Cheat sheet

Definition 1 (Average causal effect)**.** $\mathbb{E}(Y^{a=0})$ vs $\mathbb{E}(Y^{a=1})$.

Definition 2 (Conditional independence)**.** $X \perp\!\!\!\perp Y \mid Z \iff F_{X,Y|Z=z}(x,y) = F_{X|Z=z}(x) \cdot F_{Y|Z=z}(y)\ \forall\ x,y,z,$
where $F_{X,Y|Z=z}(x,y) = P(X \leq x, Y \leq y \mid Z = z)$.

Definition 3 (Additive interaction)**.** There is additive interaction if

$$\mathbb{E}(Y^{a=0,e=0}) - \mathbb{E}(Y^{a=1,e=0}) \neq \mathbb{E}(Y^{a=0,e=1}) - \mathbb{E}(Y^{a=1,e=1}).$$

Definition 4 (Multiplicative interaction)**.** There is multiplicative interaction if

$$\frac{\mathbb{E}(Y^{a=0,e=0})}{\mathbb{E}(Y^{a=1,e=0})} \neq \frac{\mathbb{E}(Y^{a=0,e=1})}{\mathbb{E}(Y^{a=1,e=1})}.$$

Definition 5 (Topological order of a DAG)**.** The nodes $V_1, V_2, \ldots$ follow a topological order relative to a DAG $\mathcal{G}$, if V_j is not an ancestor of V_i whenever $j > i$.

Definition 6 (Bayesian network)**.** A Bayesian Network with respect to a DAG $\mathcal{G}$ with nodes $V = (V_1, \ldots, V_m)$ is a statistical model for the random vector V specifying that V belongs to the collection of laws $\mathcal{B}$ satisfying the Markovian factorization

$$p(v) = \prod_{j=1}^{m} p(v_j \mid pa_j)$$

Here, $p(x \mid y) \equiv P(X = x \mid Y = y)$.

Theorem 1 (Graphoid axioms)**.** Let X, Y, Z, W be random variables on a Cartesian product space. Conditional independence satisfies

(1) $X \perp\!\!\!\perp Y \mid Z \implies Y \perp\!\!\!\perp X \mid Z$ (Symmetry)
(2) $X \perp\!\!\!\perp Y, W \mid Z \implies X \perp\!\!\!\perp Y \mid Z$ (Decomposition)
(3) $X \perp\!\!\!\perp Y, W \mid Z \implies X \perp\!\!\!\perp W \mid Y, Z$ (Weak union)
(4) $X \perp\!\!\!\perp W \mid Y, Z$ and $X \perp\!\!\!\perp Y \mid Z \implies X \perp\!\!\!\perp Y, W \mid Z$ (Contraction)
(5) If $p(x,y,z,w) > 0$, then $X \perp\!\!\!\perp W \mid Y, Z$ and $X \perp\!\!\!\perp Y \mid W, Z \implies X \perp\!\!\!\perp Y, W \mid Z$ (Intersection)

Definition 7 (d-separation of a path)**.** A path r is d-separated by a set of nodes Z iff

(1) r contains a chain $V_i \to V_j \to V_k$ or a fork $V_i \leftarrow V_j \to V_k$ such that V_j is in Z, or
(2) r contains a collider $V_i \to V_j \leftarrow V_k$ such that V_j *is not* in Z and such that no descendant of V_j is in Z.

Otherwise the path is d-connected.

Definition 8 (d-separation of two nodes)**.** Nodes V_i and V_k are d-separated by a set of nodes Z if all trails between V_i and V_k are d-separated by Z. We write d-separation as

$$(V_i \perp\!\!\!\perp V_k \mid Z)_G .$$

If V_i and V_k are not d-separated, they are d-connected and we write

$$(V_i \not\perp\!\!\!\perp V_k \mid Z)_G .$$

Definition 9. A law $\mathbb{P}$ is faithful to a DAG $\mathcal{G}$ if for any disjoint set of nodes A,B,C we have that $A \perp\!\!\!\perp C \mid B$ under $\mathbb{P}$ implies $(A \perp\!\!\!\perp C \mid B)_{\mathcal{G}}$.

Definition 10 (d-separation of a path in a SWIG)**.** A path r is d-separated by a set of nodes Z iff

(1) r contains a chain $V_i \to V_j \to V_k$ or a fork $V_i \leftarrow V_j \to V_k$ such that V_j is in Z, or
(2) r contains a collider $V_i \to V_j \leftarrow V_k$ such that V_j *is not* in Z and such that no descendant of V_j is in Z.

If a path is not d-separated by Z **and** there is no fixed node on the path, then the path is d-connected given Z.

Definition 11 (g-formulae)**.**

$$b_{\overline{a}'}(y) = \sum_{\overline{a}_K} \sum_{\overline{l}_K} p(y \mid \overline{l}_K, \overline{a}_K) \prod_{j=0}^{K} p(l_j \mid \overline{l}_{j-1}, \overline{a}_{j-1}) I(a_j = a'_j),$$

$$b_g(y) = \sum_{\overline{a}_K} \sum_{\overline{l}_K} p(y \mid \overline{l}_K, \overline{a}_K) \prod_{j=0}^{K} p(l_j \mid \overline{l}_{j-1}, \overline{a}_{j-1}) p^g(a_j \mid \overline{l}_j),$$

where $\overline{l}_k = (l_0, \ldots, l_k)$, $k \leq K$, are instantiations of **observed** variables and $p^g(a_j \mid \overline{l}_j)$ is the density of A_k^{g+} given $\overline{L}_k^g$, which is determined by g_k.