Y \mid Z \\ \text{and} \\ W \preceq Y \end{array}\right\} \quad \Rightarrow \quad X \perp\!\!\!\perp Y \mid (W, Z)$$

$$\text{P5} \quad : \quad \left.\begin{array}{c} X \perp\!\!\!\perp Y \mid Z \\ \text{and} \\ X \perp\!\!\!\perp W \mid (Y, Z) \end{array}\right\} \quad \Rightarrow \quad X \perp\!\!\!\perp (Y, W) \mid Z.$$

- Interpretations in terms of "irrelevance"

Use as Axioms

"Nearest neighbour" property of a Markov Chain:



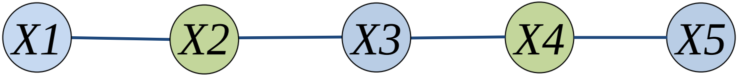
Suppose:

(i). $X_3 \perp\!\!\!\perp X_1 \mid X_2$

(ii). $X_4 \perp\!\!\!\perp (X_1, X_2) \mid X_3$

(iii). $X_5 \perp\!\!\!\perp (X_1, X_2, X_3) \mid X_4$

Then $X_3 \perp\!\!\!\perp (X_1, X_5) \mid (X_2, X_4)$.



Proof

Applying P4 and P1 in turn to (ii), we obtain

$$X_1 \perp\!\!\!\perp X_4 \mid (X_2, X_3), \tag{1}$$

while from (i) and P1 we have

$$X_1 \perp\!\!\!\perp X_3 \mid X_2. \tag{2}$$

On applying P5 to (2) and (1), we now deduce

$$X_1 \perp\!\!\!\perp (X_3, X_4) \mid X_2 \tag{3}$$

whence, by P4 and P1,

$$X_3 \perp\!\!\!\perp X_1 \mid (X_2, X_4). \tag{4}$$

Also, by (iii) and P4 we have

$$X_5 \perp\!\!\!\perp (X_1, X_3) \mid (X_2, X_4) \tag{5}$$

and so, by P4 and P1,

$$X_3 \perp\!\!\!\perp X_5 \mid (X_1, X_2, X_4). \tag{6}$$

The result now follows on applying P5 to (4) and (6).

An additional axiom? [12, 13]

If

- $X \perp\!\!\!\perp Y \mid (Z, W)$
- $Z \perp\!\!\!\perp W \mid X$
- $Z \perp\!\!\!\perp W \mid Y$
- $X \perp\!\!\!\perp Y$

then

- $Z \perp\!\!\!\perp W \mid (X, Y)$
- $X \perp\!\!\!\perp Y \mid Z$
- $X \perp\!\!\!\perp Y \mid W$
- $Z \perp\!\!\!\perp W$

Probabilistic conditional independence has no finite axiomatisation.

Variation Independence [8]

Now no probability distribution.

The *range*, $R(X)$, of X is $X(\Omega) = \{X(\omega) : \omega \in \Omega\}$.

The *conditional range*, $R(X \mid y)$, of X, given $Y = y$, is $\{X(\omega) : \omega \in \Omega,\, Y(\omega) = y\}$ [for $y \in R(Y)$].

We say X is variation independent of Y given Z, and write $X \perp\!\!\!\perp_v Y \mid Z \; [\Omega]$, if:

$$
\begin{aligned}
R(X, Y \mid z) &\equiv R(X \mid z) \times R(Y \mid z) \\
R(X, Y \mid z) &\equiv A(z) \times B(z)
\end{aligned}
$$

$$
\begin{aligned}
R(X \mid y, z) &\equiv R(X \mid z) \\
R(X \mid y, z) &\equiv A(z)
\end{aligned}
$$

- *"Embedded Multivalued Dependency"*
- The general properties P1–P5 also hold for $\perp\!\!\!\perp_v$.
  - but not the "additional axiom"
- Other additional properties hold in this case
- No finite axiomatisation

Specialisation:

Ω indexes the rows in a team, and each variable X is a column in the team.

Abstraction: The Separoid [9]

Join semilattice $(\mathcal{S}, \leq)$: $x \vee y$ is least upper bound of $\{x, y\}$.
A ternary relation $\perp\!\!\!\perp$ on $\mathcal{S}$ is a separoid if:

$$\text{P1} : \quad x \perp\!\!\!\perp y \mid z \quad \Rightarrow \quad y \perp\!\!\!\perp x \mid z$$

$$\text{P2} : \quad x \perp\!\!\!\perp y \mid x$$

$$\text{P3} : \quad \left. \begin{array}{c} x \perp\!\!\!\perp y \mid z \\ \text{and} \\ w \leq y \end{array} \right\} \quad \Rightarrow \quad x \perp\!\!\!\perp w \mid z$$

$$\text{P4} : \quad \left. \begin{array}{c} x \perp\!\!\!\perp y \mid z \\ \text{and} \\ w \leq y \end{array} \right\} \quad \Rightarrow \quad x \perp\!\!\!\perp y \mid (w \vee z)$$

$$\text{P5} : \quad \left. \begin{array}{c} x \perp\!\!\!\perp y \mid z \\ \text{and} \\ x \perp\!\!\!\perp w \mid (y \vee z) \end{array} \right\} \quad \Rightarrow \quad x \perp\!\!\!\perp (y \vee w) \mid z.$$

Strong Separoid

Now $(\mathcal{S}, \leq)$ is a *lattice*, and in addition to P1–P5, we require:

$$\text{P6} \quad : \quad \left.\begin{array}{l} x \perp\!\!\!\perp y \mid z \\ x \perp\!\!\!\perp z \mid y \end{array}\right\} \;\Rightarrow\; x \perp\!\!\!\perp (y \vee z) \mid (y \wedge z)$$

This holds universally for variation independence, but only under additional conditions (e.g., that every elementary outcome has positive probability) for probabilistic independence.

Graphical Models [4]

Let $\mathcal{G}$ be an undirected graph with node set N. Take $(\mathcal{S}, \leq)$ to be the collection of subsets of N, ordered by inclusion. For $A, B, C \subseteq N$, write $A \perp\!\!\!\perp_{ug} B \mid C \; [\mathcal{G}]$ if every path from a node in A to one in B intersects C.

This relation defines a strong separoid.

Now let $\mathcal{D}$ be a directed acyclic graph on N. For $M \subseteq N$, $\mathrm{an}(M)$, the *ancestral graph* of M, is the subgraph of $\mathcal{D}$ induced by M together with all nodes from which a path leads into M. The *moralization* $\mathcal{D}_m$ of $\mathcal{D}$ is the undirected graph obtained from $\mathcal{D}$ by adding (if necessary) an edge between each pair of nodes that both point to the same node, and then ignoring directions.

For $A, B, C \subseteq N$, write $A \perp\!\!\!\perp_{dg} B \mid C \; [\mathcal{D}]$ if

$$A \perp\!\!\!\perp_{ug} B \mid C \quad [\{\mathrm{an}(A \cup B \cup C)\}_m].$$

Again, this relation defines a strong separoid.

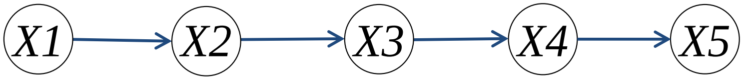
Theorem Proving Machine

- Sometimes we can represent a collection of conditional independence properties as exactly those described by graph separation, in a suitable undirected graph or DAG.
- Then we can see which properties hold just by inspection of the graph.
- Given some "input collection" of conditional independence properties we may be able to construct a graph displaying these — and then simply read off further, implied, conditional independence properties by inspection of the graph.
  - *e.g.*, for a recursive collection of the form

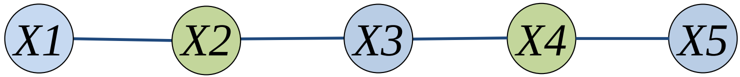
$$X_i \perp\!\!\!\perp (X_1, \ldots, X_{i-1}) \mid S_i \quad (i = 1, \ldots, N),$$

    with $S_i \subseteq (X_1, \ldots, X_{i-1})$, the relevant DAG has arrows into X_i from each member of S_i.

Markov Chain



Original directed acyclic graph



Moralised ancestral graph

Example: Criminal Case [5]

Example: Criminal Case

EYE WITNESS EVIDENCE

- An **unknown number of offenders** entered commercial premises late at night through a hole which they cut in a metal grille.
- Inside, they were confronted by a **security guard** who was able to set off an alarm before one of the intruders punched him in the face, causing his **nose to bleed**.
- **The security guard said that there were four men** but the light was too poor for him to describe them and he was confused because of the blow he had received.
- About 10 minutes later the police found **the suspect** trying to "hot wire" a car in an alley about a quarter of a mile from the incident. The suspect denied having anything to do with it.

Example: Criminal Case

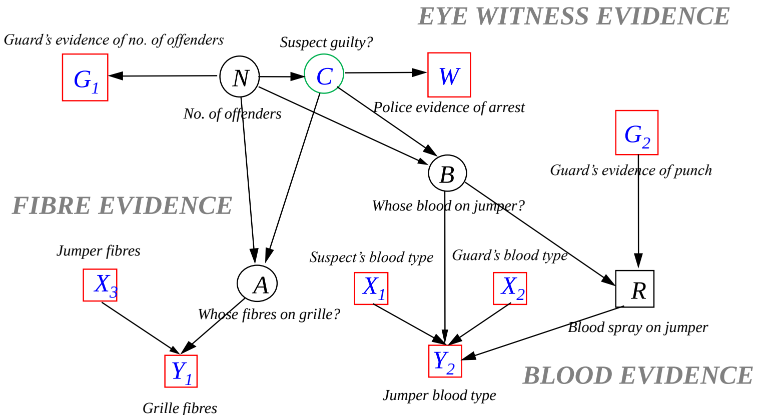
FIBRE EVIDENCE

- A tuft of **red acrylic fibres** was found on the jagged end of one of the cut edges of the grille.
- The suspect's **jumper** was **red acrylic**. The **tuft** was **indistinguishable** from the fibres of the **jumper** by eye, microspectrofluorimetry and thin layer chromatography.

BLOOD EVIDENCE

- A **spray pattern of blood** was found on the front and right sleeve of the suspect's **jumper**.
- The blood on the **jumper** was of a **different** type from that of the **suspect**, but the **same** as that of the security **guard**.

Example: Criminal Case



Moralization: 1

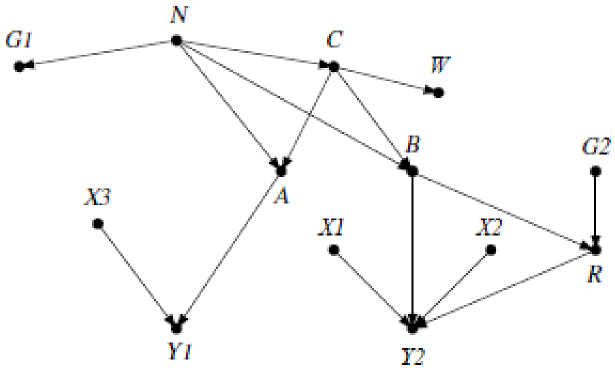


Figure 6.1: Directed graph $\mathcal{D}$ for criminal evidence

Moralization: 1

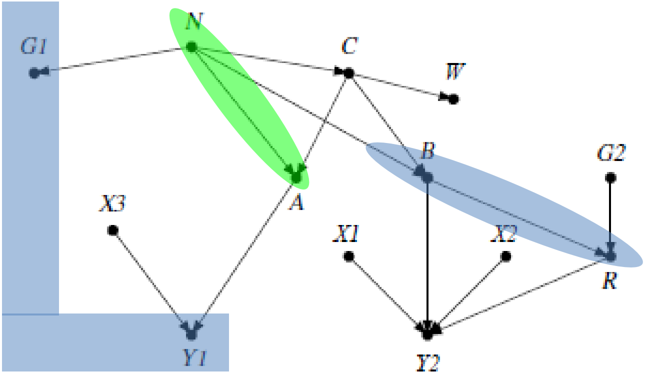


Figure 6.1: Directed graph $\mathcal{D}$ for criminal evidence

$$(B, R) \perp\!\!\!\perp (G1, Y1) \mid (A, N) \ ?$$

Moralization: 2

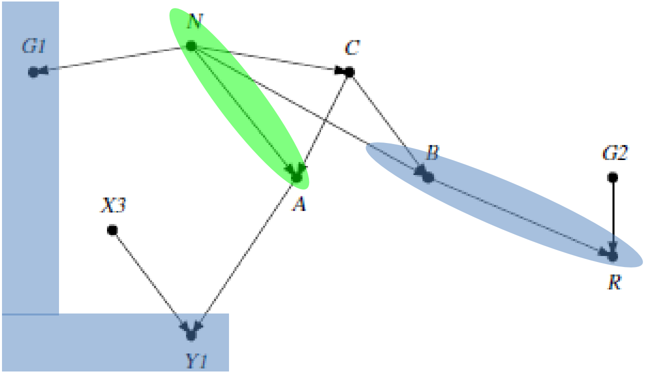


Figure 6.2: Ancestral subgraph $\mathcal{D}'$

$$(B, R) \perp\!\!\!\perp (G1, Y1) \mid (A, N) \ ?$$

Moralization: 3

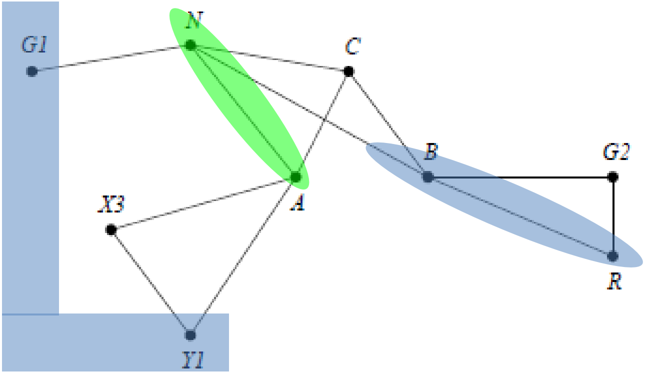


Figure 6.3: Moralized ancestral subgraph $\mathcal{G}'$

$$(B, R) \perp\!\!\!\perp (G1, Y1) \mid (A, N) \ ?$$

Possibility Function

Let $(\mathcal{S}, \leq)$ be a join semilattice.

- $\phi : \mathcal{S} \to \mathbb{R}$ is a *possibility function* if

$$\phi(x \vee y) = \max\{\phi(x), \phi(y)\}.$$

- Define $\perp\!\!\!\perp_d$ by:

$$x \perp\!\!\!\perp_d y \mid z \Leftrightarrow \min\{\phi(x), \phi(y)\} \leq \phi(z).$$

- Then $\perp\!\!\!\perp_d$ defines a separoid on $\mathcal{S}$.

Natural Independence [11]

- $\kappa : \Omega \to \{0, 1, \ldots\}$
- For $A, B \subseteq \Omega$, define

$$\begin{aligned} \kappa(A) &:= \min\{\kappa(\omega) : \omega \in A\} \\ \kappa(A \mid B) &:= \kappa(A \cap B) - \kappa(B) \end{aligned}$$

  (natural conditional function / implausibility function)

- $-\kappa$ is a possibility function on 2^Ω
  - (all we really need)

For functions X, Y, Z on Ω, write $X \perp\!\!\!\perp_N Y \mid Z\ [\kappa]$ if

(i). $X \perp\!\!\!\perp_v Y \mid Z\ [\Omega]$, and

(ii). $\kappa(x, y, z) \equiv \kappa(x, z) + \kappa(y, z) - \kappa(z)$
$\kappa(x \mid y, z) \equiv \kappa(x \mid z)$

Then $\perp\!\!\!\perp_N$ defines a strong separoid.

Further Examples: Mathematics

- Lattice theory:
  - distributivity
  - modularity
  - semimatroid
- Vector spaces:
  - linear independence
  - orthogonality

Further Examples: Statistics and AI

- Statistical models:
  - statistical independence $X \perp\!\!\!\perp_S Y \mid Z \; [\mathcal{P}]$:

$$X \perp\!\!\!\perp_p Y \mid Z \; [P], \text{ all } P \in \mathcal{P}$$

  - meta independence $X \perp\!\!\!\perp_\mu Y \mid Z \; [\mathcal{P}]$:

$$X \perp\!\!\!\perp_S Y \mid Z \; [\mathcal{P}] \text{ and } P_{(X,Z)} \perp\!\!\!\perp_v P_{(Y,Z)} \mid P_Z \; [\mathcal{P}]$$

  - hyper independence $X \perp\!\!\!\perp_H Y \mid Z \; [\Pi]$:

$$X \perp\!\!\!\perp_\mu Y \mid Z \; [\mathcal{P}] \text{ and } P_{(X,Z)} \perp\!\!\!\perp_p P_{(Y,Z)} \mid P_Z \; [\Pi]$$

- Covariance independence
- Belief independence

- Non-symmetric independence
- Causal dependence

References I

Dawid, A. P. (1979). Conditional independence in statistical theory (with Discussion). *J. Roy. Statist. Soc.* B **41**, 1–31.

Dawid, A. P. (1979). Some misleading arguments involving conditional independence. *J. Roy. Statist. Soc.* B **41**, 249–252.

Dawid, A. P. (1980). Conditional independence for statistical operations. *Ann. Statist.* **8**, 598–617.

Lauritzen, S. L., Dawid, A. P., Larsen, B. N. and Leimer, H. G. (1990). Independence properties of directed Markov fields. *Networks* **20**, 491–505.

Dawid, A. P. and Evett, I. W. (1997). Using a graphical method to assist the evaluation of complicated patterns of evidence. *J. Forensic Sci.* **42**, 226–231.

References II

- Dawid, A. P. (1998). Conditional independence. *Encyclopedia of Statistical Sciences*, Update Volume 2, edited by S. Kotz, C. B. Read and D. L. Banks. Wiley-Interscience, 146–155.
- Dawid, A. P. and Studený, M. (1999). Conditional products: An alternative approach to conditional independence. In *Artificial Intelligence and Statistics 99*, edited by D. Heckerman and J. Whittaker. Morgan Kaufmann, 32–40.
- Dawid, A. P. (2001). Some variations on variation independence. In *Artificial Intelligence and Statistics 2001*, edited by T. Jaakkola and T. Richardson. Morgan Kaufmann, 187–191.
- Dawid, A. P. (2001). Separoids: A mathematical framework for conditional independence and irrelevance. *Ann. Math. Artificial Intelligence* **32**, 335–372.

References III

Dawid, A. P. (2002). Influence diagrams for causal modelling and inference. *Intern. Statist. Rev.* **70**, 161–189. Corrigenda, *ibid*., 437.

Spohn, W. (1988). Ordinal conditional functions: a dynamic theory of epistemic states. In *Causation in Decision, Belief Change, and Statistics*, eds. W. L. Harper and B. Skyrms. Kluwer Academic, Dordrecht, 105–134.

Studený, M. (1989). Multi-information and the problem of characterization of conditional independence relations. *Problems Control and Inf. Theory* **18**, 3–16.

Studený, M (1992). Conditional independence relations have no finite complete characterization. *Trans. Eleventh Prague Conf. Inf. Theory Statist. Decision Functions and Random Processes B.* Academia, Prague, 377–396.