Conditional independence

Nooitgedagt, H.

H. Nooitgedagt

Conditional Independence

Bachelorscriptie, 18 augustus 2008

Scriptiebegeleider: prof.dr. R. Gill
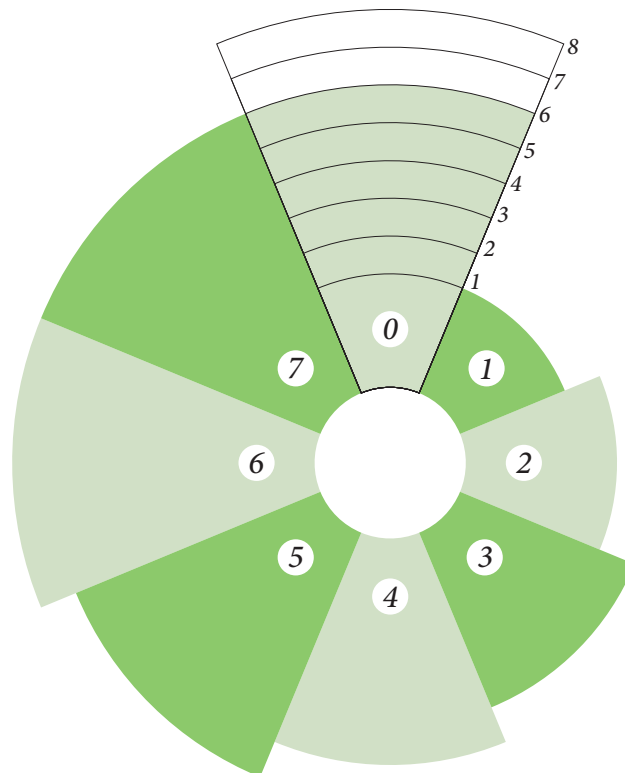
Preface

In this thesis I'll discuss Conditional Independencies of Joint Probability Distributions (here after called CI's respectively JPD's) over a finite set of discrete random variables. Remember that for any such JPD we can write down a list of all CI's, between two subsets of variables given a third. Such a list is called a CI-*trace*. An arbitrary list of CI's is called a CI-*pattern*, without a priori knowing if there will exist a corresponding JPD with this CI-pattern.

For simplicity and without loss of generality we take all JPD's over $n + 1$ variables and label them by the integers $0, 1, \ldots, n$. A CI-trace now becomes a set of triples consisting of subsets of $[n]$, the random variables (with $[n]$ I denote the set $\{0, 1, \ldots, n\}$). For example (A, B, C) with A, B and $C \subset [n]$ is such a triple, it can also be denoted as $A \perp B | C$, which means that the random variables of A are independent of the random variables of B given any outcome on the random variables of C.

It was believed that CI-traces could be characterised by some finite set of rules, called Conditional Independence rules, CI-rule.

Such a CI-rule would state that if a CI-trace contains a certain pattern of triplets it should also contain a certain other triple. Furthermore such a pattern of a CI-rule should itself be finite; it should consist of k CI's, called the antecedents that would validate another $k + 1$'th CI, called the consequent. The *order* of a CI-rule is the number k of its antecedents.

This idea would imply that the set of all CI-traces is equal to the set of all CI-patterns closed under the CI-rules. In 1992 Milan Studený wrote an article on this subject called *Conditional Independence Relations have no finite complete characterisation*. He proved that such a characterisation is not possible. Now the main goal of my thesis was to understand this article and to work out a readable version of the theorem and the proof.

The proof is based on two major parts. First of all the existence of a particular JPD and its CI-pattern on $n + 1$ variables and secondly on a proposition about CI-patterns based on entropies. The remainder of my thesis will contain sections on these two major parts, Studený's theorem and a small summary of the changes I made.

Contents

1 Introduction

In this thesis I will be discussing joint probability distributions, JPD's, over $n + 1$ discrete random variables, $X_0, X_1, \ldots X_n$, taking a finite value each. W.l.o.g. we can label these random variables with the set $[n] = \{0, 1 \ldots, n\}$. Let X_i denote the random variable with index i and suppose it takes values in $\mathcal{E}_i$, which is a finite non-empty set. For $A \subset [n]$, X_A will denote the random vector $(X_j : j \in A)$. If $A = \emptyset$ then it will be the degenerate random variable taking a fixed value, denoted by $x_\emptyset$. X will be the short notation for $X_{[n]}$. The random vector X takes values in $\mathcal{E} = \mathcal{E}_{[n]} = \times_{i \in [n]} \mathcal{E}_i$. A generic value of X_A is denoted by x_A. The JPD of X is denoted by P. The class of all JPD's over $n + 1$ random variables will be denoted by $\mathcal{P}([n])$

With $T([n])$ I shall denote the *triplet set*, of $[n]$, which is the set of all ternary-tuples $u = (A, B, C)$ such that A, B and C are disjoint subsets of $[n]$. We define the context of a triple u as $[u] = A \cup B \cup C$. A triple $u = (A, B, C)$ is called a *trivial triple* if and only if A and/or B is empty. The set of non-trivial triples is denoted by $T_*([n])$.

Now we can give the following definitions,

Definition 1.1 [1] *Let P be a JPD over $n+1$ random variables $X_0, X_1, \ldots, X_n$ and $A, B, C \subset [n]$. The set X_A is said to be conditional independent of X_B given X_C, shortly written as $X_A \perp X_B | X_C$, if and only if,*

$$P(X_A = x_A | X_B = x_B, X_C = x_C) = P(X_A = x_A | X_C = x_C),$$

whenever $P(X_B = x_B, X_C = x_C) > 0$. Note that this can also be written as

$$P(X_A = x_A, X_B = x_B, X_C = x_C)P(X_C = x_C) =$$
$$P(X_B = x_B, X_C = x_C)P(X_A = x_A, X_C = x_C).$$

For all $x_A \in \mathcal{E}_A$, $x_B \in \mathcal{E}_B$ and $x_C \in \mathcal{E}_C$.

In the remainder of the thesis I will abuse notation and use the following notation,

$$p(x_A, x_B, x_C) = P(X_A = x_A, X_B = x_B, X_C = x_C),$$
$$p(x_A | x_B, x_C) = P(X_A = x_A | X_B = x_B, X_C = x_C),$$
$$p(x_A | x_C) = P(X_A = x_A | X_C = x_C),$$
$$\text{etc...}$$

---

[1]See *Causality*, by Pearl, page 11. To fit this thesis small changes in notation have been made.

Definition 1.2 *Given a JPD, P, over $n+1$ random variables, $X_0, X_1, \ldots, X_n$, labelled by $[n]$. I define its corresponding CI-trace, $I_P \subset T([n])$ as*

$$(A, B, C) \in I_P \iff X_A \perp X_B | X_C,$$

with respect to the JPD. $X_A \perp X_B | X_C$ can shortly be written as $A \perp B | C$.

As mentioned in the preface, it was long believed that taking an arbitrary set $I \in T([n])$ and closing it under some finite set of CI-rules, would give a CI-traceof some JPD.

A CI-rule contains a certain pattern of k triples of $T([n])$ and has exactly one triple as consequent. The random variables represented by the sets in the consequent can only contain random variables that were also represented by the sets in the antecedents. The triples contained in the antecedents are always non-trivial, furthermore none of the CI-rules is a deducible from previous CI-rules. The order of a CI-rule, $\mathcal{R}$, is equal to the number of its k antecedents, denoted by $\mathsf{ord}(\mathcal{R}) = k$. For example some CI-rules are listed below

Symmetry	$A \perp B | C$	$\Rightarrow$	$B \perp A | C,$
Contraction	$A \perp B | C$ & $A \perp D | (B, C)$	$\Rightarrow$	$A \perp (B, D) | C,$
Decomposition	$(A, D) \perp B | C$	$\Rightarrow$	$D \perp B | C,$
Weak Union	$A \perp (B, D) | C$	$\Rightarrow$	$A \perp B | (C, D),$

with respectively order 1,1,2 and 1.

An important operator in the remainder of this thesis will be the operator successor. Successor is defined on $[n] \setminus \{0\}$ as follows

$$\mathsf{suc}(j) = \left\{ \begin{array}{ll} j + 1 & \text{if } j \in [n] \setminus \{0, n\}, \\ 1 & \text{if } j = n, \end{array} \right\}$$

i.e., the cyclic permutation on $[n] \setminus \{0\}$.

2 Proposition

2.1 Entropies

The theory of entropies was created to measure the mean amount of information gained from some random variable X after obtaining its actual value, i.e. we average over the possible values of X. We measure this at the hand of the random variable probabilities rather than on its actual outcome. (Note that in this thesis we will only discuss discrete cases.) The *Shannon entropy* of a discrete random variable X will be defined as:

$$H(X) := -\sum_x p(x) \log p(x)$$

with $\log := \log_2$ and by convention $0 \log 0 := 0$. For this thesis in particular we shall need the notion of the *joint entropy*. The *joint entropy* over two random variables X and Y is defined as:

$$H(X,Y) := -\sum_{x,y} p(x,y) \log p(x,y),$$

and can be extended to any number of random variables in the obvious way. The joint entropy measures the total uncertainty of the pair X, Y. Now suppose we know the value of Y and hence know $H(Y)$ bits of information of the pair (X, Y). The remaining uncertainty is due to X, given what we know about Y. The entropy of X conditional on knowing Y is therefore defined as:

$$
\begin{aligned}
H(X \mid Y) &:= H(X,Y) - H(Y) \\
&= -\sum_{x,y} p(x,y) \log p(x,y) + \sum_y p(y) \log p(y) \\
&= -\sum_{x,y} p(x,y) \log p(x,y) + \sum_{x,y} p(x,y) \log p(y) \\
&= -\sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(y)}
\end{aligned}
$$

Finally the *mutual entropy* of X and Y which measures the information that X and Y have in common is defined as:

$$H(X : Y) := H(X) + H(Y) - H(X,Y)$$

$$= -\sum_x p(x) \log p(x) - \sum_y p(y) \log p(y)$$

$$+ \sum_{x,y} p(x,y) \log p(x,y)$$

$$= -\sum_{x,y} p(x,y) \log p(x) - \sum_{x,y} p(x,y) \log p(y)$$

$$+ \sum_{x,y} p(x,y) \log p(x,y)$$

$$= -\sum_{x,y} p(x,y) \log \frac{p(y)p(y)}{p(x,y)},$$

where we subtract the joint information of the pair (X, Y) and now the difference is the information on X and Y that has been counted twice and is therefore called the mutual information on X and Y.

The following theorem states a few features of the Shannon Entropy, only the sixth statement will be proven because it will be used in the proof of lemma 2.2 on page 5.

Theorem 2.1 (Properties of the Shannon Entropy) *Let X, Y, Z be random variables then we have,*

1. $H(X, Y) = H(Y, X), H(X : Y) = H(Y : X)$.

2. $H(Y \mid X) \geq 0$ and thus $H(X : Y) \leq H(Y)$, with equality iff Y is a function of X, $Y = f(X)$.

3. $H(X) \leq H(X, Y)$, with equality iff Y is a function of X.

4. **Sub-additivity**: $H(X, Y) \leq H(X) + H(Y)$ with equality iff X and Y are independent random variables.

5. $H(X \mid Y) \leq H(X)$ and thus $H(X : Y) \geq 0$, with equality iff X and Y are independent random variables.

6. **Strong sub-additivity**: $H(X, Y, Z) + H(Y) \leq H(X, Y) + H(Y, Z)$, with equality iff Z is conditional independent of X given Y.

7. **Conditioning reduces entropy**: $H(X \mid Y, Z) \leq H(X \mid Y)$.

Proof of 6

To prove this statement we use the fact that $-\log x \geq (1 - x)/\ln 2$, with

equality if and only if $x + 1$ and $\ln := \log_2$. We get

$$H(X, Y) + H(Y, Z) - H(X, Y, Z) - H(Y)$$

$$= -\sum_{x,y} p(x, y) \log p(x, y) - \sum_{y,z} p(y, z) \log p(y, z)$$

$$+ \sum_{x,y,z} p(x, y, z) \log p(x, y, z) + \sum_{y} p(y) \log p(y)$$

$$= -\sum_{x,y,z} p(x, y, z) \log p(x, y) - \sum_{x,y,z} p(x, y, z) \log p(y, z)$$

$$+ \sum_{x,y,z} p(x, y, z) \log p(x, y, z) + \sum_{x,y,z} p(x, y, z) \log p(y)$$

$$= -\sum_{x,y,z} p(x, y, z) \log \frac{p(x, y) p(y, z)}{p(x, y, z) p(y)}$$

$$\geq \frac{1}{\ln 2} \sum_{x,y,z} p(x, y, z) \left(1 - \frac{p(x, y) p(y, z)}{p(x, y, z) p(y)} \right)$$

$$= \frac{1}{\ln 2} \sum_{x,y,z} \left(p(x, y, z) - \frac{p(x, y) p(y, z)}{p(y)} \right)$$

$$= \frac{1}{\ln 2} \sum_{x,y,z} p(y) (p(x, z|y) - p(x|y) p(z|y))$$

$$\geq 0$$

We see that this is true because the summation will always be bigger then 0, unless if $p(x, z|y) = p(x|y) p(z|y)$ for all $x, y, z \in X, Y, Z$ which is exactly the case if and only if X is conditional independent of Z given Y. Thus as required we get $H(X, Y, Z) + H(Y) \leq H(X, Y) + H(Y, Z)$, with equality if and only if X is conditional independent of Z given Y.

$\diamond$

2.2 Lemma's

Lemma 2.2 *Let P be a JPD over n random variables and I_P its CI-trace. Then the following holds,*

$$(A, B, C) \in I_P \Leftrightarrow H(X_A|X_B, X_C) = H(X_A|X_C).$$

Proof
Re-write $H(X_A|X_B, X_C) = H(X_A|X_C)$ as $H(X_A, X_B, X_C) - H(X_B, X_C) = H(X_A, X_C) - H(X_C)$ and then as $H(X_A, X_B, X_C) + H(X_C) = H(X_A, X_C) + H(X_B, X_C)$ and use 6 of thm 2.1 on page 4.

2.3   The proposition

Proposition 2.3 *Let I be a CI-trace over $n > 3$ random variables and $2 \leq k \leq n$. Consider the operator successor on $[k]$. Then the following two statements are equivalent;*
  i) $\forall j \in [k] : (0, j, \mathsf{suc}(j)) \in I$,
  ii) $\forall j \in [k] : (0, \mathsf{suc}(j), j) \in I$.

  Proof We will use the following two statements for the proof,
  1) $H(X_A, X_B, X_C) + H(X_C) \leq H(X_B, X_C) + H(X_A, X_C)$
  2)   1) holds with equality iff $(A, B, C) \in I$
The first expression we have already seen in section 2.1 and the proof of the second statement is given in lemma 2.2.
i)$\Rightarrow$ ii) : Now using 2) on 1) given that the first statement i) is true we get the following for all $j \in [k]$:

$$H(X_0, X_j, X_{\mathsf{suc}(j)}) + H(X_{\mathsf{suc}(j)}) - H(X_j, X_{\mathsf{suc}(j)}) - H(X_0, X_{\mathsf{suc}(j)}) = 0$$

Next we shall take the summation and see that by a using simple summation rules and small adjustment of the integers we will get the same summation that suggest another kind of dependencies.

$$
\begin{aligned}
0 &= \sum_{j=1}^{k} H(X_0, X_j, X_{\mathsf{suc}(j)}) + H(X_{\mathsf{suc}(j)}) - H(X_j, X_{\mathsf{suc}(j)}) - H(X_0, X_{\mathsf{suc}(j)}) \\
&= \sum_{j=1}^{k} H(X_0, X_j, X_{\mathsf{suc}(j)}) + \sum_{j=1}^{k} H(X_{\mathsf{suc}(j)}) \\
&\qquad - \sum_{j=1}^{k} H(X_j, X_{\mathsf{suc}(j)}) - \sum_{j=1}^{k} H(X_0, X_{\mathsf{suc}(j)}) \\
&= \sum_{j=1}^{k} H(X_0, X_j, X_{\mathsf{suc}(j)}) + \sum_{j=1}^{k} H(X_j) \\
&\qquad - \sum_{j=1}^{k} H(X_j, X_{\mathsf{suc}(j)}) - \sum_{j=1}^{k} H(X_0, X_j) \\
&= \sum_{j=1}^{k} H(X_0, X_{\mathsf{suc}(j)}, X_j) + H(X_j) - H(X_j, X_{\mathsf{suc}(j)}) - H(X_0, X_j)
\end{aligned}
$$

By 1) we know that for all $j \in \{1, \ldots, k\}$,

$$H(X_0, X_{\mathsf{suc}(j)}, X_j) + H(X_j) - H(X_j, X_{\mathsf{suc}(j)}) - H(X_0, X_j) \leq 0,$$

thus by the equations above they should all be equal to 0. But that means by 2) that for all $j \in \{1, \ldots, k\}$ $(0, \mathsf{suc}(j), j) \in I$ and thus that the second statement, ii), is also true.

ii)$\Leftarrow$ i) : The proof is completely analogue to the one above. Only the integers j and $\mathsf{suc}(j)$ should be interchanged.

$\diamond$

3 Constructing the JPD

3.1 Requirements

Lemma 3.1 *Let I, J be two CI-traces both over $n + 1$ random variables. Then $I \cap J$ is also a CI-trace.*

Proof.

Suppose that $P \in \mathcal{P}([n])$ is a JPD on the random variables $X_0, X_1, \ldots, X_n$ and $Q \in \mathcal{P}([n])$ is a JPD on the random variables $Y_0, Y_1, \ldots, Y_n$ with resp. CI-traces I_P and I_Q. Define the JPD $R \in \mathcal{P}([n])$ on the random variables $Z_0, Z_1, \ldots, Z_n$, with $Z_i = (X_i, Y_i)$, i.e. $z_i = (x_i, y_i)$, for all $i \in [n]$ as follows:

$$R(Z = z) = P(X = x)Q(Y = y), \quad \text{with} x \in \mathcal{E}_X, y \in \mathcal{E}_Y, \text{ and } z \in \mathcal{E}_Z$$

We only need to show that R has $I_R = (I_P \cap I_Q)$ as CI-trace. Suppose $(A, B, C) \in I_R$ then $Z_A \perp Z_B | Z_C = (X_A, Y_A) \perp (X_A, Y_B) | (X_C, Y_C)$. By *decomposition* we have $X_A \perp X_B | (X_C, Y_C)$. Now we find

$$p(x_A, x_B, x_C, y_C)p(x_C, y_C) = p(x_A, x_B, x_C)p(x_C)q(y_C)^2$$

Furthermore

$$
\begin{aligned}
p(x_A, x_B, x_C, y_C)p(x_C, y_C) &= p(x_A, x_C, y_C)p(x_B, x_C, y_C) \\
&= p(x_A, x_C)p(x_B, x_C)p(y_C)^2
\end{aligned}
$$

Hence

$$p(x_A, x_B, x_C)p(x_C) = p(x_A, x_C)p(x_B, x_C),$$

so $(A, B, C) \in I_P$. Completely analogue we can find that $(A, B, C) \in I_Q$ and as a result we have that $I_R \subset I_P \cap I_Q$.

On the other hand if $(A, B, C) \in I_P \cap I_Q$ we have

$$
\begin{aligned}
& p(x_A, x_B, x_C)p(x_C)p(y_A, y_B, y_C)p(y_C) \\
&= p(x_A, y_A, x_B, y_B, x_C, y_C)p(x_C, y_C), \\
&= p(z_A, z_B, z_C)p(z_C).
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
& p(x_A, x_B, x_C)p(x_C)p(y_A, y_B, y_C)p(y_C) \\
&= p(x_A, x_C)p(x_B, x_C)p(y_A, y_C)p(y_B, y_C), \\
&= p(x_A, y_A, x_C, y_C)p(x_B, y_B, x_C, y_C), \\
&= p(z_A, z_C)p(z_B, z_C).
\end{aligned}
$$

Hence

$$p(z_A, z_B, z_C)p(z_C) = p(z_A, z_C)p(z_B, z_C),$$

so $(A, B, C) \in I_R$. As needed to be proven $I_R = I_P \cap I_Q$ is indeed a CI-trace.

3.1.1 Parity Construction

Let $n \geq 3$ and $D \subset [n]$ such that $|D| \geq 2$. Then the following CI-trace I^D exists,

$$I^D = \{(A, B, C) \; ; \; A \cap D = \emptyset \text{ or } B \cap D = \emptyset \text{ or } D \not\subset A \cup B \cup C\}$$

3.1.2 Four State Construction

Let $n \geq 3$. Then there always exists a CI-trace K such that
- $(0, i, j) \in K$ whenever $i, j \in [n] \setminus \{0\} : i \neq j$,
- $(i, j, 0) \notin K$ whenever $i, j \in [n] \setminus \{0\} : i \neq j$,
- $(A, B, \emptyset) \notin K$ whenever $A, B \neq \emptyset$.

3.1.3 Increase-Hold Construction

Let $n \geq 3$. Then there exists a CI-trace J such that
- $(0, j, \mathsf{suc}(j)) \in J$ whenever $j \in [n] \setminus \{0, n\}$,
- $(0, \mathsf{suc}(j), j) \notin J$ whenever $j \in [n] \setminus \{0, n\}$.

3.2 The CI-trace

Lemma 3.2 *Let $n \geq 3$ and $s \in [n] \setminus \{0\}$, then*

$$I = [T([n]) \setminus T_*([n])] \cup \left(\bigcup_{j \in [n] \setminus \{0, s\}} \{(0, j, \mathsf{suc}(j)), (j, 0, \mathsf{suc}(j))\} \right)$$

is a CI-trace.

Proof

W.l.o.g. we can assume $s = n$ Thus we denote

$$I = [T([n]) \setminus T_*([n])] \cup \left(\bigcup_{j=1}^{n-1} \{(0, j, \mathsf{suc}(j)), (j, 0, \mathsf{suc}(j))\} \right).$$

To show that I is a CI-trace we put $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$ where

$\mathcal{D}_1 \quad = \quad \{D \subset N; |D| = 4\}$

$\mathcal{D}_2 \quad = \quad \{D \subset N; |D| = 3 \wedge D \neq \{0, j, \mathsf{suc}(j)\}$ for every $j = 1, \ldots, n-1\}$.

Consider the dependency models I^D for $D \in \mathcal{D}$, K and J constructed above in resp. 3.1.1, 3.1.2 and 3.1.3. By Lemma 3.1 we have $L = K \cap J \cap \left(\cap_{D \in \mathcal{D}} I^D \right)$ is a CI-trace. It is easy to see that $I \subset L$.

9

To prove that $L \subset I$ we need to show that for any $u = (A, B, C) \notin I$ we have $u \notin L$. That means that the following two cases should be ruled out for any $u \in L \setminus I$; $C = \emptyset$ and $|[u]| \geq 3$:

$C = \emptyset$:

Then $u \notin K$ by the third condition of the construction of K and thus $u \notin L$. In particular this shows that any non-trivial $(A, B, C) \in T([n]) \setminus I$ with $|[u]| = 2$ is not an element of L.

$|A \cup B \cup C| = 3$:

We consider two cases; either $[u] \in \mathcal{D}_2$ or $[u] \notin \mathcal{D}_2$. In the first case we know $\exists D \in \mathcal{D}_2$ such that $D \subset [u]$ but that implies that $u \notin \cap_{D \in \mathcal{D}_2} I^D$ so $u \notin L$.

Secondly suppose $[u] \notin \mathcal{D}_2$ and $C \neq \emptyset$ then $[u] = \{0, j, \mathsf{suc}(j)\}$ for some $j \in \{1, \ldots, n-1\}$. We can have the following two cases;

- $u \in \{(j, \mathsf{suc}(j), 0), (\mathsf{suc}(j), j, 0)\}$. But that means that $u \notin K$ because of the second condition of the construction of K and thus $u \notin L$.

- $u \in \{(0, \mathsf{suc}(j), j), (\mathsf{suc}(j), 0, j)\}$. But that means that $u \notin J$ because of the second condition of the construction of J and thus $u \notin L$.

$|A \cup B \cup C| \geq 4$:

Then $\exists D \subset N$ such that $D \in \mathcal{D}_1$ and $D \subset [u]$. This implies $u \notin \cap_{D \in \mathcal{D}_1} I^D$. But then we also have that $u \notin L$.

So if $u \in T([n]) \setminus I$ then $u \notin L$. Thus $I = K \cap J \cap \left(\cap_{D \in \mathcal{D}} I^D \right)$.

$\diamond$

3.3   Proofs of the Constructions

Proof of the Parity Construction

If we can construct a JPD such that I^D is its CI-trace the construction is proven correct. To do so take $n + 1$ random variables, such that X_i takes values in

$$\mathcal{E}_i = \left\{ \begin{array}{ll} \{0, 1\} & \text{if } i \in D, \\ \{0\} & \text{if } i \notin D. \end{array} \right\}$$

Furthermore we take a JPD $P_D \in \mathcal{P}([n])$ as follows

$$p(x) = \left\{ \begin{array}{cl} 2^{1-|D|} & \text{if } \sum_{i \in D} x_i \text{ is even,} \\ 0 & \text{if } \sum_{i \in D} x_i \text{ is odd.} \end{array} \right\}$$

First we will show that $I^D \subset I_P$. Let $u = (A, B, C) \in I^D$ then we have three possibilities, if:

$A \cap D = \emptyset$:

We know that A contains only indices i for which we have $\mathcal{E}_i = \{0\}$. This means that all random variables indexed by A are deterministic and hence we know $p(x_A, x_B, x_C) = p(x_B, x_C)$ and $p(x_A, x_C) = p(x_C)$ (i.e. $p(x_A, x_B, x_C)p(x_C) = p(x_A, x_C)p(x_B, x_C)$). So $u \in I_P$.

$B \cap D = \emptyset$:

Analogue to the proof above.

$D \not\subset A \cup B \cup C$:

This implies $\exists j \in D$ such that $j \notin [u]$. Now take the marginal distribution on $[n] \setminus \{j\}$ other random variables. It's immediate that this is a uniform distribution on the outcomes of the $[n] \setminus \{j\}$ remaining random variables. Hence all random variables of $[n] \setminus \{j\}$ are independent. so certainly $A \perp B | C$, hence $u \in I^D$

To show that $I_P \subset I^D$ we only need to prove that I_P does not contain anything else. Suppose that $u = (A, B, C) \in I_P$ but $u \notin I^D$ than we know that neither A and B are empty and they both contain at least one random variable that is also in D and that $D \subset [u]$. But then A can never be conditional independent of B given C because the $\sum_{i \in D} x_i = even$.

Thus $I_P = I^D$ and indeed $I^D \in \text{CIR}(N)$.

$\diamond$

Proof of the Four-State Construction

Take $n + 1$ random variables such that X_i takes values in $\mathcal{E}_i = \{0, 1\}$ for $i \in [n]$ and let $P \in \mathcal{P}([n])$ be defined as follows

$$\begin{aligned} P(X_0 = 0, X_{[n] \setminus \{0\}} = \bar{0}) &= a_1, \\ P(X_0 = 1, X_{[n] \setminus \{0\}} = \bar{0}) &= a_2, \\ P(X_0 = 0, X_{[n] \setminus \{0\}} = \bar{1}) &= a_3, \\ P(X_0 = 1, X_{[n] \setminus \{0\}} = \bar{1}) &= a_4, \end{aligned}$$

such that $a_j > 0$ for $j \in \{1, 2, 3, 4\}$, $\sum_j a_j = 1$ and $a_1 a_4 \neq a_2 a_3$.

Take $(0, i, j) \in T([n])$ such that $i \neq j$ then there are eight possibilities outcomes I will list only two. The others are similar:

$$
\begin{aligned}
a_1(a_1 + a_2) &= P((X_0, X_i, X_j) = (0, 0, 0))P(X_j = 0) \\
&= P((X_0, X_j) = (0, 0))P((X_i, X_j) = (0, 0)) = a_1(a_1 + a_2) \\
0 \times (a_1 + a_2) &= P((X_0, X_i, X_j) = (1, 1, 0))P(X_j = 0) \\
&= P((X_0, X_j) = (1, 0))P((X_i, X_j) = (1, 0)) = a_1 \times 0
\end{aligned}
$$

So $(0, i, j) \in I_P$.

Take $(i, j, 0) \in T([n])$ such that $i \neq j$. We have

$$
\begin{aligned}
P((X_i, X_j, X_0) = (1, 0, 0)) &= 0, \\
P((X_0, X_j) = (0, 0)) &= a_1, \\
P((X_0, X_i) = (0, 1)) &= a_3.
\end{aligned}
$$

Hence $(i, j, 0) \notin I_P$.

Take $(A, B, \emptyset) \in T([n]) \setminus T_*([n])$. If A and $B \subset [n] \setminus \{0\}$ its immediate that there are not independent. So suppose w.l.o.g. that $0 \in A$. Then we have

$$
\begin{aligned}
P(X_A = \bar{0}) &= a_1, \\
P(X_B = \bar{1}) &= a_3 + a_4, \\
P(X_A = \bar{0}, X_B = \bar{1}) &= 0,
\end{aligned}
$$

Hence $p(x_A)p(x_B) \neq p(x_A, x_B)$ which implies $(A, B, \emptyset) \notin I_P$.
So P is a JPD that has a CI-trace that has the properties of K.

$\diamond$

Proof of the Increase-hold construction
Take $n + 1$ random variables, such that X_i takes values in

$$
\mathcal{E}_i = \left\{ \begin{array}{ll} \{1, \ldots, i\} & \text{for } i \in [n] \setminus \{0\}, \\ \{1, \ldots, n\} & \text{for } i = 0. \end{array} \right\}
$$

Furthermore we take a JPD $P \in \mathcal{P}([n])$ as follows $p(a^k) = \frac{1}{n}$, with $a^k = (a_0^k, a_1^k, \ldots, a_n^k) \in \mathcal{E}$, such that

$$
a_i^k = \left\{ \begin{array}{ll} \min\{i, k\} & \text{for } i \in [n] \setminus \{0\} \\ k & \text{for } i = 0. \end{array} \right\}
$$

Take $(0, j, \mathsf{suc}(j)) \in T([n])$ such that $j \in \{1, \dots, n-1\}$ and $k \in [n] \setminus \{0\}$. We distinguish two cases; either $k \geq j+1$ or $k < j+1$. In the first case $x_j = j$ or $x_j = k$ and we get

$$
\begin{aligned}
\tfrac{1}{n}\tfrac{1}{n} &= P((X_0, X_j, X_{\mathsf{suc}(j)}) = (k, j, k))P(X_{\mathsf{suc}(j)} = k) \\
&= P((X_0, X_{\mathsf{suc}(j)}) = (k, k))P((X_j, X_{\mathsf{suc}(j)}) = (j, k)) \quad = \quad \tfrac{1}{n}\tfrac{1}{n} \\
\tfrac{1}{n}\tfrac{1}{n} &= P((X_0, X_j, X_{\mathsf{suc}(j)}) = (k, k, k))P(X_{\mathsf{suc}(j)} = k) \\
&= P((X_0, X_{\mathsf{suc}(j)}) = (k, k))P((X_j, X_{\mathsf{suc}(j)}) = (k, k)) \quad = \quad \tfrac{1}{n}\tfrac{1}{n}
\end{aligned}
$$

In the second case $k \geq j+1$:

$$
\begin{aligned}
\tfrac{1}{n}\tfrac{n-j}{n} &= P((X_0, X_j, X_{\mathsf{suc}(j)}) = (k, j, j+1))P(X_{\mathsf{suc}(j)} = j+1) \\
&= P((X_0, X_{\mathsf{suc}(j)}) = (k, j+1))P((X_j, X_{\mathsf{suc}(j)}) = (j, j+1)) \\
&= \tfrac{1}{n}\tfrac{n-j}{n}
\end{aligned}
$$

and thus indeed we have $(0, j, \mathsf{suc}(j)) \in I_P$

Take $(0, \mathsf{suc}(j), j) \in T([n])$ such that $j \in \{1, \dots, n-1\}$ and $k = j+1$. Then

$$
\begin{aligned}
\tfrac{1}{n}\tfrac{n-j+1}{n} &= P((X_0, X_{\mathsf{suc}(j)}, X_j) = (k, j+1, j))P(X_j = j) \\
&\neq P((X_0, X_j) = (k, j))P((X_{\mathsf{suc}(j)}, X_j) = (j+1, j)) \quad = \quad \tfrac{1}{n}\tfrac{n-j}{n}
\end{aligned}
$$

So $(0, \mathsf{suc}(j), j) \notin I_P$. So there exists a CI-trace J with these properties.

$\diamond$

4 Studený's Theorem

Theorem 4.1 (Studený '92) *No finite set of CI-rules, say $\mathcal{R}_0, \mathcal{R}_1, \ldots, \mathcal{R}_p$, can characterise all CI-traces, i.e. the following does not hold*

$$(I \subset T([n]) \text{ is a } CI - trace \iff I \text{ closed under } \mathcal{R}_0, \mathcal{R}_1, \ldots, \mathcal{R}_p)$$

In words: an arbitrary set of independencies can not be extended to a CI-trace by closing it under a finite set of CI-rules.

Proof

Suppose that we can characterise all CI-traces with a finite set of CI-rules, $\mathcal{R}_0, \mathcal{R}_1, \ldots, \mathcal{R}_p$. Take $n \in \mathbb{N}_{>m}$, with $m = \max_{i \in [p]}(\mathsf{ord}(\mathcal{R}_i))$. We define the following CI-pattern

$$I = [T([n]) \setminus T_*([n])] \cup \left(\bigcup_{j \in [n] \setminus \{0\}} \{(0, j, \mathsf{suc}(j)), (j, 0, \mathsf{suc}(j))\} \right).$$

Let $K = \{u_1, \ldots, u_{\mathsf{ord}(\mathcal{R}_i)} \in I\}$ for $i \in [p]$. If $\mathcal{R}_i$ can be applied on K call its consequent u_c. The set K contains triplets involving at most m of the $n > m$ pairs $j, \mathsf{suc}(j)$. So for some s, both the triplets $(0, s, \mathsf{suc}(s))$ and $(0, \mathsf{suc}(s), s)$ are not in K. Now Lemma 3.2 on page 9 gives us the following CI-trace

$$J = [T([n]) \setminus T_*([n])] \cup \left(\bigcup_{j \in [n] \setminus \{0,s\}} \{(0, j, \mathsf{suc}(j)), (j, 0, \mathsf{suc}(j))\} \right).$$

Now we have $K \subset J \subset I$. So if CI-rule $\mathcal{R}_i$ can be applied on K, with consequent, say u_c, it means that $u_c \in J$ and hence $u_c \in I$. But that means that I is closed under $\mathcal{R}_0, \mathcal{R}_1 \ldots, \mathcal{R}_p$ and thus I should be a CI-trace. Which is in contradiction with Proposition 2.3 on page 6. And the statement is true:

No finite set of CI-rules can characterise all CI-traces

$\diamond$

5 Conclusion

There exists no set of finite CI-rules such that we can characterise all CI-traces without a priori knowing its corresponding JPD. Milan Studený proved this already in 1992 and so did I. Of course that would not have been possible without his work. However I do believe that I have written an easier readable version then he had done.

The proof was based on a contradiction containing the following steps

1. Suppose there exist such a characterisation.

2. Construct a set I of triplets over at least $n + 1$ random variables, such that n is larger then the largest order of the antecedents, say m. Consisting of all trivial triplets and all triplets of the form $(0, j, \mathsf{suc}(j))$ and of course their mirror images.

3. For each subset K of I of size $|K| \leq m$, there exist a s such that $(0, s, \mathsf{suc}(s))$ and its mirror image are not contained in K.

4. Construct a CI-trace with Lemma 3.2 that contains that subset and hence its consequent. This CI-trace is a subset of I hence the I contains the consequent. So I should be a CI-trace.

5. Proposition 2.3 tells us that I should also contain the triplets of the form $(0, \mathsf{suc}(j), j)$ and its mirror images. Nevertheless those are not included, hence I should not be a CI-trace.

6. This contradiction tells us that the assumption is false and hence we find that there does not exists a finite characterisation of all CI-traces.

This thesis does not contain more then the theorem itself and all elements needed to proof the statement. Studený did the same in 7 pages, though he did almost need a whole other article to explain Proposition 2.3, didn't give the proofs to the three constructions needed for Lemma 3.2 and I dare say couldn't explain the concept about the CI-rules very clearly.

My first contribution is made at point 5. The requirements to prove Proposition 2.3 are made simple and clear in only three pages. I took the idea behind another article of Studený's, *Multiinformation and the problem of characterisation of conditional independence relations*, which is the theory behind *Shannon Entropies*. With that idea the statements in the Proposition were made immediately clear.

My second and last contribution is the whole thesis as one piece, at least I hope that anybody that reads my thesis will understand what Studený did in his article more easily.

6 Bibligraphy

M.Studený (1988). *Multiinformation and the problem of characterization of conditional independence relations*, Problem of Control and Information Theory, Vol 18 (1), pp. 3-16, Prague.

M. Studený (1992). *Conditional Independence Realisation have no finite characterization*, Prague.

T.L. Fine (1973). *Theories of Probability an examination of foundations*, Academic Press, New York.

J.A. Rice (1995). *Mathematical Statistics and Data Analysis*, Duxbury Press, California.

J.F.C. Kingman, S.J. Taylor (1966). *Introduction to Measure and Probability*, Cambridge University Press.

M.A. Nielsen, I.L Chuang (2000). *Quantum Computation and Quantum Information*, Cambridge University Press.

S.L. Lauritzen (1996). *Graphical Models*, Oxford University Press.

J. Pearl (2000). *Causality*, Cambridge University Press.