15.1 Conditional Independence

Formal properties must be defined so that we can convert from a graph to a probability distribution, and convert from a distribution back to a graph.

Recall the definition of conditional independence:

\[ X \perp\!\!\!\!\!\!\perp Y \mid Z \text{ if } f(x, y, z) = g(x, z)h(y, z) \quad \forall x, y, z \]

15.1.1 Properties of Conditional Independence

Note the following properties apply to sets of variables as well as single variables.

(C1) if \( X \perp\!\!\!\!\!\!\perp Y \mid Z \) then \( Y \perp\!\!\!\!\!\!\perp X \mid Z \)

(C2) if \( X \perp\!\!\!\!\!\!\perp Y \mid Z \) and \( U = h(X) \), then \( U \perp\!\!\!\!\!\!\perp Y \mid Z \)

(C3) if \( X \perp\!\!\!\!\!\!\perp Y \mid Z \) and \( U = h(X) \), then \( X \perp\!\!\!\!\!\!\perp Y \mid (Z, U) \)

(C4) if \( X \perp\!\!\!\!\!\!\perp Y \mid Z \) and \( X \perp\!\!\!\!\!\!\perp W \mid (Y, Z) \), then \( X \perp\!\!\!\!\!\!\perp (W, Y) \mid Z \)

Consider the case analogous to (C4) where the distributions are not conditioned on \( Z \):

If \( X \perp\!\!\!\!\!\!\perp Y \) and \( X \perp\!\!\!\!\!\!\perp W \mid Y \) then \( X \perp\!\!\!\!\!\!\perp (W, Y) \). This implies the following graph:

```
  X
 / \    /
/    \  /   \
Y     W
```

Figure 15.1:
Note that in general it does not imply that $W \perp \perp Y$, shown by the following graph:

![Graph](image)

Figure 15.2:

There are times that a set of conditional independence statements imply a given single graph. Note, however, that sometimes a set of conditional independence statements could be compatible with (and therefore imply) multiple graphs.

(C5) $X \perp \perp Y | Z$ and $X \perp \perp Z | Y$, then $X \perp \perp (Y, Z)$ This property only holds under certain conditions.

**Example of when the property fails:**

Define:

\[
P(X = 0) = P(X = 1) = \frac{1}{2}
\]

\[
P(X = Y = Z = 1) = P(X = Y = Z = 0) = \frac{1}{2}
\]

Other configurations of $X$, $Y$, and $Z$ have 0 probability.

This gives:

\[
p(X|Z) = \frac{p(X, Z)}{p(Z)} = \frac{1}{2} \frac{p(X=Z)}{p(Z)} = 1_{\{X=Z\}}
\]

\[
p(Y|Z) = 1_{\{Y=Z\}}
\]

Where $1_{\{A\}}$ is the indicator function. The indicator function is equal to 1 when $A$ is true and 0 when $A$ is false.

\[
p(X, Y|Z) = \frac{p(X, Y, Z)}{p(Z)} = \frac{1}{2} \frac{1_{\{X=Y=Z\}}}{p(Z)} = 1_{\{X=Y=Z\}}
\]

This gives us $p(X, Y|Z) = p(X|Z)p(Y|Z)$. By the same math, just switching the $Y$ and $Z$, we also get $p(X, Z|Y) = p(X|Y)p(Z|Y)$. Thus,

\[
X \perp \perp Y | Z \text{ and } X \perp \perp Z | Y
\]

It is not true that $X \perp (Y, Z)$ (since $X = Y$, given $Y$, we know $X$, so $X$ cannot be independent of $Y$). Note that 0 probabilities exist in the distribution, so the density is not positive.
Theorem 15.1. When the density is strictly positive such that $f(x, y, z) > 0 \ \forall x, y, z$, then (C5) holds.

Proof. If $f(x, y, z) > 0$, $X \perp \!\!\!\!\!\!\perp Y|Z$, and $X \perp \!\!\!\!\!\!\perp Z|Y$ then there exists strictly positive functions $k, l, g, h$ such that

$$f(x, y, z) = k(x, z) l(y, z) = g(x, y) h(z, y)$$

$$\Rightarrow g(x, y) = \frac{k(x, z) l(y, z)}{h(z, y)}.$$

Fix $z = z_0$

$$\text{define } \pi(x) = k(x, z_0), q(y) = \frac{l(y, z_0)}{h(y, z_0)}$$

$$\Rightarrow g(x, y) = \pi(x) q(y)$$

$$\Rightarrow f(x, y, z) = g(x, y) h(y, z) = \pi(x) (q(y) h(y, z))$$

$$\Rightarrow X \perp \!\!\!\!\!\!\perp Y, Z$$

Gaussians, or any density function from the exponential family of probability distributions, satisfy the positive and continuous condition.

15.1.2 Conditional Independence Analogy

We can use an analogy to help us understand the concepts of conditional independence. Think of $X$, $Y$, and $Z$ as “information” or “knowledge” sources. Define “$\perp \!\!\!\!\!\!\perp$” = “irrelevant” and “$|$” = “knowing”

Then the conditional independence rules can be read as if they are English. For instance, suppose we are talking about learning German. Then rule (C2) can be read as “If, given you are learning German ($Z$), it so happens that whether or not you’ve read textbook $X$ is irrelevant to whether or not you’ve read textbook $Y$ (perhaps they cover two totally different aspects of the language), then reading a modification of textbook $X$, such as just reading one chapter, or reading it backwards (this is $U=f(x)$) is also irrelevant to whether or not you’ve read textbook $Y$”.

15.1.3 Rules (C1) through (C4) form an algebraic system

(C1) through (C4) can be properties of some algebraic system.

**Definition 15.2.** A semi-graphoid is any algebraic system that satisfies the above ternary relations (C1)-(C4).

**Definition 15.3.** A graphoid is any algebraic system that satisfies the above ternary relations (C1)-(C5).

Undirected graphs and simple separation on satisfies (C1)-(C4). If the subsets are disjoint C5 is also satisfied. Let $A, B, C$ be subsets of $V$, where $G=(V,E)$. We use $A \perp \!\!\!\!\!\!\perp B|C$ to mean “$C$ separates $A$ and $B$”. e.g. any path from a node in $A$ to a node in $B$ must go through a node in $C$.

(S1) If $A \perp \!\!\!\!\!\!\perp B|C$ then $B \perp \!\!\!\!\!\!\perp A|C$
(S2) If \(A \perp B|C\) and \(U \subseteq A\), then \(U \perp B|C\)

(S3) If \(A \perp B|C\) and \(U \subseteq B\), then \(A \perp (C \cup U)|B\)

(S4) If \(A \perp B|C\) and \(A \perp D|(B \cup C)\), then \(A \perp (B \cup D)|C\)

(S5) If all subsets are disjoint, \(X \perp\perp Y|Z\), and \(X \perp\perp Z|Y\) then \(X \perp(Y, Z)\)

We will now prove the block independence lemma. This gives us the reverse of (S5). That is, if we know \(Y\) is separated from \(Z\) by \(X\) and \(Z\), and also that \(Y\) is separated from \(Z\) by \(X\) and \(Z\), we can conclude that \(Y\) is separated from \(Z\) given \(X\).

**Block independence lemma**

If \((X, Y, Z_1, Z_2)\) is a partitioned disjoint vector, and if \(f(X, Y, Z_1, Z_2)\) is positive everywhere then the following are equivalent:

a) \(Y \perp(Z_1, Z_2)|X\)

b) \(Y \perp Z_1|\{X, Z_2\}, Y \perp Z_2|\{X, Z_1\}\)

We have already shown that in our proof of property (C5) that \(b\) implies \(a\) under positivity. We will now show that \(a\) implies \(b\).
\[ \hat{g}(Z_2) = 0 \cdot Z_2 = 0 \]
\[ \log f(X, Y, Z_1, Z_2) = g(X, Y) + h(X, Z_1, Z_2) \]
\[ = g(X, Y) + \hat{g}(Z_2) + h(X, Z_1, Z_2) \]
\[ = g(X, Y, Z_2) + h(X_1, Z_1, Z_2) \]
Where we define \( g'(X, Y, Z_2) = g(X, Y) + \hat{g}(Z_2) \)

This factorization implies \( Y \perp \! \! \! \perp Z_1 | \{ X_1, Z_2 \} \). Similarly, one can set \( \hat{g}(Z_1) \) and show \( Y \perp \! \! \! \perp Z_2 | \{ X_1, Z_2 \} \). Note that positivity was not needed to prove this direction.

## 15.2 Graph Semantics

This section discusses a set of properties called Markov Properties, which may or may not hold on undirected graphical models. We will use the following notation in this section.

A graph \( G = (V, E) \), where \( V, E \) are sets of vertices and edges.

Let \( X_\alpha \) be a set of random variables, with \( \alpha \in V \).

For \( A \subseteq V \), define \( X_A = \bigcup_{\alpha \in A} X_\alpha \) and \( X = X_V \).

For \( A \subseteq V \), define \( x_A = \bigcup_{\alpha \in A} x_\alpha \) and \( x = x_V \).

Note, \( X_\alpha \) can be a vector or a scalar.

### 15.2.1 Markov Properties

We now define a set of Markov Properties (also called “semantics”) that a probability measure \( P \) might obey. A probability measure \( P \) is said to obey the following Markov Properties:

(P) Pairwise Markov Property, relative to \( G \), if for any pair \( (\alpha, \beta) \),

\[ \alpha \not\sim \beta \Rightarrow \alpha \perp \! \! \! \perp \beta | V \setminus \{\alpha, \beta\} \]

![Diagram of the Pairwise Markov Property](image)

Figure 15.7: Diagram of the Pairwise Markov Property

This means that any pair of non-adjacent vertices are independent given the rest of the vertices. Also recall that in Gaussians \( X_i \perp \! \! \! \perp X_j | X_{1:N \setminus \{i,j\}} \) if and only if \( K_{i,j} = 0 \). This is an example of the pairwise Markov property. Each zero in the concentration matrix corresponds to a missing edge, making \( X_i \) and \( X_j \) non-adjacent. The pairwise Markov property says that \( X_i \) is independent of \( X_j \) given the rest of the graph.
(L) Local Markov Property, relative to $G$, if for any $\alpha \in V$,

$$\alpha \perp \perp \left(V \setminus \text{cl}(\alpha)\right) \mid \text{bd}(\alpha)$$

One example of a model which has the Local Markov Property is the Markov Random Field.

![Figure 15.8: Diagram of the Local Markov Property. We see that $\alpha$ is independent of the remainder of the graph $V \setminus \text{cl}(\alpha)$ given $\alpha$’s boundary.](image)

(G) Global Markov Property, relative to $G$, if for any tuple $(A, B, S)$ of disjoint subsets of $V$, with $S$ separating $A$ and $B$ in $G$,

$$A \perp \perp B \mid S$$

![Figure 15.9: Diagram of the Global Markov Property.](image)

**Theorem 15.4.** For undirected graphs $G$, and any probability distribution over $X$, $(G) \Rightarrow (L) \Rightarrow (P)$

**Proof.** $(G) \Rightarrow (L)$

Define $S=\text{bd}(\alpha)$, $B=\alpha$, $A=$ everything else. It is obvious that $\text{bd}(\alpha)$ separates $\alpha$ from $V \setminus \text{cl}(\alpha)$. In the same way, we immediately get that $(G) \Rightarrow (P)$. We still need to show that $(L) \Rightarrow (P)$.

**Proof.** $(L) \Rightarrow (P)$

Pick some $\alpha$. Choose $\beta \in V \setminus \text{cl}(\alpha)$ ($\Rightarrow \alpha \not\sim \beta$)

Notice that $\text{bd}(\alpha) \cup ((V \setminus \text{cl}(\alpha)) \setminus \{\beta\}) = V \setminus \{\alpha, \beta\}$
Figure 15.10: Demonstration for why $bd(\alpha) \cup ((V \setminus cl(\alpha)) \setminus \{\beta\}) = V \setminus \{\alpha, \beta\}$. The thing to note is that $\alpha$’s closure is just its boundary with itself included.

\[
(L) \Rightarrow \alpha \perp\!
\perp V \setminus cl(\alpha) \cup bd(\alpha)
\]

and

\[
(C3) \Rightarrow X \perp\!
\perp Y \mid Z, U = h(X) \Rightarrow X \perp\!
\perp Y \mid U
\]

Let $X = \alpha, Y = (V \setminus cl(\alpha)), Z = bd(\alpha), U = f(Y) = Y \setminus \{\beta\}$

Then, by combining the above, we get

\[
\alpha \perp\!
\perp (V \setminus cl(\alpha)) \cup (bd(\alpha) \cup (V \setminus cl(\alpha)) \setminus \{\beta\})
\]

\[
\Rightarrow \alpha \perp\!
\perp (V \setminus cl(\alpha)) \setminus \{\alpha, \beta\}
\]

but $\beta \in V \setminus cl(\alpha) \Rightarrow$ (using C2)

\[
\alpha \perp\!
\perp \beta \mid V \setminus \{\alpha, \beta\}
\]

\[
(P)
\]

Note, it is not necessarily true that $(P) \Rightarrow (L)$ or $(L) \Rightarrow (G)$. However, we do have the following theorem:

**Theorem 15.5.** $(P) \equiv (L) \equiv (G)$ when for all disjoint subsets $A, B, C, D$:

\[
[A \perp B] (C \cup D) \text{ and } A \perp C] (B \cup D) \Rightarrow A \perp (B \cup C) \mid D
\]

This theorem is true when the density functions are positive. This is similar to the property (C5), but with the addition of set D.

**Proof.** We will simply show that $(P) \Rightarrow (G)$, since we already know that $(G) \Rightarrow (L) \Rightarrow (P)$.

Assume $(P)$ and that $S$ separates $A$ and $B$ in $G$, and assume that

\[
[A \perp B] (C \cup D) \text{ and } A \perp C] (B \cup D) \Rightarrow A \perp (B \cup C) \mid D
\]
holds and A and B are non-empty. We will use backwards induction on S to prove the theorem.

**Base Case:**

\[ |S| = n = |V| - 2 \Rightarrow |A| = |B| = 1 \]

By (P), \( |A| = |B| = 1 \Rightarrow \alpha \perp \beta|S \Rightarrow (G) \)

So, for any S such that \( |S| = |V| - 2 \), we have (G)

**Induction:**

Let \( |S| = n < |V| - 2 \). Assume this is true for \( |S| > n \). We have two cases:

**Case I:** \( V = A \cup B \cup S \)

**Case II:** \( A \cup B \cup S \subset V \)

choose \( \alpha \in V \setminus (A \cup B \cup S) \).
(note, we can’t have both $A$ and $B$ connect to $\alpha$ because then $S$ would not be a separator of $A$ and $B$)

Without loss of generality, we assume the first case, which gives use

\[ \alpha \perp \perp B | A \cup S \]

Using (C5) we get:

\[ B \perp \perp (A \cup \{\alpha\}) \mid S \]
\[ \Rightarrow B \perp \perp A \mid S \]
\[ \Rightarrow (G) \]

\[ \square \]

All three Markov Properties (P),(L),(G), hold if we are using positive probability distributions (such as Gaussians, or anything from the exponential family of distributions, or mixtures of them).